université PARIS-SACLAY

U-statistics of row-column exchangeable matrices

Application to ecological network analysis

U-statistiques de matrices échangeables ligne-colonne
Application à l'analyse de réseaux écologiques

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°574, mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées
Graduate School : Mathématiques. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **UMR MIA Paris-Saclay (Université Paris-Saclay, AgroParisTech, INRAE)**, sous la direction de **Stéphane ROBIN**, Professeur, la co-direction de **Sophie DONNET**, Directrice de recherche, et la co-direction de **François MASSOL**, Directeur de recherche

Thèse soutenue à Paris-Saclay, le 13 novembre 2023, par

Tâm LE MINH

Composition du jury

Membres du jury avec voix délibérative

Sylvain ARLOT Professeur, Université Paris-Saclay	Président
Yohann DE CASTRO Professeur, École Centrale de Lyon	Rapporteur & Examinateur
Gesine REINERT Professeure, Université d'Oxford	Rapporteure & Examinatrice
François MUNOZ Professeur, Université Grenoble-Alpes	Examinateur
Tabea REBAFKA Maîtresse de conférences, Sorbonne Université	Examinatrice

THESE DE DOCTORAT
NNT : 2023UPASM027

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre : U-statistiques de matrices échangeables ligne-colonne : application à l'analyse de réseaux écologiques

Mots clés : U-statistiques, échangeabilité, statistique asymptotique, modèles de graphes aléatoires, réseaux écologiques

Résumé : Le travail présenté dans cette thèse est essentiellement théorique, mais motivé par des applications écologiques. Les réseaux d'interactions écologiques représentent le fonctionnement d'un écosystème. L'étude de la variabilité des réseaux d'interactions permet de comprendre comment les écosystèmes sont affectés par des facteurs externes. Cette thèse propose une méthodologie d'analyse des réseaux bipartites, applicable aux réseaux écologiques mutualistes. Cette méthodologie est basée sur les U-statistiques de matrices échangeables ligne-colonne. Les matrices échangeables ligne-colonne sont des matrices aléatoires dont la distribution de probabilité jointe est invariante par permutations simultanées des lignes et des colonnes. Les U-statistiques correspondent à la classe des statistiques définies comme la moyenne empirique d'une fonction d'un sous-ensemble, sur tous les sous-ensembles d'observations. Les U-statistiques des matrices sont la moyenne de la fonction d'une sous-matrice sur l'ensemble des matrices. En analyse de réseaux, les matrices échangeables ligne-colonne sont les matrices d'adjacence de réseaux bipartites à nœuds échangeables et les U-statistiques peuvent être utilisées comme estimateurs de quantités d'intérêt. Cette thèse se concentre sur le comportement asymptotique des U-statistiques des matrices échangeables ligne-colonne. Dans la première partie, les martingales inverses sont utilisées pour obtenir un théorème limite sur les U-statistiques de matrices échangeables ligne-colonne. Dans la deuxième partie, une décomposition de type Hoeffding est établie pour ces matrices, ce qui étend le théorème limite précédent. Inspiré par cette décomposition, un estimateur de la variance asymptotique est également suggéré, permettant de proposer une méthode générale pour effectuer des tâches d'inférence statistique sur des modèles de réseaux échangeables. La troisième partie de la thèse étend la méthodologie aux U-statistiques dégénérées, qui ont un taux de convergence plus rapide. Ces développements statistiques sont appliqués à l'analyse des réseaux bipartites, y compris les réseaux écologiques mutualistes. De nombreuses questions écologiques s'intéressent à la structure générale des réseaux plutôt qu'à la liste des espèces présentes. Les modèles de réseaux aléatoires à nœuds échangeables, dont les matrices d'adjacence sont échangeables ligne-colonne, sont donc bien adaptés à l'analyse de ces réseaux. Les U-statistiques sont utilisées comme estimateurs de quantités d'intérêt telles que l'hétérogénéité des degrés, les densités de motifs ou des métriques sur les graphons. Il est possible d'obtenir des garanties statistiques pour ces estimateurs, par exemple sous la forme d'intervalles de confiance, grâce aux résultats théoriques et à la méthodologie développée dans cette thèse. Quelques exemples de modèles de réseaux aléatoires échangeables et de U-statistiques sont donnés, répondant à des questions écologiques réelles. Des études de simulation sont utilisées pour valider l'utilisation de cette méthodologie sur ces exemples.

université PARIS-SACLAY

ÉCOLE DOCTORALE
de mathématiques
Hadamard (EDMH)

Title: U-statistics of row-column exchangeable matrices : application to ecological network analysis

Keywords: U-statistics, exchangeability, asymptotic statistics, random graph models, ecological networks

Abstract: The work presented in this thesis is essentially theoretical, but motivated by ecological applications. Ecological interaction networks represent the functioning of an ecosystem. Investigating the variability of interaction networks enables us to understand how the ecosystems are affected by external factors. This thesis suggests a methodology to analyze bipartite networks, applicable to ecological mutualistic networks. This methodology is based on U-statistics of row-column exchangeable matrices. Row-column exchangeable matrices are random matrices, the joint probability distribution of which is invariant by simultaneous permutations of rows and columns. U-statistics correspond to the class of statistics defined as the empirical mean of a function of a subset, over all subsets of observations. U-statistics of matrices are the average of a submatrix function over the entire matrices. In network analysis, row-column exchangeable matrices are the adjacency matrices of bipartite node-exchangeable networks and U-statistics can be used as estimators of quantities of interest. This thesis focuses on the asymptotic behavior of the U-statistics of row-column exchangeable matrices. In the first part, backward martingales are used to derive a limit theorem on U-statistics of row-column exchangeable matrices. In the second part, a Hoeffding-type decomposition is established for them, which extends the previous limit theorem. Inspired by this decomposition, an estimator of the asymptotic variance is also suggested, making it possible to propose a general method for performing statistical inference tasks on exchangeable network models. The third part of the thesis extends the methodology to degenerate U-statistics, which have a faster rate of convergence. These statistical developments are applied to the analysis of bipartite networks, including mutualistic ecological networks. Many ecological questions are interested in the general structure of networks rather than the collection of present species. This makes exchangeable random network models, the adjacency matrices of which are row-column exchangeable, well-suited to analyze these networks. U-statistics are used as estimators of quantities of interest such as the degree heterogeneity, motif densities or graphon metrics. It possible to obtain statistical guarantees on these estimators, for example in the form of confidence intervals, owing to the theoretical results and the methodology developed in this thesis. Some examples of exchangeable random network models and U-statistics are given, answering real ecological questions. Simulation studies are used to validate the use of this methodology for these examples.


AgroParisTech


INRAE


université
PARIS-SACLAY


Fondation mathématique
FMJH
Jacques Hadamard


AGENCE NATIONALE DE LA RECHERCHE
ANR


MATH
INNOV


Région
îledeFrance

Remerciements

Mes premiers mots de remerciements doivent revenir à mes directeurs de thèse Stéphane, Sophie et François, qui m'ont accompagné pendant cette thèse et avant cela, pendant le stage de master. J'ai pris grand plaisir à travailler avec vous, nos discussions ont été d'une richesse incroyable. Je ne pourrais assez vous remercier pour votre confiance, votre soutien sans faille et surtout votre bienveillance extrême, même quand je doutais de mon travail. Vous m'avez offert un environnement de recherche exceptionnel et je n'aurais pas pu espérer mieux pour ces trois années. Merci aussi pour vos nombreux conseils de carrière, j'espère qu'on aura l'occasion de poursuivre longtemps notre collaboration.

Je suis extrêmement reconnaissant envers Yohann De Castro et Gesine Reinert d'avoir accepté de rapporter cette thèse et de leurs relectures approfondies de mon manuscrit. Je remercie également Sylvain Arlot, François Munoz et Tabea Refabka qui ont accepté de faire partie de mon jury de thèse. Merci pour ces discussions scientifiques fort intéressantes, ouvrant à de nombreuses pistes pour des travaux futurs. Merci à Pierre Latouche, Elisa Thébault et Nicolas Verzelen pour leurs conseils, qui nous ont grandement aidé à aiguiller cette thèse.

Parmi tous les mondes possibles, je n'aurais peut-être jamais pensé faire une thèse si je n'avais pas rencontré Wen-Chin Li et Michele Belotti. Je vous remercie d'avoir cru en moi à des moments où j'étais en proie aux doutes sur mon avenir, et c'est donc en partie grâce à vos encouragements que je suis arrivé là où j'en suis.

Pour cette thèse, j'ai eu la chance d'être accueilli au MIA Paris-Saclay. Merci aux membres du labo d'avoir ensoleillé ces trois années par leur gentillesse et leur humour. Je suis heureux d'avoir croisé vos routes et j'espère que cela se produira à nouveau à l'avenir : Éric A., Farida, Isabelle, Julie, Pierre B., Liliane, Julien, Lucia, Camille, Stéphane, Christelle, Hugo, Jade, Pierre G., José, Sébastien, Nicolas, Joon, Gabriel, Jean-Benoist, Tristan, Gaspar, Sarah, Éric P., Laure, Théodore.

Je dois spécialement mentionner mes frères et sœurs doctorants et post-doctorants. D'abord, merci à mes ainés, qui ont été une grande source d'inspiration : Saint-Clair, Felix, Raphaëlle, Martina; puis, merci à ceux qui sont arrivés avec moi ou en cours de périple, également pour aussi tout ce temps passé ensemble dans et hors du labo : Emre, Bastien, Barbara, Mathilde, Caroline, Annaïg, Armand, Marina, Jérémy, Marion, Mary, Florian, Jules; enfin, merci aux

nouveaux ou futurs doctorants, que je regrette ne de pas avoir pu cotoyer plus longtemps, mais grâce à qui la relève est assurée, j'en suis sûr : Marie, Louis, Jeanne, François, Hayato.

Un grand merci également à toute la troupe toulousaine pour votre amitié indéfectible : Clément, Marc, Ursule, ainsi que Vincent B., Pierre-Marie, Vincent M., Alice, Vojislav. Merci aussi à Barthélémy, Arnaud, Bettina, Arthur, Paul, Killian, Melvyn, Typhaine "M.", Aurélien, Typhaine "R.", Quentin, ainsi que Matéo, Zyn, Florian, Ryusei. Ces trois années auraient été bien moins agréables sans vous.

Enfin, je ne puis jamais assez remercier ma famille : Ngoc Lan, Tuân, et mes parents, sur lesquels j'ai toujours pu compter. Merci de votre soutien inconditionnel. Sans vous, je ne serais probablement jamais arrivé là où j'en suis et je vous en suis éternellement reconnaissant.

Contents

List of Figures

List of Tables

Table of notations

As most chapters have been written as independent articles, notations may vary from one chapter to another. However, some notations are universal. Below is a non-exhaustive list of notations used throughout the thesis.

Abbreviations

a.e.	almost everywhere
a.s.	almost surely
c.d.f.	cumulative distribution function
i.i.d.	independent and identically distributed
AHK	Aldous-Hoover-Kallenberg
RCE	row-column exchangeable
BA	Barabási-Albert model
BEDD	bipartite expected degree distribution model
CM	configuration model
EDD	expected degree distribution model
ER	Erdős-Rényi model
ERGM	exponential random graph model
LBM	latent block model
SBM	stochastic block model
WS	Watts-Strogatz model
CLT	central limit theorem
LIL	law of the iterated logarithm
SLLN	strong law of large numbers
WLLN	weak law of large numbers

Sets and tuples

$A \subseteq B$	A is a subset of B
$A \subset B$	$A \subseteq B$ and $A \neq B$
$\mathbb{S}_n$	The symmetric group of order n
$[\![n]\!]$	The set $\{1, ..., n\}$
$\mathcal{P}_k(A)$	The set of the subsets of A of cardinal k
$(x_1, ..., x_K) \leq (y_1, ..., y_K)$	$\forall 1 \leq k \leq K$, $x_k \leq y_k$
$(x_1, ..., x_K) < (y_1, ..., y_K)$	$(x_1, ..., x_K) \leq (y_1, ..., y_K)$ and $(x_1, ..., x_K) \neq (y_1, ..., y_K)$

Probability

Let X and Y be two square-integrable random variables.

$\mathbb{E}[X]$	The expectation of X
$\mathbb{V}[X]$	The variance of X
$\mathrm{Cov}(X,Y)$	The covariance of X and Y
$\langle X,Y\rangle$	The scalar product of X and Y in the L_2 probability space, i.e. $\mathbb{E}[XY]$
$\mathrm{Cov}(X,Y \mid \mathcal{F})$	The covariance of X and Y conditioned by $\mathcal{F}$, i.e. $\mathbb{E}[XY \mid \mathcal{F}] - \mathbb{E}[X \mid \mathcal{F}]\mathbb{E}[Y \mid \mathcal{F}]$
$\mathcal{B}(p)$	A Bernoulli distribution of parameter $p \in [0,1]$
$\mathcal{P}(\lambda)$	A Poisson distribution of parameter $\lambda > 0$

U-statistics

Let Y be an infinite RCE matrix.

$Y_{(i_1,...,i_p;j_1,...,j_q)}$	The submatrix of Y extracted from the rows $(i_1,...,i_p)$ and the columns $(j_1,...,j_q)$
Y^T	The transpose of Y
$\lvert Y\rvert_1$	$\sum_{ij}\lvert Y_{ij}\rvert$
$Y^{\odot k}$	Y elevated to the element-wise power k, i.e. $Y_{ij}^{\odot k} = Y_{ij}^k$
$\mathrm{Tr}(Y)$	The trace of Y
h	A kernel function of $\mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$
$h(Y_{\{i_1,...,i_p\},\{j_1,...,j_q\}})$	$Y_{(i_1,...,i_p;j_1,...,j_q)}$, where the order of the indices do not matter due to the symmetry of h
$X_{\{i_1,...,i_p\},\{j_1,...,j_q\}}$	$h(Y_{\{i_1,...,i_p\},\{j_1,...,j_q\}})$ in Chapters 2 and 3
$h_{\{i_1,...,i_p\},\{j_1,...,j_q\}}$	$h(Y_{\{i_1,...,i_p\},\{j_1,...,j_q\}})$ in Chapter 4
$U^h_{m,n}(Y)$	The U-statistic of kernel h on the m first rows and n first columns of Y
$\boldsymbol{\xi}$, $\boldsymbol{\eta}$, $\boldsymbol{\zeta}$	The families of AHK variables $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ and $(\zeta_{ij})_{i\geq 1,j\geq 1}$

Bipartite graphs

Let $G = (V_1(G), V_2(G), E(G))$ be a bipartite graph.

$V_1(G)$	The set of the row nodes of G
$V_2(G)$	The set of the column nodes of G
$E(G)$	The set of the edges of G
$A(G)$	The adjacency (or incidence) matrix of G
$v_1(G)$	The number of row nodes in G
$v_2(G)$	The number of column nodes in G
$e(G)$	The number of edges in G
$\lvert\mathrm{Aut}(G)\rvert$	The number of automorphisms of G

Preface

This thesis presents the work I carried out during three years as a doctoral candidate at the MIA Paris laboratory under the supervision of Stéphane Robin, Sophie Donnet and François Massol. This preface is by no means essential to the thesis, but its intend is to prepare the reader before getting to the heart of the matter. First, it sets out the general context of my research, including the questions and elements that have guided the course of this work. Then it introduces some basic notions and definitions that may be useful in understanding the rest of the thesis. Finally, it briefly explains the structure of the thesis, so that the reader is aware of how the chapters relate to each other.

Context

This thesis focuses on the asymptotic behavior of U-statistics on row-column exchangeable matrices. This work is mainly theoretical, but is strongly motivated by applications to the analysis of ecological networks. Indeed, in ecology, most studies do not focus on a single network, but a collection of networks sampled under different conditions, at different locations or at different times. By jointly analyzing the networks in the collection, we hope to obtain information on network variability in space and time, and the influence of external perturbations. However, there is no general methodology for such studies. In the ecological literature, methods are specific to the available data and to the different biological issues under investigation. The aim of this thesis is to propose a more versatile method for studying ecological networks.

There are many types of networks in ecology. The most widely studied are species interaction networks. These networks often focus on a particular type of interaction. These networks have different properties, depending on the type of interaction. We have chosen to focus mainly on bipartite networks, such as pollination networks (Fig. 1). The starting point for this work was therefore to develop a comparison method for these networks. When compared to networks of other fields, ecological networks have specific properties.

Firstly, these networks are the aggregation of individual interaction data. Depending on how we choose to aggregate the data, we may obtain different networks (Fig. 2). For example, we can group interactions at different taxonomic ranks (by species, genus, family, etc.), over

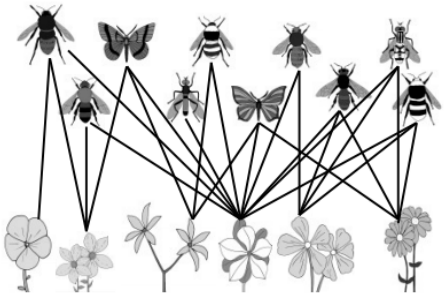

Figure 1 – A binary pollination network links plants to the insects that visit them. Figure taken from Fontaine (2009).

different periods (by day, month, year, etc.), over different geographical areas (by observation site, region, etc.).

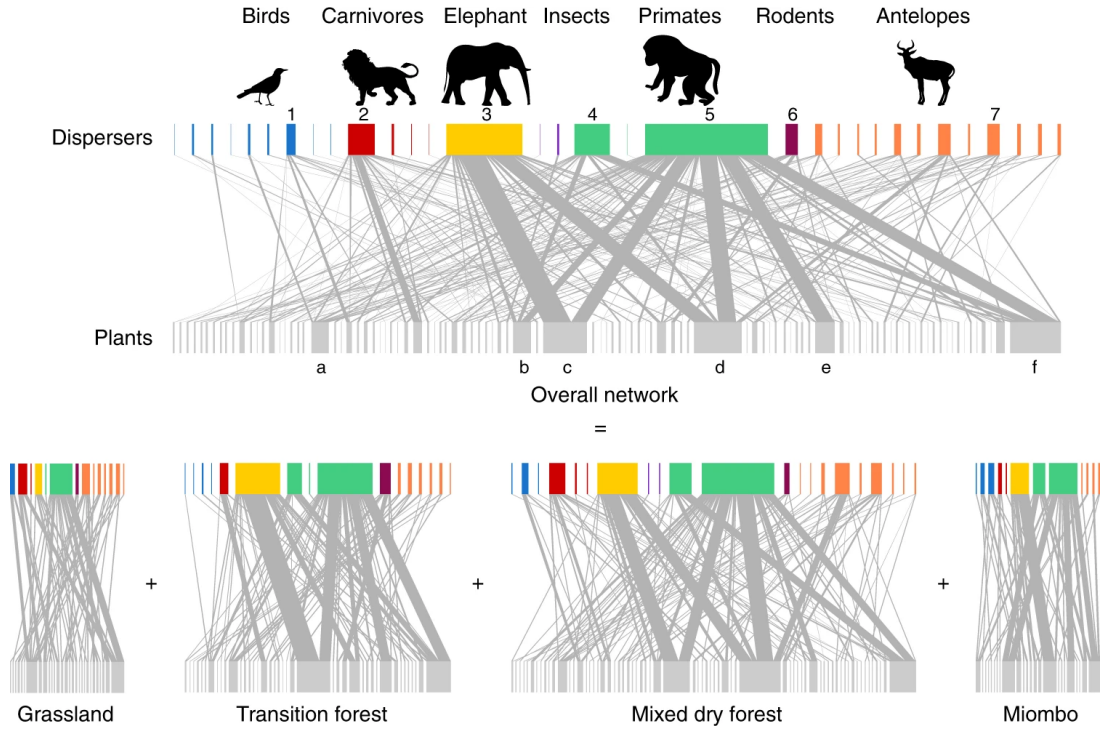

Figure 2 – The question of data aggregation is a recurrent problem in ecological networks. Top: a seed dispersal network. Bottom: the interaction data has been split by vegetation types, leading to different structures. Figure taken from Timóteo et al. (2018).

The sampling method also affects the topology of the network. For example, in pollination networks, a single insect can be sampled as interacting with only one plant species, e.g. if caught using hand net after observation of the visit. The same insect can also be observed to have visited several flowers of this plant, resulting in many interactions, instead of only a

single, on this plant. This same insect could also be sampled as interacting with many plants, if interactions are analyzed through the pollen found on the insect. Despite being the same individual, interactions can be sampled differently, which lead to different data.

Above all, sampling effort plays a crucial role in ecological interaction networks. The reconstructed network is often only part of the complete network, and it is difficult to know how much effort is needed to be certain of having sampled the entire network. There is therefore a notion of uncertainty inherent in the data.

Finally, because of the issue of sampling, or also more simply because the networks are observed in different places or at different times, they do not necessarily involve the same species. Therefore, we often end up with a collection of networks, in which the individual networks have different sets of nodes, potentially of different sizes.

My approach to designing a methodology was therefore motivated by three guidelines. Firstly, the method must not only be able to characterize a network, but above all it must be able to analyze several networks alongside each other. For example, such a method must be able to be used to compare networks. Secondly, it must take into account the specific features of ecological interaction networks, in particular the characteristics described above. Finally, it must be able to answer ecological questions. It is easy to say whether two observed networks are identical, but it is more difficult to know to what extent they differ, whether these differences are significant and above all, how to interpret them from an ecological point of view.

The methodology devised in this thesis is supported by two main pillars: exchangeable random graph models and U-statistics. The first, random graph models, allows us to consider that each observed network is the realization of a certain random model. This makes it possible to capture the source of variability in the networks that is due solely to chance. The legitimacy of using exchangeable models in particular will be justified in the introductory chapter of the thesis. Notably, these models make it possible to obtain interesting properties for U-statistics, which are the second pillar of this methodology. These are a class of statistics generalizing the concept of empirical mean, which will be mainly used as estimators. To construct my methodology, I have established theoretical properties for the mathematical objects involved, i.e. exchangeable random graph models and the U-statistics applied to these models. Nevertheless, the initial application objective has never been lost sight of.

Mathematical notation and basic definitions

Here, I present some notations used in the thesis, then I give some general definitions of network and graph theory. A few notions are briefly discussed, but the experienced reader will be able to skim over this part, which aims more to introduce notations than to explain these notions in detail.

Sets and k-tuples

Let B be a set. We denote $|B|$ or alternatively $\mathrm{Card}(B)$ the cardinal of B. If B is an infinite set, we write $|B| = \infty$ or $\mathrm{Card}(B) = \infty$. In this thesis, $\mathbb{N}$ is the set of positive integers $\{1, 2, ...\}$. Unless explicitly specified, it does not include 0. $[\![n]\!]$ is the set $\{1, ..., n\}$ of the n first elements of $\mathbb{N}$. Let C be another set. If C is a subset of B, we denote $C \subseteq B$. If futhermore, $C \neq B$, then we write $C \subset B$. For $k \in \mathbb{N}$, we denote $\mathcal{P}_k(B)$ the set of all subsets of B with cardinal k.

Let I be a set. A family $(b_i)_{i\in I}$ is a collection of elements b_i belonging to B and indexed by I. We denote $\{b_i : i \in I\}$ the set of elements of this family. If I is ordered, then $(b_i)_{i\in I}$ is a process. In addition, if I is countable, then $(b_i)_{i\in I}$ is a sequence.

An n-tuple is an ordered collection of n elements. These elements can belong to different sets. Suppose that for each $k \in [\![n]\!]$, b_k belongs to a set B_k. The sets of all the n-tuples $(b_1, ..., b_n)$ is denoted $B_1 \times ... \times B_n = \prod_{k=1}^{n} B_k$. If all the sets B_k are ordered, for $1 \leq k \leq n$, and $(a_1, ..., a_n)$ is another n-tuple of $\prod_{k=1}^{n} B_k$, then the notation $(a_1, ..., a_n) \leq (b_1, ..., b_n)$ means that $a_k \leq b_k$, for all $1 \leq k \leq n$. In addition, the notation $(a_1, ..., a_n) < (b_1, ..., b_n)$ means that $(a_1, ..., a_n) \leq (b_1, ..., b_n)$ and $(a_1, ..., a_n) \neq (b_1, ..., b_n)$.

Some notations of analysis and probability

Let E be a set and $f : E \to \mathbb{R}$ a real-valued function. Let $c \in \mathbb{R}$. $f \equiv c$ means that f is the constant function with value c everywhere on its domain E.

Let g be another real-valued function defined on E. $f \overset{a.e.}{=} g$ means that f and g are equal almost everywhere, i.e. they are equal on all their domain except for a set of measure 0.

Let $\gamma(E)$ be an element of E or its boundary, or infinity. $f(e) = o_{e\to\gamma(E)}(g(e))$ means that for all constants $C > 0$, there is a neighborhood of $\gamma(E)$ in which $|f(e)| \leq C|g(e)|$. $f(e) = O_{e\to\gamma(E)}(g(e))$ means that for any neighborhood of $\gamma(E)$, there is a constant $M > 0$ such that $|f(e)| \leq M|g(e)|$ in this neighborhood. $f(e) \asymp_{e\to\gamma(E)} g(e)$ means that we have both $f(e) =$

$O_{e\to\gamma(E)}(g(e))$ and $g(e) = O_{e\to\gamma(E)}(f(e))$. Often, we use these notions for $E = \mathbb{N}$ and $\gamma(E) = \infty$. In this case, there is no ambiguity and we will write more simply $f(n) = o(g(n))$, $f(n) = O(g(n))$ and $f(n) \asymp g(n)$.

Let X and Y be two random variables. The equality of these two variables in distribution is denoted $X \overset{\mathcal{D}}{=} Y$. The almost sure equality, i.e. $\mathbb{P}(X = Y) = 1$, is denoted $X \overset{a.s.}{=} Y$.

Let $(X_n)_{n\geq 1}$ be a sequence of random variables and $(a_n)_{n\geq 1}$ a sequence of constants. $X_n = o_P(a_n)$ means that for all $C > 0$, we have $\lim_{n\to\infty} \mathbb{P}(|X_n| \leq C|a_n|) = 1$. $X_n = O_P(a_n)$ means that for all $\epsilon > 0$, there is a constant $M > 0$ such that $\forall n \geq 1$, we have $\mathbb{P}(|X_n| \leq M|a_n|) < \epsilon$.

Networks and graphs

Networks A *network* is a system of interconnected entities. As such, it is made up of *nodes*, representing the entities, connected by *links*, representing the relationships between the entities. These relationships can sometimes be quantified. In this case, the links can be associated with real values (beyond 0 and 1) and the network is said to be *valued* or *weighted*. Otherwise, the network is said to be *binary*. These relationships can also have a direction. In this case, the network is said to be *directed*. In this thesis, I only consider *undirected* networks, meaning that all interactions are symmetric.

A network is said to be *multipartite* if it has several sets of nodes and links only connect pairs of nodes from different sets. These sets of nodes are sometimes called *levels* and multipartite networks can sometimes be found under the name of *multilevel* networks. In this thesis, only cases with a maximum of two different sets will be considered. These networks are referred to as *bipartite*. Sometimes, to distinguish the simple case from the multipartite case, we say that a network is *unipartite* if it is not multipartite. Figure 3 shows an example of tripartite network.

Graphs The natural mathematical object used to study networks is the *graph*. A graph is made up of two sets: a set of *vertices* and a set of *edges*. Obviously, the vertices of a graph represent the nodes of a network and the edges represent the links, so we can use these terms interchangeably.

Let G be a graph. We write $G = (V(G), E(G))$, where $V(G)$ is the set of the nodes of G and $E(G) \subseteq \mathcal{P}_2(V(G))$ the set of edges. Without loss of generality, we can label each node by a natural number so we assume $V(G) \subset \mathbb{N}$. We denote $v(G) = |V(G)|$ and $e(G) = |E(G)|$ the respective number of nodes and of edges of G. A *weighted* or *valued graph* taking values in B is a graph with an associated *weight* function $w : E(G) \to B$. This function associates each edge of G to its corresponding value. If $w : e \longmapsto w(e) = 1$, then w can be omitted and the graph is not

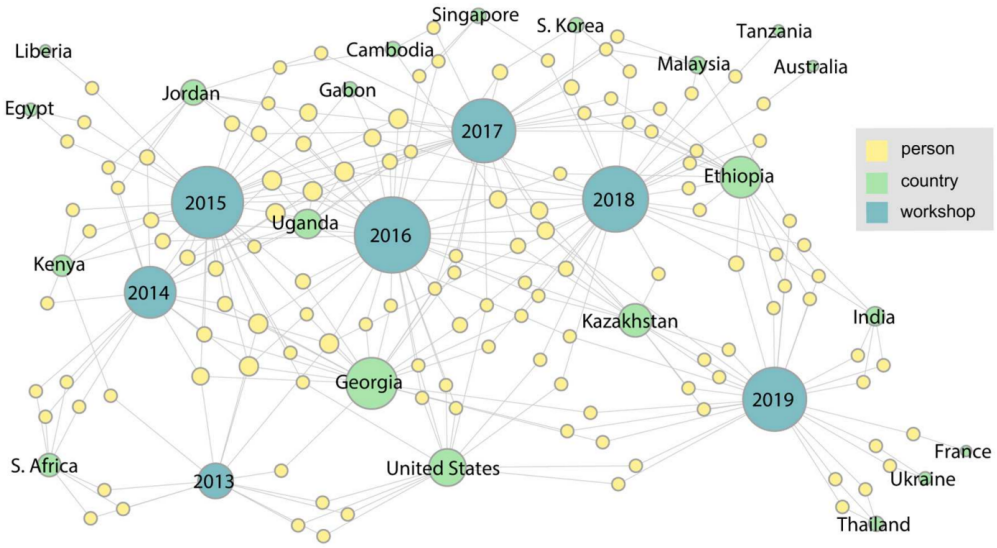

Figure 3 – A tripartite binary network representing researchers, their countries and attendance to international workshops (years). The size of the nodes indicate degrees, i.e. the sums of links stemming from the nodes. Figure taken from Ambrosiano et al. (2020).

valued. To distinguish the valued case and the non-valued case, we say that $G = (V(G), E(G))$ is a *binary graph* in the latter case.

A *subgraph* of G is a graph F such that $V(F) \subset V(G)$ and $E(F) \subset \mathcal{P}_2(V(F)) \cap E(G)$. We write $F \subset G$. A *path* in G is a sequence of nodes $i_1, ..., i_K$ such that for all $1 \leq k \leq K-1$, i_k and i_{k+1} are connected by an edge, i.e. $\{i_k, i_{k+1}\} \in E(G)$. A graph is *connected* if all pairs of nodes can be linked by a path. A *connected component* F of G is a maximal connected subgraph of G, i.e. a connected subgraph such that adding any other node of $G \backslash F$ and all its edges makes the new subgraph not connected.

A graph can also be used to represent a bipartite network. In a bipartite network, we can find a partition of $V(G)$ into two disjoint sets $V_1(G)$ and $V_2(G)$, i.e. $V_1(G) \cap V_2(G) = \varnothing$ and $V_1(G) \cup V_2(G) = V(G)$, such that $E(G) \cap (\mathcal{P}_2(V_1(G)) \cup \mathcal{P}_2(V_2(G)) = \varnothing$. However, the graph object is only moderately suitable for modelling bipartite networks because these constraints are complex and yet necessary. We therefore propose to define a simpler *bipartite* graph object with a simple constraint.

By placing an order between the two sets of nodes, we can define $G = (V_1(G), V_2(G), E(G))$ where $V_1(G)$ is the "first" set of nodes, $V_2(G)$ the "second" set of nodes and $E(G) \subset V_1(G) \times V_2(G)$ the set of edges. By analogy with the adjacency matrices that will be defined later, we also call $V_1(G)$ the *row nodes* and $V_2(G)$ the *column nodes*. This definition imposes an artificial direction on the edges, since each edge is a couple instead of a set of two nodes. However, the first element

of the couple is always part of the first set and the second is always part of the second set. This is just a benign consequence of the fact that we have decided that there is a "first" and a "second" set of nodes. This definition also allows the labels of the two sets of nodes $V_1(G)$ and $V_2(G)$ to overlap. This is not a problem because the roles of the two sets are already distinct by definition. The cardinals of the new sets of nodes can be denoted by $v_1(G) = |V_1(G)|$ and $v_2(G) = |V_2(G)|$.

Valued bipartite graphs can also be defined by associating a weight function in the same way as for unipartite graphs. A subgraph F of the bipartite graph G is a graph such that $V_1(F) \subset V_1(G)$, $V_2(F) \subset V_2(G)$ and $E(F) \subset (V_1(F) \times V_2(F)) \cap E(G)$. The notions of path and connected component remain the same. The only subtlety lies in the fact that a path is a sequence alternating between elements of $V_1(G)$ and elements of $V_2(G)$.

Useful operations are the union and intersection of graphs. We denote the union of two bipartite graphs G_1 and G_2 by $G_1 \cup G_2 = (V_1(G_1) \cup V_1(G_2), V_2(G_1) \cup V_2(G_2), E(G_1) \cup E(G_2))$. Analogously, the intersection of two bipartite graphs G_1 and G_2 is $G_1 \cap G_2 = (V_1(G_1) \cap V_1(G_2), V_2(G_1) \cap V_2(G_2), E(G_1) \cap E(G_2))$.

Adjacency matrices The other useful mathematical object for representing networks is the *adjacency matrix*. This representation will be extensively used in this thesis because it is particularly convenient. It provides a simpler representation of networks, particularly for defining random models on them. It also allows us to use linear algebra tools directly to carry out operations on networks. $\mathcal{M}_{m,n}(B)$ denotes the set of matrices of size $m \times n$ with values in B and $\mathcal{M}_n(B)$ the set of square matrices of size $n \times n$ with values in B.

Without loss of generality, let us label here the nodes of a unipartite graph G by the first integers $V(G) = [\![v(G)]\!]$. The adjacency matrix $A(G)$ of the graph G, with weight function w with values in B ($w \equiv 1$ if G is a binary graph) is the matrix of $\mathcal{M}_{v(G)}(B)$ defined by

$$A(G)_{ij} = \begin{cases} w(\{i,j\}), & \text{if } \{i,j\} \in E(G) \\ 0, & \text{otherwise,} \end{cases}$$

for $(i,j) \in V(G)^2$. The adjacency matrix of a unipartite (undirected) graph is therefore symmetrical. Each row and each column represents a node. The symmetry comes from the fact that the i-th row represents the same node as the i-th column. Figure 4 shows a unipartite graph and its adjacency matrix.

For a bipartite graph $G = (V_1(G), V_2(G), E(G))$, we assume without loss of generality that nodes are labelled by positive integers starting at 1, i.e. $V_1(G) = [\![v_1(G)]\!]$ and $V_2(G) = [\![v_2(G)]\!]$. G is associated with a weight function w with values in B ($w \equiv 1$ if G is a binary graph). We

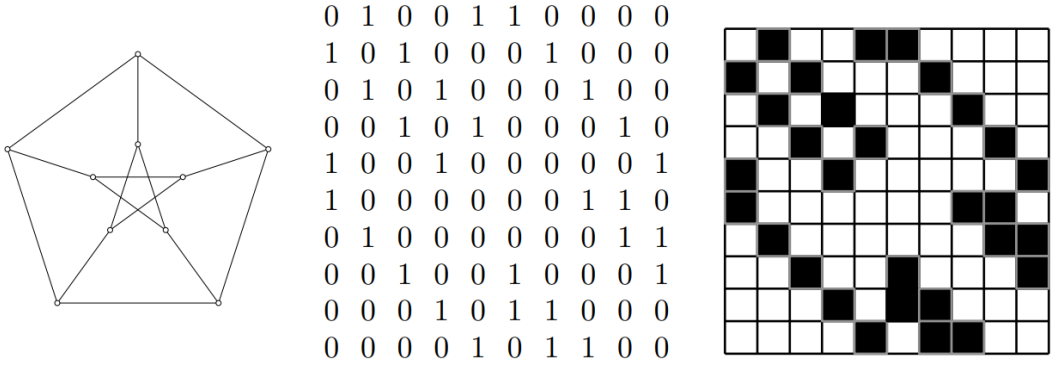

Figure 4 – A binary unipartite graph (left) and its adjacency matrix (middle). Equivalently, since the graph is binary, the adjacency matrix can be represented by a pixel map in $[0,1]^2$ (right), where $[0,1]^2$ is subdivided into $n \times n$ patches, and black patches represent the 1s of the matrix. Figure taken from Lovász (2012).

define the adjacency matrix $A(G) \in \mathcal{M}_{v_1(G),v_2(G)}(B)$ as follows:

$$A(G)_{ij} = \begin{cases} w(i,j), & \text{if } (i,j) \in E(G) \\ 0, & \text{otherwise.} \end{cases}$$

Note that the adjacency matrix of a bipartite graph is not symmetrical and it may be rectangular. This is because the rows represent the nodes of $V_1(G)$ (row nodes) and the columns represent the nodes of $V_2(G)$ (column nodes), i.e. different nodes with distinct roles.

Figure 5 illustrates the duality between networks and matrices. The original dataset consists of a list of countries and their performance indices for each of the 17 Sustainable Development Goals defined by the United Nations. This is more naturally represented by a matrix, but this matrix can also be considered as the adjacency matrix of a network, which enables the use of network science tools to analyze the data (Sciarra et al., 2021).

Random graph or network models A network is *random* if its corresponding graph is a random variable. This means that the graphs of observed networks are the result of a random experiment. A random network (or graph) *model* defines the distribution of this random variable. To define models, it is sometimes simpler to use adjacency matrices, i.e. to define the distribution of $Y = A(G)$ instead of first defining the distribution of G and then deducing the distribution of $A(G)$.

A random graph model can be defined by this distribution on the matrix Y. We often use the term *joint distribution* to distinguish between the distribution of the entire matrix and

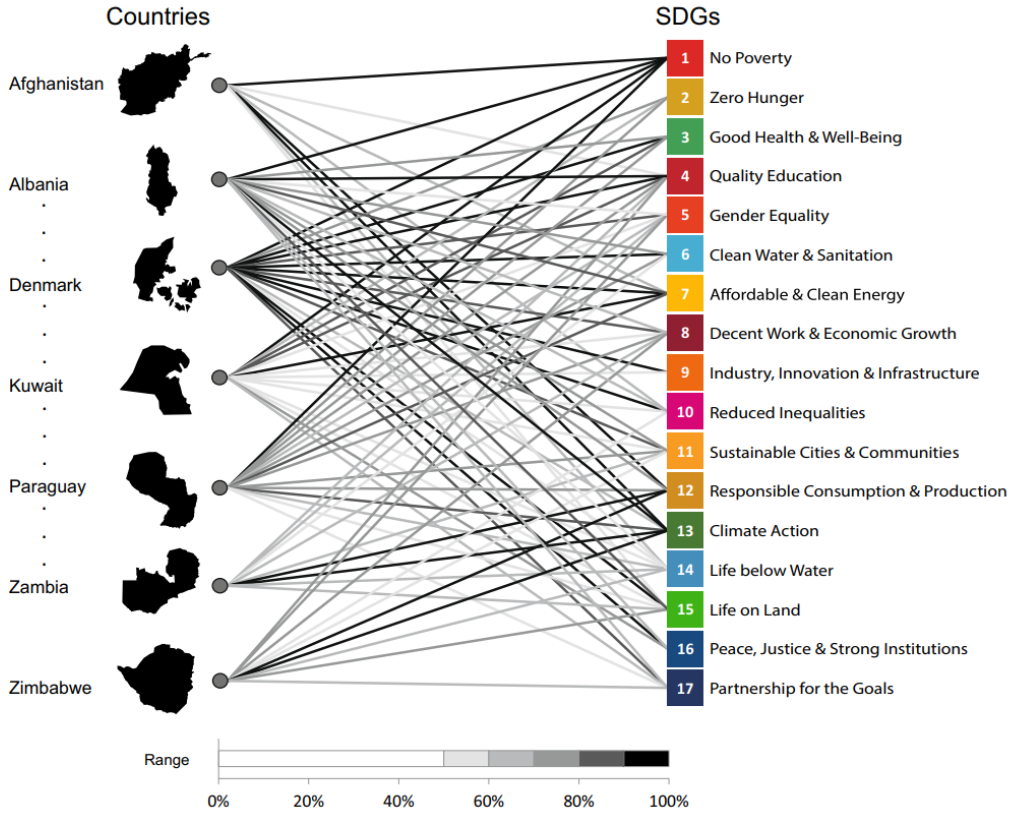

Figure 5 – A bipartite weighted network connecting seven countries and the 17 Sustainable Development Goals (SGDs) defined by the United Nations. The links are weighted with a performance index between 0 and 1 as defined by Lafortune et al. (2018) based on indicators defined by UN General Assembly (2017). The dataset is naturally represented by a matrix, but Sciarra et al. (2021) used network analysis methods to rank countries depending on their performance. Data for 2020 from Sachs et al. (2020), figure taken from Sciarra et al. (2021).

that of one entry Y_{ij} of the matrix. The joint distribution contains information about both the distribution of all Y_{ij} and their dependency structure. In the general case, it is not necessarily easy to find the distribution of a single Y_{ij}. This must be obtained by "marginalizing" $\mathbb{P}(Y)$, i.e. by integrating it with respect to all the other entries of the matrix.

Structure of the thesis

The thesis consists of an introductory chapter, three chapters grouping together various contributions, and a concluding chapter. The introductory chapter presents the problem this thesis is interested in studying, as well as the various contributions of this thesis. The main mathematical notions will also be introduced in this chapter.

The following three chapters are not organized by theme, but correspond to successive works in chronological order, some intended for publication in scientific articles. The first chapter lays the foundation stone for the proposed methodology. It corresponds to a theoretical article, published in ESAIM: Probability & Statistics. Then, as time went by, new theoretical results enabled us to considerably improve this methodology. These improvements will be presented in a second article, which has been submitted in a statistics journal. A chapter is dedicated to this article. Finally, the last contribution chapter was written in the month prior to submission of this thesis and presents my latest work, which establishes results that significantly complement the methodology developed in the second article. Because the final methodology is the culmination of all these developments, the most complete version of this methodology will be explained in this chapter. This could be the subject of a future article. These three chapters reflect my progress in solving the initial problem. As the first two are independent articles, some elements are redundant. This is not the case with the last chapter, which was written specifically for this thesis.

Finally, the concluding chapter of this thesis offers suggestions for further work. For the most part, these are avenues that I have begun to explore but not developed in more detail.

Chapter 1

Introduction

Contents

1.1. Networks and models

A network is a system consisting of multiple entities (nodes) and their pairwise relationship (links). Networks are omnipresent in real-life data and therefore they can be found in a large variety of contexts. However, they are complex objects and their analysis has been limited by the lack of mathematical theory, at least until the 19th century and the development of graph theory, combinatorics and probability theory. The exponentially increasing complexity of the interconnected world in the most recent century has led to the development of network theory to investigate complex systems. Despite the different systems they represent, networks share a common trait. Relationships between entities hold at least as much importance as the collection of entities. This is apparent as the natural mathematical object used to represent the networks is the graph, in which entities are represented by the vertices and relations by the edges of a graph. The key to most questions raised in network analysis is to understand regularities or rules in the pattern of links between nodes. With that in mind, network science has emerged as a field itself which benefits a multitude of apparently disparate science areas, including physics (especially statistical mechanics), biology, sociology, computer science, etc. (Barthélemy, 2011; Perna and Latty, 2014; Saavedra et al., 2014; Fortunato et al., 2018). Figure 1.1 shows two examples of trade networks, where nodes are countries and links are export activity. One might expect the two networks to show similarities, but the question is to what extent and, if there are significant differences, how to interpret them. Figure 1.2 represents a biological network, where nodes are proteins and links are their interactions. One interesting question is to understand the structure of this network, which gives information about the mechanisms determining the interactions.

One of the main objectives around network analysis is to link the observed networks to theory, i.e. to understand the processes producing the networks and their variability. Network models can be used to integrate these processes and how they affect networks. The aim of random network models is to add controlled sources of heterogeneity and variability. Before presenting any analysis strategy involving models, I first give a few examples of historical random network models taken from classical network theory. These models have been used to analyze networks in many areas of science.

Network models can be classified into two families that I call procedure-based models and

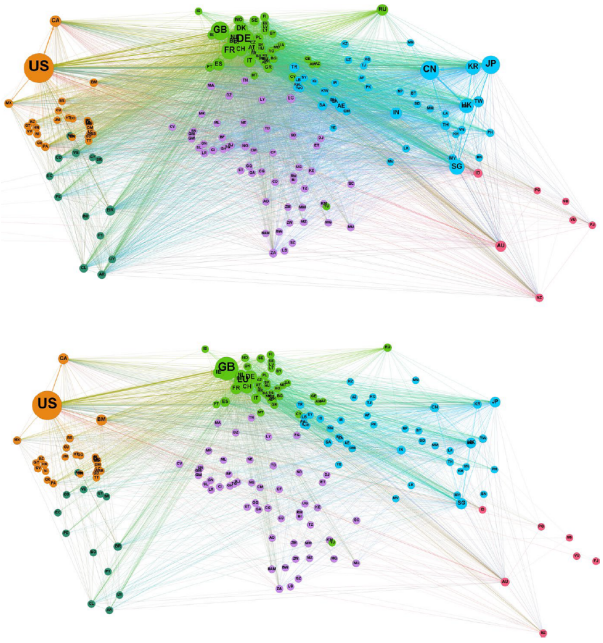

Figure 1.1 – World export activity networks for transportation (top) and financial (bottom) services from the OECD-WTO Balanced Trade in Services database (Fortanier et al., 2017). Figure taken from Tajoli et al. (2021).

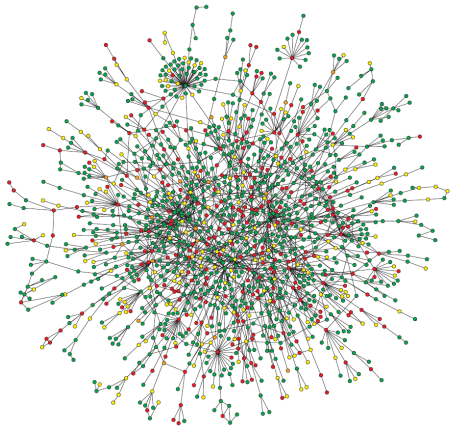

Figure 1.2 – Protein interaction network in the yeast *Saccharomyces cerevisiae*, colours indicates the effects of removing the corresponding protein in the yeast (red = lethal, green = non-lethal, orange= slow growth, yellow=unknown). Figure taken from Barabasi and Oltvai (2004).

likelihood-based models. The distinction between these two families lies in the definition of the model. A procedure-based model is defined by an initial network which is iteratively modified by some algorithm, possibly stochastic. A likelihood-based model is directly defined by a probability distribution on the network adjacency matrix. The likelihood of a network is the probability that the model generates that network. Obviously, a procedure-based model yields a distribution on

the network adjacency matrix and a likelihood-based model can be simulated with an algorithm. However, neither the probability distribution of a network generated by a procedure-based model nor the algorithm to simulate a network according to a likelihood-based model are necessarily straightforward. When both are actually possible, the model can be considered as both likelihood and procedure-based.

1.1.1. The Erdős-Rényi model and some variations

The following models are introduced as they originally appear in the literature. Therefore, most are unipartite binary graph models. However, most can be transposed into a bipartite or weighted setup without difficulty.

The Erdős-Rényi (ER) model is often considered as the simplest random graph model. In reality, two related but distinct models are referred to as the ER model. Erdős and Rényi (1959) introduced the $G_{n,M}$ model with two parameters n, the number of nodes, and M, the number of edges. Then, the M edges are distributed to connect M pairs of nodes. The other model, denoted $G_{n,p}$ has been introduced by Gilbert (1959). It has two parameters n the number of nodes and p a probability. The random graph is generated by connecting each pair of nodes independently with probability p (Fig. 1.3). When $n \to \infty$ and $M = np$ where p is fixed, both models are asymptotically equivalent.

However, for some finite n, they cannot be confused with each other. Whereas the $G_{n,p}$ model is both procedure and likelihood-based, the $G_{n,M}$ model is only procedure-based. In the $G_{n,p}$ model, all edges are i.i.d. with a Bernoulli distribution by definition. In the $G_{n,M}$, the final graph has been uniformly sampled from the set of graphs with n nodes and M edges. The joint distribution of its adjacency matrix can be calculted, but it is complex due to the dependency induced by the constraint of fixed number of edges. This property makes the $G_{n,p}$ model easier to study and therefore more widely used. In this thesis, unless explicitly specified, the ER model will refer to the $G_{n,p}$ model.

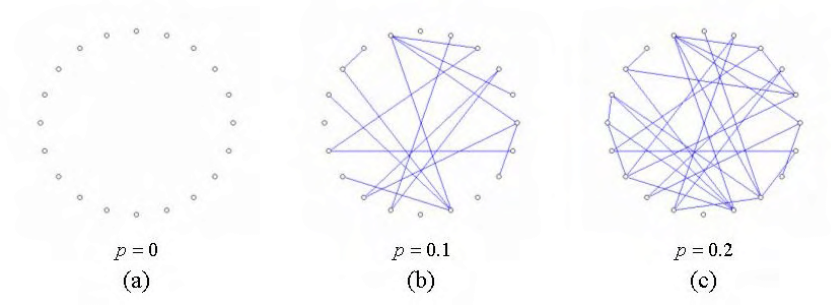

Figure 1.3 – $G_{n,p}$ graphs for fixed n, different values of p. Figure taken from Newth (2006).

The graphs generated by these models are particularly homogeneous. As a result, they do not describe a large number of empirical networks very well. Real-life networks commonly exhibit two properties.

First, Barabási and Albert (1999) have highlighted the property of scale-free invariance. Scale-free invariance refers to degree distributions, which in the case of real networks, often follow a power law. This property is related to the notions of growth and preferential attachment (de Solla Price, 1976): when a new node is added to an existing network, it is more likely to connect to nodes that are already strongly connected (high degree nodes). Thus, edges cannot be drawn uniformly at random and degrees are inhomogeneous.

Second, the small-world property states that the length of the path between two randomly selected nodes is small (Watts and Strogatz, 1998), i.e. it has high clustering coefficient. This property was first identified in sociology, where it was shown that in the studied populations, only a short chain of relationships separates on average two individuals who do not know each other (Milgram, 1967). Indeed, it is often observed that the expected length of this chain only increases logarithmically with the size of the population. Scale-free networks also exhibit the small-world property (Cohen and Havlin, 2003). This property is notably observed in plant-pollinator networks, where this growth is sometimes even slower (Olesen et al., 2006). Nevertheless, the sample size required to verify such a property is rarely reached in ecology.

Networks with the small-world property, and therefore the scale-free property, often feature "hubs" (Fig. 1.4). Hubs are highly connected nodes, which in consequence are some kind of central nodes. They play significant roles in both the scale-free property and the small-world property. They are partially responsible for the asymmetry in the degree distribution, leading to the typical power law distribution of scale-free networks. Hubs also accounts for the small-world property when the network is said to be assortative, i.e. highly connected nodes tend to connect with each other. In this case, the paths that connect any pair of individuals only need to pass through a few hubs. Hubs do not appear in ER graphs.

Variations of the ER model have been suggested to incorporate these properties featured in real networks. The Watts-Strogatz model (WS) (Watts and Strogatz, 1998) and the Barabási-Albert model (BA) (Barabási and Albert, 1999) are two among the most famous ones. These two models are procedure-based. The WS starts with a ring lattice and use random rewiring to generate a graph with the small-world property. The BA starts with a small fully connected graph and nodes are iteratively added. Each newly added node connects to each existing ones with a higher probability if the degree of the existing node is higher. This yields a network with the scale-free property. These two models are examples of models designed to mimic a few selected macroscopic properties of real-life networks.

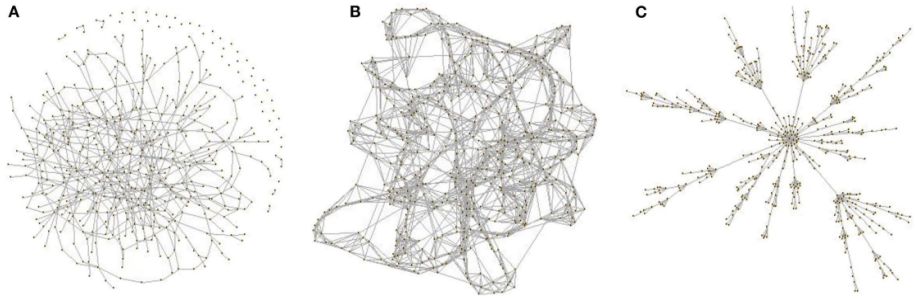

Figure 1.4 – Left: a $G_{n,p}$ network. Middle: a Watts-Strogatz (WS) network. Right: a Barabási-Albert (BA) network. Although the WS and BA generate networks with different structures, they both lead to the formation of hubs (highly connected nodes). Figure taken from Koutrouli et al. (2020).

1.1.2. Configuration models

The configuration model (CM) has origins in ecology and null models, where Connor and Simberloff (1979) used such a model to study island species distribution with presence-absence data, binary matrices describing which species is present on which island. These matrices can be viewed as adjacency matrices of bipartite networks. The CM aims to generate random networks with particular fixed degree sequences, that are also called marginals. It is related to the $G_{n,M}$ model as it comes to uniformly pick a network from the set of networks respecting a set of constraints (the number of edges in the $G_{n,M}$ model, with the addition of the marginals in the CM). A related mathematical question has been investigated by Bender and Canfield (1978) and Bollobás (1980), on models of random matrices with fixed marginals.

Simulating such networks had been a complex issue for a long time, although it has been partially solved. Many algorithms have been proposed. Gotelli and Entsminger (2001, 2003) described several types of "fill" and "swap" algorithms (Fig. 1.5). Fill algorithms build a network by sequentially drawing edges between the nodes. Because the network has to satisfy the marginal constraints, these algorithms have to backtrack when no more edge can be added without breaking the constraints, which explains their computational cost. To solve this problem, swap algorithms start with an observed network and at each iteration, they swap a couple of edges. After any number of iterations, the modified network still satisfies the constraints. However, the distribution of the sequence of networks obtained at each iteration is a Markov chain, so the distribution of the network obtained even after a large number of iterations may not be uniform, despite achieving stationarity. Miklós and Podani (2004) proved that the trial-swap algorithm, a slight modification of a swap algorithm, is unbiased. However, its complexity makes it difficult to apply in practice. The current state-of-the-art method has been found by Strona

et al. (2014), who designed an algorithm to generate networks with unbiased distribution in fewer iterations than simpler swap algorithms. This algorithm is relatively fast for networks of moderate size ($10^4 \times 10^4$ nodes). Therefore, the simulation issue is circumvented when at least one network reproducing the constraints, needed to initialize the algorithm, is already known. This requirement is fulfilled in null model analysis, a common analysis approach in ecology that will be presented later.

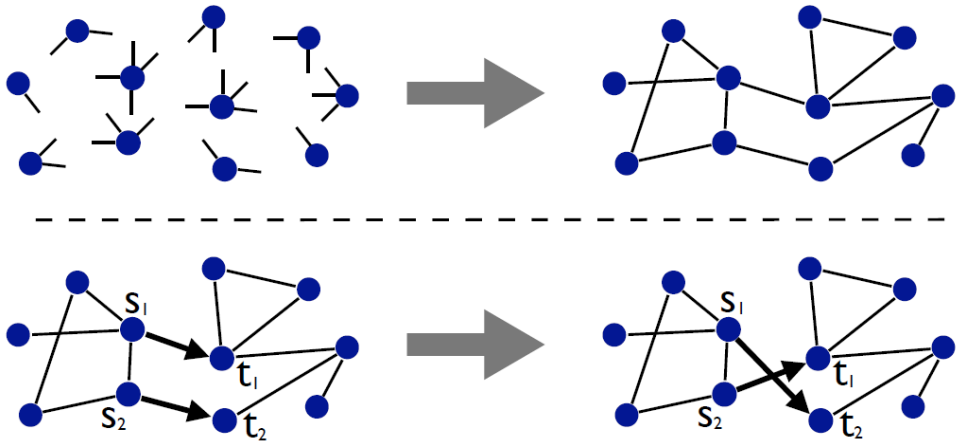

Figure 1.5 – Two procedures generating networks from a configuration model. Top: a fill algorithm starts with nodes and stubs corresponding to a given degree sequence, ends with all the stubs connected. Bottom: a swap algorithm starts with a network, at each iteration, two edges are selected and are swapped, e.g. (s_1, t_1) and (s_2, t_2) are swapped to (s_1, t_2) and (s_2, t_1). Figure taken from Zamora-López (2009).

The marginal constraints have been softened in several variations of the CM. Instead of using specific degree sequences, Newman et al. (2002) used distributions of degrees. Each node draws independently its degree from a multinomial distribution, so that the sequence of degrees is a random sample. Then, networks can be sampled using the previously listed algorithms. Chung and Lu (2002) added another type of variability by using "expected" degree sequences. Its bipartite version can be formulated as follows. The two sets of nodes of the bipartite network have respective sequences of expected degrees $(d_i)_{1\leq i\leq m}$ and $(e_i)_{1\leq j\leq n}$ such that $\sum_{i=1}^{m} d_i = \sum_{j=1}^{n} e_j = L$ and $L \geq \max_{i,j}(d_i, e_j)$, the node i of the first set and j of the second set are connected with probability $p_{ij} = \frac{d_i e_j}{L}$. Therefore, the Chung-Lu model allows fluctuations of the degrees around the specified expected degrees. Combining both sources of variability, the expected degree distribution model (EDD) (Picard et al., 2008) lets each node draw its expected degree in a distribution. In contrast to the degree sequence models, the expected degree sequence models (the Chung-Lu model and the EDD) are both procedure-based and likelihood-based, so generating networks is simple and expressing the joint distribution of the network is easier.

1.1.3. Exponential random graph models

Holland and Leinhardt (1981) suggested the p_1 model, a likelihood-based model for unipartite directed networks. The adjacency matrices of these networks are square, but potentially asymmetric. For a pair of nodes (i, j), $Y_{ij} = 1$ if there is a directed connection from i to j, $Y_{ij} = 0$ otherwise. The pairs of nodes are independent and there are thus 4 possibilities for $A_{ij} = (Y_{ij}, Y_{ji})$, to which the following probabilities are assigned:

$$\begin{aligned}
\log \mathbb{P}(A_{ij} = (0,0)) &= \lambda_{ij} \\
\log \mathbb{P}(A_{ij} = (1,0)) &= \lambda_{ij} + \alpha_i + \beta_j + \gamma \\
\log \mathbb{P}(A_{ij} = (0,1)) &= \lambda_{ij} + \alpha_j + \beta_i + \gamma \\
\log \mathbb{P}(A_{ij} = (1,1)) &= \lambda_{ij} + \alpha_i + \beta_j + \alpha_j + \beta_i + 2\gamma + \rho,
\end{aligned}$$

hence, for an observed pair of nodes

$$\log \mathbb{P}(a_{ij}) = \lambda_{ij} + y_{ij}(\alpha_i + \beta_j + \gamma) + y_{ji}(\alpha_j + \beta_i + \gamma) + y_{ij}y_{ji}\rho.$$

The coefficients represent several effects that are well known in sociology. α_i is the propensity of node i to be the starting point of a connection (expansiveness), and conversely β_i is its propensity to receive one (attractiveness). γ is a diffusion effect that controls the density of interactions and ρ is the propensity of reciprocity, which makes that for a pair of individuals, interactions in one direction and in the other are not independent. Finally, λ_{ij} is a normalization coefficient so that these probabilities sum to 1.

Their model can be transposed to a bipartite undirected setup, with rectangular adjacency matrices Y of size $m \times n$, where m and n are the number of nodes of the two sets. Let γ, $(\alpha_i)_{1\leq i\leq m}$, $(\beta_j)_{1\leq j\leq n}$ and $(\lambda_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ such that

$$\begin{aligned}
\log \mathbb{P}(Y_{ij} = 0) &= \lambda_{ij} \\
\log \mathbb{P}(Y_{ij} = 1) &= \lambda_{ij} + \alpha_i + \beta_j + \gamma.
\end{aligned}$$

hence, for an observed pair of nodes

$$\log \mathbb{P}(y_{ij}) = \lambda_{ij} + y_{ij}(\alpha_i + \beta_j + \gamma).$$

Since all the pairs of nodes are independent, the likelihood of an observed adjacency matrix y is

$$\begin{aligned}
\mathbb{P}(y) &= \prod_{i,j} \mathbb{P}(y_{ij}) \\
&= \exp\left(\sum_i y_{i+}\alpha_i + \sum_i y_{+i}\beta_i + y_{++}\gamma\right) \times \exp\left(\sum_{i,j} \lambda_{ij}\right),
\end{aligned}$$

where y_{i+} is the degree of the i-th node of the first set, y_{+j} is the degree of the j-th node of the second set and y_{++} is the number of edges in the network. These terms are easily obtainable network descriptors. Group them in a "score" vector $x(y) = (y_{1+}, ..., y_{m+}, y_{+1}, ..., y_{+n}, y_{++})$ and group the parameters $\theta = (\alpha_1, ..., \alpha_m, \beta_1, ..., \beta_n, \gamma)$. The exponential term depending on the $(\lambda_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ is a multiplicative normalizing constant depending on θ. Therefore, the likelihood of an observed network given by (1.1) can be written as the likelihood of an exponential family distribution

$$\mathbb{P}(y) = \frac{\exp(\theta^T x(y))}{Z(\theta)}, \tag{1.1}$$

where $Z(\theta) = \exp\left(-\sum_{i,j} \lambda_{ij}\right)$.

Observe that the joint probability distribution of the network falls under the exponential family where the sufficient statistics are the network descriptors contained in the score vector $x(y)$. Holland and Leinhardt (1981) suggested to generalize the model by incorporating other network statistics into the score vector, including the count of particular subgraphs that seem relevant for characterizing networks. However, the p_1 model makes the strong assumption that the pairs of nodes are independent. This is necessary to obtain this probability distribution, but this is often too crude an approximation. To relax this assumption, Frank and Strauss (1986) defined a dependence structure between the nodes of the network and their neighbors, called a Markov graph: two pairs of nodes are now dependent if they share a node. This model defines the class of exponential random graph models (ERGMs) or p^* models, for which it is possible to embed other types of variables such as global metrics in the score vector $x(y)$. The strength of ERGMs is to be able to predict an interaction on the basis of what is happening around it: the expression of the likelihood of a graph is the same as previously defined by the model p_1 and the conditional probability of one edge knowing the rest of the network can be deduced from it:

$$\begin{aligned} p(y_{ij} = 1 \mid y^{-ij}) &= \frac{\mathbb{P}(y^+)}{\mathbb{P}(y^+) + \mathbb{P}(y^-)} \\ &= \frac{\exp\left(\theta^T (x(y^+) - x(y^-))\right)}{1 + \exp\left(\theta^T (x(y^+) - x(y^-))\right)} \\ &= \text{logit}^{-1}\left(\theta^T (x(y^+) - x(y^-))\right), \end{aligned}$$

where Y^{-ij} is the whole matrix Y excluding Y_{ij}, Y^+ (respectively Y^-) is the whole matrix Y with $Y_{ij} = 1$ (respectively $Y_{ij} = 0$).

Despite the flexibility of ERGMs, their use is not generalized to all domains besides sociology. The lack of a suitable estimation procedure is one of the major problems encountered in the application of ERGMs, this will be explained later. Another inherent caveat of ERGMs is that they do not represent well the local interactions between small subsets of nodes. Indeed, the descriptors involved, such as network metrics and subgraph frequencies, summarize the global

structure of the networks well but struggle to capture local information. This also has the consequence that the model is often degenerate and unstable (Handcock, 2003; Chatterjee and Diaconis, 2013). Because of this, on one hand, ERGMs are mostly suited to represent large networks but on the other hand, their computation cost is too large for these networks.

1.1.4. Latent space models

In many fields, but particularly in social network analysis, it has been common to use latent structures such as latent groups to characterize a population structure, where individuals within a group play an equivalent role in the network. For instance, groups could be made of people with the same social position (White et al., 1976; Faust and Skvoretz, 2002). The same concepts have been investigated in ecology, where these groups of species with similar roles in an ecosystem have also been called functions or guilds (Luczkovich et al., 2003; Baskerville et al., 2011).

It is reasonable to use this concept to study networks. Formalizing it mathematically, the nodes are partitioned into K groups (for a bipartite network, K_1 row groups and K_2 column groups). The set of edges between the nodes of group q and the nodes of group ℓ defines a block (q, ℓ). Thus, Wang and Wong (1987) extended the previously mentioned p_1 model so that the edges which belong to the same block share common parameters. Nowicki and Snijders (2001) assumed that the groups are unknown and they used latent variables to encode the memberships of the nodes in the groups. Suppose that there are n individuals in a unipartite network and $(Z_i)_{1\leq i\leq n}$ are these latent variables, e.g. $Z_k = q$ means the k-th individual belongs to the q-th group. Then the interaction between individuals i and j follows a distribution $\mathcal{L}$, the parameters of which only depend on the groups to which they belong:

$$Y_{ij} \mid Z_i = q, Z_j = \ell \sim \mathcal{F}(\alpha_{q\ell}). \tag{1.2}$$

This model is called the stochastic block model (SBM) which is more adapted to the analysis of network data, because the groups are not observed most of the time. This model induces a relatively simple latent structure, since all nodes are equivalent within the same group, which constitutes a mixture of simple models. Typically, for binary networks, $\mathcal{F}$ is a Bernoulli distribution and the SBM is a mixture of ERs. Figure 1.6 shows a panel of network structures that can be described by the discrete latent space defined by SBMs.

The SBM can be considered as a special case of a large class of random graph models called the latent space models (Goldenberg et al., 2010; Matias and Robin, 2014; Jacobs and Clauset, 2014). In these models, the underlying process producing the structure of the graph is determined by a hidden node structure. In the SBM, this hidden structure is the membership to a group ($Z_i \in \{1, ..., K\}$). Other models define more complex latent spaces, allowing for more

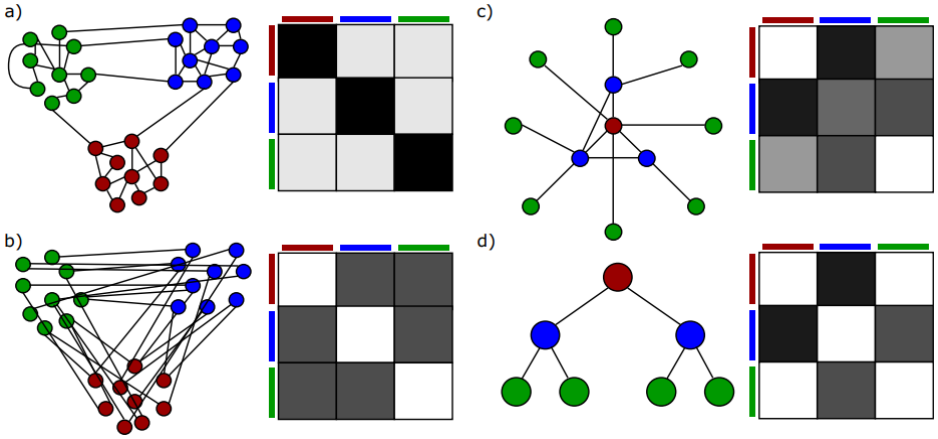

Figure 1.6 – Different network stuctures explained by binary SBMs with $K = 3$ and their probability matrix $\boldsymbol{\alpha}$. The color of a node indicates its membership of a group. However, these groups are generally not known before performing inference on the model. a) a modular structure, where groups correspond to modules, b) a tripartite structure, where nodes from each group only interact with nodes from other groups, c) a core-periphery structure, where groups correspond to successive layers, d) a hierarchical tree structure, where each group corresponds to a level in the tree. Figure taken from Funke and Becker (2019).

flexibility. Notably, Airoldi et al. (2008) extended the SBM to the case where each individual belongs to a mixture of groups, so the interactions are given by a block average, weighted by the proportions of this mixture.

The latent space can also be continuous, implying that, conversely to the SBM with a discrete space, all individuals are potentially different. In latent position models, the notions of groups and equivalence of SBMs has been abandoned in favour of a mapping of the nodes into a continuous space, for example $Z_i \in \mathbb{R}^K$. The interactions are then determined by these hidden node positions, for example by the euclidian distances between them (Hoff et al., 2002) or their dot product (Young and Scheinerman, 2007). An hybrid approach by Handcock et al. (2007) adds elements of SBM to latent position models, determining the positions of individuals by a mixture of normal distribution symbolizing their membership in different groups. Finally, to refine the differences between groups or positions, some models use a discrete hierarchical latent space, like a tree. The interactions are then also characterized by the positions of the individuals in the hierarchy (Clauset et al., 2008). The hierarchical latent space may also be continuous, like a hyperbola (Asta and Shalizi, 2015).

More generally, the adjacency matrix has values in a set E (e.g. $E = \{0, 1\}$ for binary graphs, $E = \mathbb{R}$ for real-valued graphs, etc.) and the latent variable belong to some latent space F. The

general form of these models is

$$Y_{ij} \mid Z_i, Z_j \sim \mathcal{L}(Z_i, Z_j), \forall 1 \leq i, j \leq n,$$

where n is the number of nodes of the network, $\mathcal{F}$ is a probability distribution on F and for all $z_i, z_j \in F$, $\mathcal{L}(z_i, z_j)$ is some probability distribution on E. Conditionally to the latent variables, each Y_{ij} has a distribution only depending on those of the corresponding nodes Z_i and Z_j. The latent variables $(Z_i)_{1 \leq i \leq n}$ can be either constant or drawn from some distribution, for example $Z_i \overset{i.i.d.}{\sim} \mathcal{F}, \forall 1 \leq i \leq n$.

One advantage of these models is that they are generative models for random graphs. These latent spaces are used to model the mechanisms producing the network, since they encode the structure of the network. In all these models, the distribution of individuals in the latent space is often of as much interest as the interactions themselves (Lubold et al., 2023). One of the key goals of using these models is therefore to be able to reveal the underlying structure that generates the network. The model fixes the constraints on this structure: the simpler it is, the more amenable it is to interpretation. In this respect, the SBM is a simple latent space model. The information contained in the SBM makes it possible to evaluate the number and the size of the groups, the memberships of the individuals in the different groups and the role and interactions between the different groups. This information therefore goes beyond the local scale and also deals with the organization of the entire population under study. For instance, McCormick and Zheng (2015) used the latent space estimated by their model to make inferences about the unobserved social position of certain individuals.

It should be noted that these models can be combined with ERGMs. In fact, the score vector of ERGMs can incorporate a space characterised by latent variables (Schmidt and Mørup, 2013). It is the existence of this latent structure that gives them the above-mentioned advantages of latent space models, although the computational burden of ERGMs has yet to be solved.

1.1.5. Estimation for likelihood-based models

The diversity and flexibility of the models presented above enable them to answer many network analysis questions. Network models can be used in several ways to explain network structure. One way for them is to attempt to reproduce the properties of observed networks. This is the case with some models, such as the BA and the WS, which have been designed to reproduce specific properties of real networks, namely the scale-free property and the small-world property. For example, in the BA, the scale-free property is generated by a preferential attachment process. Preferential attachment can be expected to occur in real-life networks if they have properties similar to those of the networks simulated by the BA. However, there

is no guarantee that these properties can be explained by these processes alone. The power law degree distributions observed in real networks often lack statistical significance (Stumpf and Porter, 2012). Many models have been shown to generate scale-free networks, but without implementing the preferential attachment process. Typically, the CM specifies a degree sequence for the generated networks, so it can simulate networks with a power-law degree distribution, i.e. with the scale-free property (Massol et al., 2017).

In contrast, in likelihood-based models, the probability distribution of networks is defined. The processes generating the networks can be incorporated and set by parameters, which lends flexibility to the models. By estimating parameters using observed networks, we hope to characterize and understand network structure. To do this, we generally study the likelihood function. If the parameters are unknown, the likelihood function of a model is often denoted by $f(y;\theta) := \mathbb{P}(Y = y;\theta)$ where Y is the random adjacency matrix of the network, y its realization and θ is the vector of parameters of the model. In this case, the likelihood of the observed networks is a function of these parameters. Given one or a set of observed networks, we can estimate θ by maximizing the likelihood function, for example. However, this optimization problem is not straightforward in general, because in many likelihood-based network models, the likelihood function is well identified, but sometimes untractable. This is the case with ERGMs and many latent space models.

The difficulty in using ERGMs lies in the normalization constant $Z(\theta)$ in equation (1.1), which is difficult to calculate except for small graphs. Since node pairs are no longer independent as in the p_1 model, all possible networks would have to be explored. The presence of this constant makes parameter estimation by maximum likelihood difficult. To get around this, Strauss and Ikeda (1990) maximizes a pseudo-likelihood function based on the conditional probabilities of the network edges, which are easier to estimate. However, this pseudo-likelihood approach leads to results that are generally different from the maximum likelihood estimator. Although the resulting estimator is consistent and asymptotically normal for a fixed number of nodes (Arnold and Strauss, 1991), Geyer and Thompson (1992) have shown that it overestimates the dependency in the graph structure and does not give reliable results in practice. For these reasons, Hunter et al. (2008) advocated approaching maximum likelihood by Monte Carlo Markov chain methods (MCMC). However, these methods are slow to converge (Bhamidi et al., 2011) and therefore sometimes inapplicable.

Estimation procedures for latent space models are also complex. These models are defined using a sequence of latent variables $Z = (Z_i)_{1\leq i\leq n}$. Usually, the model defines the complete likelihood $f(y,z;\theta) := \mathbb{P}(Y = y, Z = z;\theta)$. The observation likelihood function is obtained by intergration $f(y;\theta) = \int f(y,z;\theta)dz$. However, depending on the latent structure, the likelihood function is often untractable and it is impossible to use simple convex optimization tools to maximize

it. Sometimes, Bayesian methods or likelihood function approximation methods are required for estimation. For example, estimation of parameters in the SBM is done through MCMC (Snijders and Nowicki, 1997; Nowicki and Snijders, 2001) or variational inference (Daudin et al., 2008; Latouche et al., 2012).

Another tricky problem with likelihood-based models is model selection, i.e. the question of choosing the form of the model. This concerns, for example, the descriptors to be incorporated in the score vector of ERGMs, or the shape of the latent space in latent space models when it is unknown. Selection criteria are also needed to choose the number of groups or dimensions in the latent space. The link between the complexity of these models and their number of parameters is non-trivial. It is therefore impossible to use conventional criteria (e.g. the AIC/Akaike Information Criterion, the BIC/Bayesian Information Criterion) directly. Other selection criteria need to be adapted, e.g. the integrated complete likelihood (ICL) criterion, initially devised for simple mixture models (Biernacki et al., 2000), used for selecting the number of groups in SBMs (Côme and Latouche, 2015).

1.2. Interaction networks in ecology

1.2.1. Ecological interactions

My work has been motivated by the study of ecological interactions between species. Understanding these interactions allows to comprehend the role of species, how they affect, or depend, on each other. In other words, the organization of an ecosystem and the processes governing its functioning are reflected by the interactions between species (Allee et al., 1949; Paine, 1966). For this reason, ecologists have been collecting interaction data for a long time (Clements and Long, 1923; Robertson, 1928). Many ecosystems and many types of biotic interactions can be represented by networks. Although interaction networks are the most common type of network in ecology, other types of networks are also considered by ecologists, including spatial network and animal social networks (Sundaresan et al., 2007; Sueur et al., 2019). Even though data from ecological surveys, i.e. presence-absence or abundancy data (Bell, 2003), are usually not studied through networks, they can be seen as graph adjacency matrices so many network approaches can apply (Thébault, 2013; Vilhena and Antonelli, 2015). By essence, all these ecological networks differ in the nature of the represented entities and relationships. Spatial networks connect locations. Animal social networks involve individuals, not species. Presence-absence data associate species to observation sites. It is to be noted that network theory can be used to study all these types of networks, for example, block models can be useful for spatial networks, as well as interaction networks. However, these data raise different ecological questions and for

that reason, the analyses that are used could ultimately vary and the interpretations are not the same. My focus is placed on interactions networks, although I will sometimes mention the other types.

Among interaction networks are many different types of interaction that can be represented. The two main types of interaction are antagonistic and mutualistic. An antagonistic interaction is characterized by one species taking advantage of another, the former causing harm to the latter. Food webs (Dunne et al., 2002a; Dobson, 2009), herbivory networks (Villa-Galaviz et al., 2012; Welti et al., 2017) and host-parasite networks (Sugiura, 2007; Hadfield et al., 2014) are the main examples. In a mutualistic interaction, the two species involved cooperate, each benefiting from the other. These include pollination (Ramírez et al., 2011; de Manincor et al., 2020a), where insects feed on nectar and help dispatch pollen, seed dispersal (Carlo et al., 2003; Schleuning et al., 2011; Heleno et al., 2013), where animals, e.g. birds, spread seeds after eating fruits, ant-plant interaction (Blüthgen et al., 2004; Passmore et al., 2012), where plants provide food and nesting space to ants who protect them from herbivores. Figure 1.7 shows an example of food web and an example of pollination network.

Given the diversity of ecosystems that can be represented by interaction networks, there is great hope that many different biological questions can be studied in the light of network analysis. In particular, my work has been motivated by the study of insect-mediated pollination networks. In these networks, interactions only occur between two types of species: pollinating insects and plants. An interaction between an insect and a plant means that the insect has been observed visiting the plant or, equivalently, that the plant has been observed being visited by the insect. Because of this constraint, pollination networks are bipartite. As mutualistic interactions generally involve two species interacting, but each benefiting in a different way, most mutualistic networks are also made up of two distinct species types and can therefore be represented in this way. The same applies to certain antagonistic networks, such as herbivory or host-parasite networks. These differ from food webs, which are unipartite networks, in which all species can a priori prey on each other. The remainder of this section provides an overview of ecological network analysis in the literature. I will put the focus on mutualistic networks, but many analyses can be freely transposed between different types of networks, mutualistic or antagonistic, unipartite or bipartite, so studies on other types of networks are also mentioned where relevant.

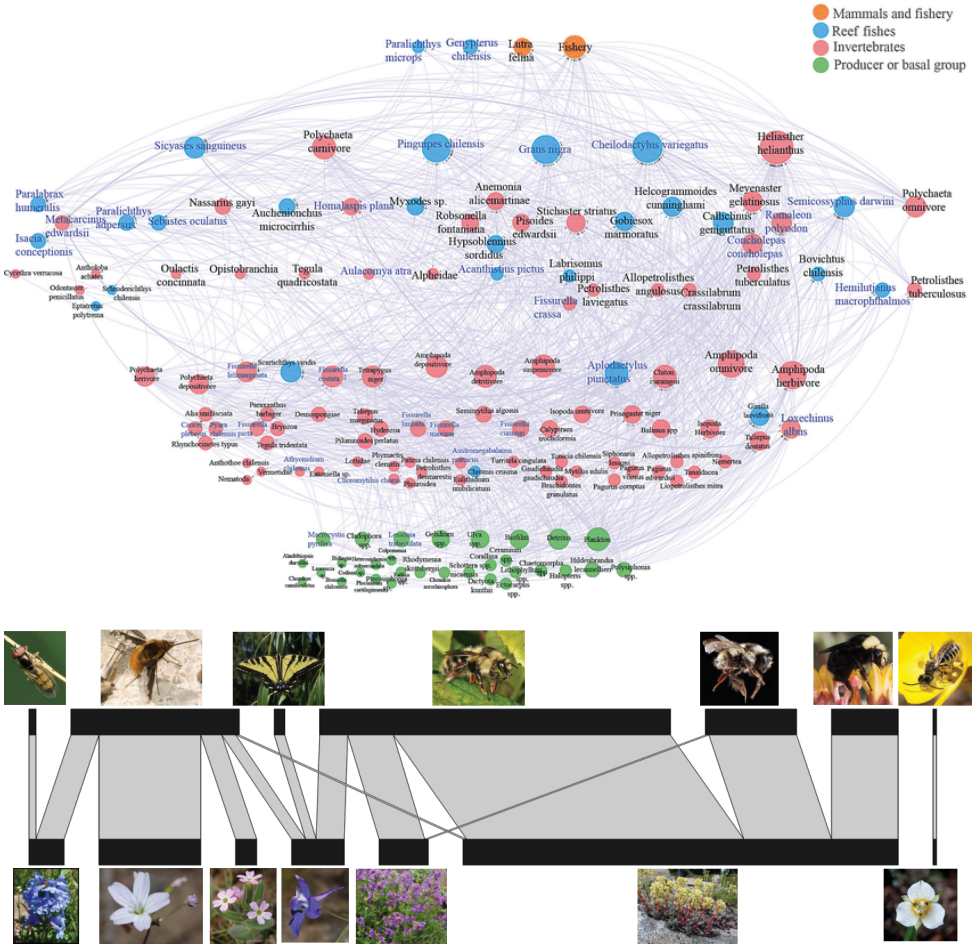

Figure 1.7 – A unipartite ecological network and a bipartite ecological network. Top: a food web from a Chilean coastal ecosystem, the vertical position of each node corresponds to its trophic level, which indicates the position of the species in the food chain. Figure taken from Pérez-Matus et al. (2017). Bottom: a pollination network between seven insect species and seven plant species, the width of the links indicates the frequency of their interaction. Figure taken from Seo and Hutchinson (2018).

1.2.2. Network structure

The answers to many ecological questions on the organization of ecosystems lie in the structure of networks. Much work has been carried out to study the determining factors of the structure of interaction networks. A classic approach is to investigate how species characteristics and preferences influence the network (Carlo et al., 2003; Blüthgen et al., 2004; Bartomeus, 2013; Eklöf et al., 2013; Van Kleunen et al., 2023). Another recurring concern about networks is about the missing links: does the absence of edge in the observed network means that two species actually never interact due to ecological reasons or is it a simple consequence of in-

complete sampling? Using prior knowledge on the species, one can mitigate the sampling bias by incorporating forbidden links, where two species cannot interact due to known constraints (incompatible phenology or traits) (Stang et al., 2006; Olesen et al., 2011; Novak et al., 2011). Jordano (2016) and Dormann et al. (2017) review the principal effects of sampling on interaction networks. However, despite the imperfection of the data, many networks share common properties that can be ecologically interpreted regardless of the involved species, see for example Jordano et al. (2003); Blüthgen et al. (2006) and Bascompte and Jordano (2013). Many descriptive metrics claim to measure these properties. Connectance is the proportion of the possible interactions in the network that actually occur (Jordano, 1987; Dunne et al., 2002b). In other science areas, connectance is also known as the density of the network. It is one of the simplest, yet one of the most used metrics. Two other widely used metrics are modularity and nestedness.

Networks can sometimes be separated into compartments (Guimera and Nunes Amaral, 2005), which indicates in ecological networks that several communities coexist (Fig. 1.8). Ecological communities are interacting species in a given area. Each compartment represents the subnetwork of one community (Dicks et al., 2002). Such structure can arise from co-evolutionary processes, where species in one community have obtained phenotypic traits favouring interactions with species of its community, for example in pollination networks, specific morphological constraints allowing insects to open flowers, matching phenologies or plant attractivity attributes (colors, odor) with respect to insect perception (Olesen et al., 2007; Dupont and Olesen, 2009; Martin Gonzalez et al., 2012; Maruyama et al., 2014; Carstensen et al., 2016). The metric used to quantify this phenomenon is called a modularity index. For a given partition of the collection of the species of a network, an index is computed comparing how much species interact more within their own group than with species of other groups. The modularity value usually retained is the maximal value across all possible partitions (Newman and Girvan, 2004), which is not usually easy to find because it is a NP-complete problem (Fortunato and Barthelemy, 2007; Fortunato, 2010).

One of the most characteristic features of mutualistic networks is their nested structure (Bascompte et al., 2003). Nestedness is the property where both generalist species (having many interaction partners) and specialist species (having few interaction partners) interact preferably with other generalist species, while interactions between specialists species are uncommon (Fig. 1.8). This phenomenon is hypothesized to partially come from the fact that generalist species are either more abundant or yield more resources, which are made available to all species including specialist species (Thompson, 2005; Ollerton et al., 2007). Many indices claim to measure nestedness in a network (Patterson and Atmar, 1986; Wright and Reeves, 1992; Brualdi and Sanderson, 1999; Almeida-Neto et al., 2008; Staniczenko et al., 2013), although they seems to be not always correlated (Podani and Schmera, 2012; Thomas et al., 2015).

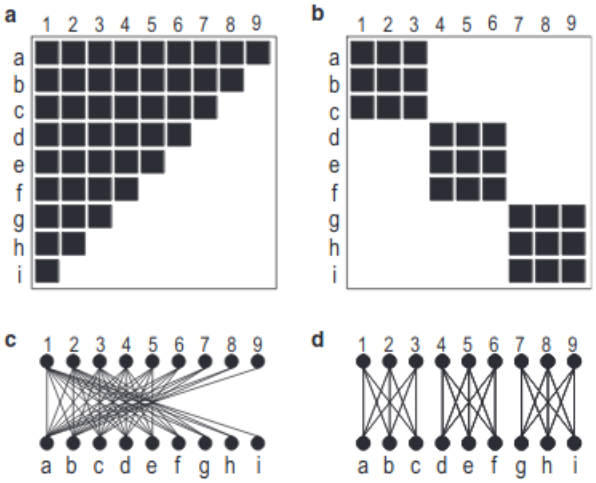

Figure 1.8 – The nested (left) and modular (right) properties for networks and their adjacency matrices. The typically nested network has an upper triangular adjacency matrix, whereas the typically modular network has a diagonal block adjacency matrix. Figure taken from Fontaine et al. (2011).

The modularity and nestedness indices are two examples of classic metrics among many used to describe ecological networks. Other examples of descriptors are the degree distributions (Jordano et al., 2003), the network specialization (Blüthgen et al., 2006) or densities of a particular pattern in the network (Simmons et al., 2019; Lanuza et al., 2023). Although some studies focus on one particular descriptor, it is not uncommon that several are jointly used to analyze networks (Fonseca and Ganade, 1996; Olesen et al., 2007; Thébault and Fontaine, 2010; de Manincor et al., 2020a). Remarkably, these descriptors are usually chosen and interpreted with care. In fact, many metrics are shown to be intercorrelated (Blüthgen et al., 2008; Fortuna et al., 2010) or heavily depend on factors which cannot be controlled, e.g. species abundance (Vázquez et al., 2007). In particular, nestedness metrics have been heavily criticized in recent literature due to their correlation to connectance (Staniczenko et al., 2013) and to the degree distributions (Astegiano et al., 2015; Payrató-Borràs et al., 2019). This can lead to biased interpretations. However, a set of wisely picked metrics is still a convenient way to characterize networks. For this reason, they can be used to capture the variability of the structure of a network, which is an essential question in ecosystem analysis.

1.2.3. Network response to environmental changes

Considering that one network is a snapshot of an ecosystem at a given time and place, one hopes to capture the variability of a network in order to understand how they evolve over time or under various abiotic conditions. Tylianakis and Morris (2017) suggests that ecological network

variability is driven by three main mechanisms: changes in species composition, changes in interaction probabilities, changes in co-evolutionary processes. For example, drastic alterations to networks can arise from invasive species (Aizen et al., 2008; Blanchard, 2015; David et al., 2017) or species removal (Pocock et al., 2012; Heinen et al., 2023).

For this reason, networks give insights on the reaction of ecosystems to external perturbations. Therefore, they can be used to predict the effects of global change (Vitousek, 1994). For example, climate change, the modification of the biogeochemical cycles, the anthropisation of natural environments or biological invasions are known to alter ecosystems (Montoya et al., 2006). On the contrary, some networks are shown to be more robust to these changes (Vila et al., 2009). Thus, the robustness of a network can indicate how fragile an ecosystem is.

Some ecosystemic services, like the pollination of plants by insects, are essential. Understanding the robustness of service-supporting ecological networks is crucial as relevant mitigation strategies and prevention measures have to be rapidly identified and implemented to preserve or restore ecosystems (Gill et al., 2016; Kaiser-Bunbury et al., 2017). Figure 1.9 summarizes the data used by Kaiser-Bunbury et al. (2017) to investigate the effects of vegetation restoration in lands degraded by human impact. They concluded that restored sites yields more resilient pollination networks. Burkle et al. (2013) used long-term observation of pollination networks to identify that these networks had been generally resilient until now. However, progressive phenology shifts alter the probabilities of interaction. They found that networks have become increasingly vulnerable and are coming closer to a tipping point, certainly leading to massive extinction. Therefore, it is paramount to direct conservation efforts to fragile ecosystems. Networks can help identify important species to prioritize as their extinction have the biggest impact on the ecosystems (Memmott et al., 2004; Biella et al., 2017).

Most works aiming to investigate the mechanisms of the structure of ecological networks study the variability of these networks in space and time (Carlo et al., 2003; Dupont et al., 2009; Burkle et al., 2013; Carstensen et al., 2016; Miele et al., 2020). Sometimes, observed networks are associated with environmental covariates. Therefore, many approaches have been developed to compare the networks alongside ecological gradients, for instance, climatic variables (temperature, precipitation) combined with altitude (Ramos-Jiliberto et al., 2010; Dalsgaard et al., 2011; Trøjelsgaard and Olesen, 2013; Pauw and Stanway, 2015) or latitude (Dalsgaard et al., 2011; Trøjelsgaard and Olesen, 2013; Sebastián-González et al., 2015; de Manincor et al., 2020b), species diversity (Dalsgaard et al., 2011; Welti et al., 2017), type of habitat (Schleuning et al., 2011), agricultural land cover (Geslin et al., 2013; Sebastián-González et al., 2015; Redhead et al., 2018; Neff et al., 2021). Commonly investigated network properties as well as adapted network analysis methods with respect to ecological gradients are reviewed in Pellissier et al. (2018). However, most approaches are idiosyncratic, heavily depending on the characteristics

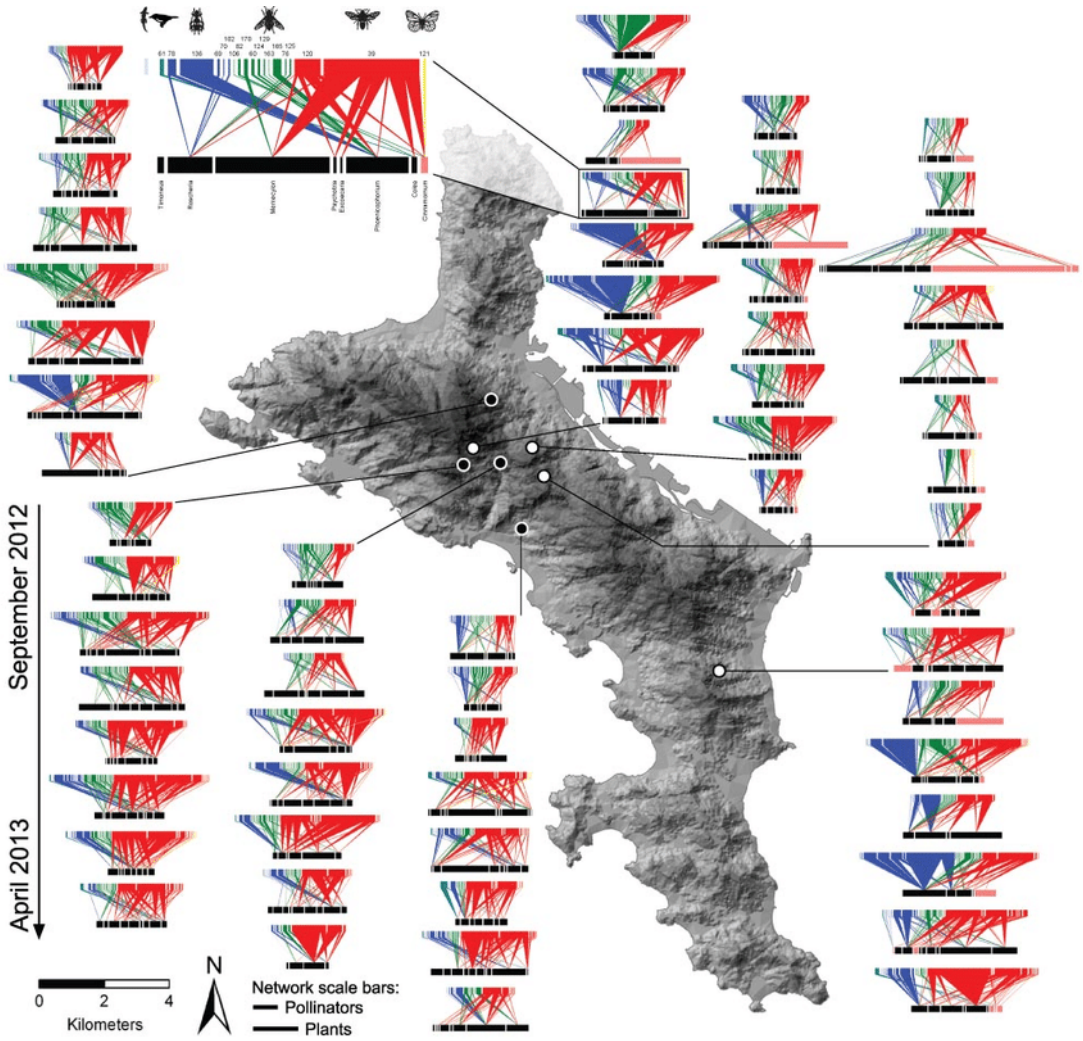

Figure 1.9 – Pollination networks observed in eight sites on the island of Mahé, Seychelles. The vegetation has been restored in four sites (white dots), unrestored in the four remaining (black dots). For each site, eight monthly networks are represented, each corresponding to one month. The width of the links indicates the frequency of interaction between the two species. Colors in the network correspond to families of animal species. Figure taken from Kaiser-Bunbury et al. (2017).

of the data like the associated covariates, the network sizes, the collection of involved species, which makes them complex to generalize to a generic framework applicable to a large range of analyses.

1.2.4. Probabilistic approaches

Ecologists are perfectly aware that an observed network does not represent the complete set of interactions. In particular, sampling bias has been shown to have a great influence on observed networks. Differences in sampling methods can lead to different networks (Bosch et al., 2009; Pornon et al., 2017; Novella-Fernandez et al., 2019; de Manincor et al., 2020a; Dubart et al., 2021). Figure 1.10 shows two networks built from two different measures of interaction strength. Since species abundance is a key determining factor for interaction networks (Dupont et al., 2003; Vázquez and Aizen, 2004; Vázquez et al., 2007), rare species are usually undersampled. As a consequence, these species appear to be specialists in the observed network (Fründ et al., 2016), regardless of their actual role in the ecosystem. Incomplete sampling and sampling effort directly affect various network metrics (Nielsen and Bascompte, 2007; Rivera-Hutinel et al., 2012; Vizentin-Bugoni et al., 2016). For these reasons, an increasing portion of the literature use probabilistic models in an attempt to take into account the uncertainty around the observed networks.

Probabilistic approaches have been extensively used by ecologists for a long time under the framework of the null model approach. More specifically, generative approaches, using random graph models, are more and more employed to analyze ecological network (Bartomeus, 2013; Poisot et al., 2016; Ohlmann et al., 2019; de Manincor et al., 2020b; Herrera et al., 2023). In random network analysis, model-based approaches are opposed to design-based approaches, although both can be combined. Design-based approaches assume the observed network to be sampled from a larger but fixed network according to a sampling design. The randomness comes from the sampling procedure itself and so therefore no random network model is required to derive estimators for quantities of interest on the complete unobserved network (see Dall'Asta et al., 2006, for an example). Conversely, in a model-based approach, the observed network is a realization of a random variable, the distribution of which is specified by a random network model.

Discussion around these two paradigms can be found in Handcock and Gile (2010). Although the sampling bias issue in ecology seems to be perfectly solved by using the design-based paradigm, model-based approaches are in fact more versatile, allowing to incorporate other randomness sources. A random network model can also be used to incorporate the mechanisms that generate the networks. In this case, generative network models can be used. This is a blessing in ecology, where one major aim of network analysis is to understand the ecological processes producing the networks. One would often like to estimate an interpretable parameter of a network model or to directly use the models themselves for prediction purposes (Seo and Hutchinson, 2018; Gravel et al., 2019; Valdovinos, 2019). More importantly, measurement er-

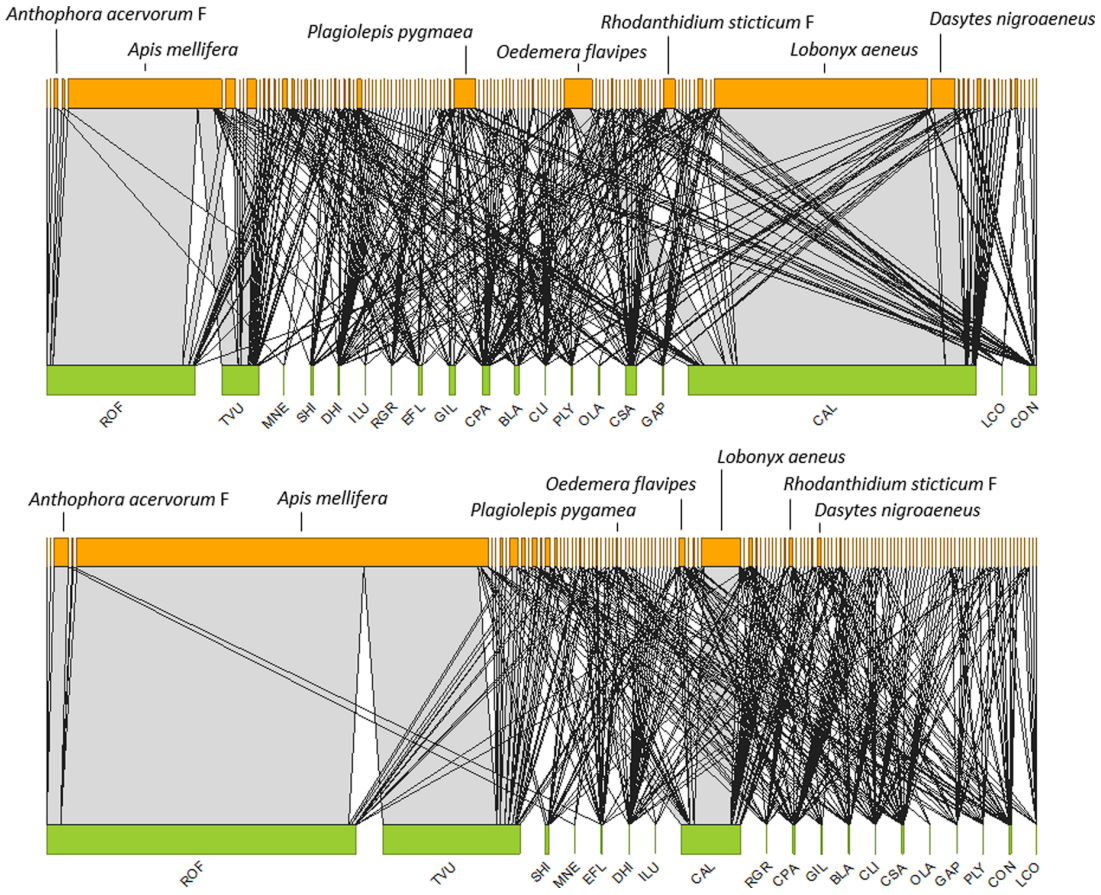

Figure 1.10 – Two networks constructed from the same data. Top: the link width indicates the number of visits of an insect on a plant species. Bottom: the link width indicates the number of flowers visited by an insect on a plant species. Figure taken from Novella-Fernandez et al. (2019).

rors can be obtained more easily than in design-based methods, because in the latter, required sampling probabilities are unobserved in most designs. In this regard, using a model brings more statistical guarantees to the estimators. Hence, for ecological applications, model-based approaches hold decisive advantages over design-based approaches. This is not surprising that the most common network analysis strategy in ecology, the null model approach, falls under the model-based family. For an overview of the background and techniques used in the design-based paradigm, see Frank (2005). This thesis will solely deal with model-based approaches.

1.3. Null model analysis

As we have just seen, in ecological network analysis, one key question is whether the observed network can be explained by ecological theory. Random network models, of which a few classes

have been introduced earlier, can sometimes be used to incorporate the mechanisms that generate the networks. The widely used null model strategy consists in testing the goodness of fit of a random network model, called the null model, with respect to a some particular network metrics. I explain this common approach in ecology and outline some of its limitations, motivating my work to develop a novel model-based methodology.

1.3.1. Principles of hypothesis testing

Before explaining the null model approach, let us recall the principles of statistical hypothesis testing. The objective of hypothesis testing is to use observed data to reject (or by default, to accept) a given hypothesis, called the null hypothesis.

First, we assume that the observed data y is the realization of a probabilistic model, i.e. a random variable Y with some distribution $\mathcal{F}$. Then, we define the null hypothesis $\mathcal{H}_0$ to be tested. Under the null hypothesis, the data Y_0 has a theoretical distribution $\mathcal{F}_0$. The idea behind hypothesis testing is to compare $\mathcal{F}$ and $\mathcal{F}_0$. Ideally, if we can say that $\mathcal{F} \neq \mathcal{F}_0$, then we can reject the null hypothesis $\mathcal{H}_0$ with certainty.

However, in practice, it is often impossible to assess from the observed data whether $\mathcal{F} = \mathcal{F}_0$ or not. We are only able to calculate a probabilistic confidence level based on some function $S(y)$ of the data, so we investigate whether $S(Y)$ has the same distribution as $S(Y_0)$. Sometimes, we only dispose of one observation. In this case, we hope to be able to calculate a p-value p, which is the probability of observing more "extreme" data than y under $\mathcal{H}_0$, e.g. $\mathbb{P}(S(Y_0) \geq S(y))$ or $\mathbb{P}(S(Y_0) \leq S(y))$. If we find a small enough p-value, then we can decide to reject $\mathcal{H}_0$.

In network analysis, y is an observed network, assumed to be a realization of some random Y. The null hypothesis is associated to a random graph model, called the null model, so that the joint distribution of the networks Y_0 generated by the null model is $\mathcal{F}_0$. Consider a network metric $S(\cdot)$ of interest. Under the null hypothesis, the distribution of this statistic is the distribution of $S(Y_0)$, where $Y_0 \sim \mathcal{F}_0$. Like previously, the p-value is defined as the probability to observe a more "extreme" statistic than $S(y)$, e.g. $\mathbb{P}(S(Y_0) \geq S(y))$ or $\mathbb{P}(S(Y_0) \leq S(y))$. The p-value can be used as a criterion to reject $\mathcal{H}_0$. For example, in an ERGM (see equation (1.1)), the elements of the score vector $x(y) = (x_i(y))_i$ can be seen as variables affecting the probability distribution of the graph. It may be interesting to test the effect of one $x(y)_i$ in the model, i.e. whether the probabilities of interaction are a function of this variable. This means that we are looking to reject or accept an hypothesis of the form $\mathcal{H}_0 : \theta_i = 0$, where θ_i is the parameter associated to the variable $x(y)_i$.

1.3.2. Null models in for ecological networks

Principles

In ecology, an interesting question is whether the observed network structure is the product of certain ecological processes or whether it is simply due to chance. An observed network may have interesting properties. However, if a simple random model reproduces these properties, they may not be so significant. Only when the random model does not reproduce these properties can it be said that they are due to other factors that would be worth investigating.

Recall that random graph models introduce heterogeneity and variability in the networks. Each model controls different sources affecting heterogeneity and variability. To determine whether a particular process had an effect on the observed network, the null model approach consists in using a model that theoretically excludes the effects of that process. Therefore, a null model randomly simulates networks that look like they would have without that particular process. The observed network is compared to the simulated networks using network descriptors. In a nutshell, the data is "noisy". We are unsure whether there is a real "signal" (the ecological process of interest) behind this noise. The null model attempts to simulate noisy data, without this potential signal. If the empirical data differs significantly from the data simulated by the null model, then this determines the presence of a signal.

For this question, the use of the statistical hypothesis testing framework is appealing. However, in studies using the null model approach, the null hypotheses are usually not rigorously defined. Still, to determine if the data differs from the distribution defined by the null model, ecologists use p-values, so this is a statistical test under disguise. Defining a null model means setting a distribution $\mathcal{F}_0$ associated to a null hypothesis $\mathcal{H}_0$. For a correct interpretation, one should correctly identify $\mathcal{H}_0$ with ecological terms, which is not as simple as it seems and it is something often overlooked in null models studies.

Models commonly used

The aim of a null model is to generate "maximally" random networks, nullifying a particular process. For ecological networks, the most common null models consist in shuffling the data (i.e. the network edges) under a set of constraints, which correspond to the processes that are retained in the model. We can call these null models constraints-based (see Tab. 1.1 for a summary of model types).

Classic constraints for network null models are the number of nodes in the network, the number of edges and the node degrees (marginals). Connor and Simberloff (1979), for the first

Random graph models		Network null models	
Procedure-based	Likelihood-based	Constraint-based	Process-based

Table 1.1 – The types of models defined in this chapter. These categories are not mutually exclusive. Random graph models can be used as network null models. A random graph model can be both procedure-based and likelihood-based. A network null model can be both constraint-based and process-based.

occurrence of the null model approach in the ecological literature, used these constraints (and others) to investigate "checkerboard" patterns in presence-absence matrices representing the distribution of birds species in the islands of an archipelago. In a checkerboard pattern, some species can be regrouped into sets. Only one of the species of a given set is present on a given island, i.e. no pair of these species is never observed on a given island. Diamond (1975) observed that these sets often regroup similar species. He interpreted these checkerboard patterns as the result of assembly rules dictated by competition between similar species, where some cannot co-occur as a result. Connor and Simberloff (1979) used a configuration model (CM) to simulate networks with the same number of links and node degrees as the observed one. This means that in the generated networks, each island has the same number of species as in the observed network, and each species are present on the same number of islands as in the observed network. They found that the checkerboard pattern is equally common in the simulated networks and in the observed one, implying that this pattern can actually arise by chance, casting doubt above Diamond's theory.

Heuristically, the constraints in the null model of Connor and Simberloff (1979) are used to make the simulated data resemble the observed data. The null model here is used to randomize the organization of the network under some constraints respected by the observed data. In terms of statistical testing, using the CM as a null model is equivalent to restrict the support of $\mathcal{F}$ and $\mathcal{F}_0$ to graphs that have the same node degrees than the observed ones. The CM stayed widely used in subsequent null model studies.

Limitations

Despite the simplicity of this approach, it is difficult to ecologically interpret the results of the above-mentioned approach without ambiguity. The ecological meaning of $\mathcal{H}_0$ is not straightforward. In statistical testing, the statistician defines $\mathcal{H}_0$ first. This is the hypothesis in which they are interested, so $\mathcal{F}_0$ is only defined as a consequence. In most null models studies, the choice of a set of constraints determines the null model, i.e. $\mathcal{F}_0$. The translation of $\mathcal{F}_0$ into ecological concepts $\mathcal{H}_0$ is only subsequent. In the CM, only the marginals are included

in $\mathcal{H}_0$. Using a particular sequence of degrees is a way to capture the heterogeneity of the observed network nodes and inject it in the null model. However, the ecological processes being controlled here are unclear. Many phenomena can affect the node degrees, e.g. species abundance, competition, phenotype, preferential attachment, etc. In fact, these phenomena might also affect the networks in other ways. Therefore, the null model partially nullifies the effect of these phenomena, but only beyond their effect on marginals. The same interrogation subsists for all the constraints that are not the direct product of an ecological process.

The second sensible caveat of this approach lies in the nature of the simulated networks. The null model simulates data from the distribution $\mathcal{F}_0$. Therefore, each simulation should be a replicate of the experiment under $\mathcal{H}_0$. Ecologically, the simulated networks should represent independent observations of an ecosystem in the same conditions. Putting such hard constraints drastically reduces the support of $\mathcal{F}$ and $\mathcal{F}_0$, so the subset of networks satisfying the marginal constraint might be very small. For some configurations, the observed network is the unique network that satisfies them, e.g. for unipartite networks, all the networks satisfying the star pattern, where 1 node have degree $n-1$ and $n-1$ nodes have degree 1, are the same up to node permutations. This is highly unrealistic since networks are observations of some particular ecosystems, so if we observe an ecosystem several times, it is highly probable that the associated networks have different row and column sums. For example, sampling effort has a strong effect on them. Most often, an observed network is only an incomplete sample of the full set of interactions. In this regard, using the exact sequence of marginals in the generated networks introduces a bias because it does not reflect the uncertainty due to this sampling.

Finally, a last point overlooked by the null model studies is that the distribution of the network simulated by procedure-based models is not so obvious. Since the distribution of networks is untractable, there is sometimes a gap between what the ecologist expects to simulate and what is actually simulated. In fact, the algorithm used by Connor and Simberloff (1979) to simulate under the CM is biased and $\mathcal{F}_0$ is not a uniform distribution on its constrained marginals support. With the current state of the art algorithms, it is now possible to sample correctly from the CM but this problem has been neglected for a long time. Even so, the reason why one wants a uniform distribution is unclear. Similar to the CM, one might be tempted to use some algorithm to generate random networks with a particular property. However, the null distribution is biased for other properties that were not accounted for. For example, the Barabasi-Albert model can be used to generate scale-free networks, i.e. networks with a power law degree distribution. However, there exists many other models that can generate scale-free networks. It has been shown that the networks generated by different models may not present the same structural properties, despite having the same power law degree distribution (Grisi-Filho et al., 2013; Tsiotas, 2019). Therefore any of these models generates scale-free networks,

but with a different distribution, i.e. they are associated with different null hypotheses $\mathcal{H}_0$. It is uncertain how one interprets these differences, since the simple constraints are the same, but the complex distributions are different.

It is often not obvious to identify the null hypothesis $\mathcal{H}_0$ for a constraint-based null model, but there are in fact cases where $\mathcal{H}_0$ can be directly clearly defined, and the null model is derived from $\mathcal{H}_0$ instead of the opposite. For an ecological question, captured by a null hypothesis $\mathcal{H}_0$, it might be possible to find a relevant null model to test exactly this hypothesis. Many variations of the CM have been employed later with relaxed constraints, partially fixing some limitations of the CM. Gilpin and Diamond (1982) use a network model where the individual probability of interactions depend on row and column parameters, defining "expected" degrees instead of fixed degrees. Their model would be the bipartite Chung-Lu model described earlier. Instead of fixing hard constraints, this model allows some variability on the degrees by only fixing constraints on the moments of the node degrees in $\mathcal{F}$ and $\mathcal{F}_0$, meaning that its support of is not restricted anymore (except for the number of nodes). Bascompte et al. (2003) used two null models, the bipartite version of the ER and a model where the interaction probability between two nodes is proportional to the average of the observed degrees of these nodes to investigate the nestedness property of mutualistic networks. Vázquez and Aizen (2004) used a model where the interaction probability between two nodes is proportional to both node degrees, a variation of the Chung-Lu model, to highlight the asymmetric specialization in plant-pollinator networks, which is a notion similar to nestedness. These models soften the constraints but their interpretation is still difficult. Dormann et al. (2009) generated "artificial" networks by sampling the node degrees in a particular distribution, then used a version of the configuration model but for weighted networks. We can argue that these approaches make a step forward in the right direction. The introduction of another source of variability helps better underline the fact that the simulated networks are replicates of an actual experiment.

Another type of null models

One way to control the ecological hypotheses included (or excluded) in $\mathcal{H}_0$ is to use generative network models, based on ecological processes. With this approach, $\mathcal{H}_0$ is first defined in ecological terms. The set of ecological processes assumed to affect the network and the ones that are being tested are identified. Because the underlying processes generating the network are directly modeled, these models are likelihood-based. The distribution $\mathcal{F}$ from which the observed network is drawn and the null distribution $\mathcal{F}_0$ are defined as follows :

- $\mathcal{F}$ is unknown but assumed to belong to some family of models so that all the ecological mechanisms that need to be controlled are included,

- $\mathcal{F}_0$ belongs to the same family but is obtained by some "restriction" (e.g. a particular combination of parameters) nullifying the processes that are excluded in $\mathcal{H}_0$ (e.g. setting some parameters to 0).

Such an approach would solve all the issues previously discussed. The main limitation is the difficulty to find a suitable model and decide which processes to include (Wilber et al., 2017). Nevertheless, the use of such models is becoming increasingly common in ecology, beyond the null model approach. Rankin et al. (2016) used ERGMs as null models for the study of animal social networks. de Manincor et al. (2020b) used latent block models (LBM), the bipartite version of the SBM, to study plant-pollinator networks along a latitudinal gradient. Wells and O'Hara (2013) goes beyond the species-level and also modeled the processes at the individual-level. Process-based models also allow to make predictions, notably with the use of covariates (Bartomeus, 2013; Valdovinos, 2019).

1.4. Exchangeable random network models

In the previous sections, I have outlined the context in which the contributions of this thesis are inserted. I have reviewed a few models used in network analysis and one of their main use in ecological network analysis, as null models. As we have seen, the null model approach suffers from some limitations. My objective is to define a novel framework to study ecological networks, overcoming these limitations. The cornerstone of my framework is exchangeable random graph models. The sources of heterogeneity and variability introduced by random graph models are quintessential to derive statistical guarantees in network analysis. We have seen that likelihood-based models partially solve the limitations of the null model approach. My framework makes use of likelihood-based network models which are easily parameterized and with an additional assumption: node exchangeability. This assumption seems reasonable for ecological interaction networks and allows to obtain convenient mathematical properties on these models. I give the mathematical definition of exchangeability and how it may apply to network data. We discuss its ecological interpretations. Finally, I introduce some notions for exchangeable network models, which will be the class of models used in my framework.

1.4.1. Mathematical definition of exchangeability

Exchangeable sequences

Exchangeability is a general probabilistic concept under which a sequence of random variables is invariant under permutations. Denote $\mathbb{S}_n$ the group of permutations of order n (only changing

the first n integers) and $\mathbb{S}_\infty = \bigcup_{n=1}^{\infty} \mathbb{S}_n$ the group of finite permutations over $\mathbb{N}$.

Definition 1.4.1. *An infinite sequence of random variables $X = (X_1, X_2, ...)$ is exchangeable if and only if for all permutation $\sigma \in \mathbb{S}_\infty$,*

$$(X_{\sigma(1)}, X_{\sigma(2)}, ...) \overset{\mathcal{D}}{=} X.$$

We observe that exchangeability is a probabilistic symmetry weaker than the i.i.d. assumption. If $(X_1, X_2, ...)$ is exchangeable, then all the X_i, for $i \geq 1$ are identically distributed. However, they are not necessarily independent. In fact, the concept of exchangeability for infinite sequences is closely related to the concept of "i.i.d.-ness". These two concepts are linked by de Finetti's theorem, which is often put into words by the following sentence *"An infinite exchangeable X is a mixture of i.i.d. sequences."*. Here is a more formal version of this theorem.

Theorem 1.4.2. *An infinite sequence $X = (X_1, X_2, ...)$ with values in a measurable set $(E, \mathcal{E})$ is exchangeable if and only if there exists a probability measure μ on $(E, \mathcal{E})$ such that for any $N \in \mathcal{N}$, for all $A = (A_1, ..., A_N) \in \mathcal{E}^N$,*

$$\mathcal{P}((X_1, ..., X_N) \in A) = \int_{\Pi(E)} \pi(A_1)...\pi(A_N)\mu(d\pi),$$

where $\Pi(E)$ is the set of probability measures on $(E, \mathcal{E})$.

This version is consistent with the above quote: it says that if X is an infinite exchangeable sequence, then X is i.i.d. conditionally to some σ-field $\mathcal{F}_\infty$. $\mathcal{F}_\infty$ is generated by latent variables that are not measurable by any X_i, $i \geq 1$. Actually, this σ-field can be shown to be the tail-field $\mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, ...)$. Therefore, a more sophisticated version of this result, appearing in Kallenberg (2005), can be stated under the form of a representation theorem.

Theorem 1.4.3. *$X = (X_1, X_2, ...)$ is exchangeable if and only if there exists a function $f : [0,1]^2 \to E$ and α and $(\xi_i)_{i \geq 1}$ i.i.d. random variables with distribution $\mathcal{U}[0,1]$ such that for all $i \geq 1$,*

$$X_i \overset{a.s.}{=} f(\alpha, \xi_i).$$

This theorem can be powerful at it gives a representation of infinite exchangeable sequences with i.i.d. variables. However, from a practical point of view, we usually only observe a finite sequence of random variables. The definition of exchangeability is slightly different for finite sequences.

Definition 1.4.4. *An finite sequence of random variables $(X_1, ..., X_n)$ is exchangeable if and only if for all permutation $\sigma \in \mathbb{S}_n$,*

$$(X_{\sigma(1)}, ..., X_{\sigma(n)}) \overset{\mathcal{D}}{=} (X_1, ..., X_n).$$

The exchangeability of a finite sequence is a weaker property than the exchangeability of an infinite sequence. Indeed, not all finite exchangeable sequence are said to be infinitely extendible, i.e. they can not always be written as the first elements of an infinite exchangeable sequence. For example, consider that $(X_1, ..., X_n)$ is the result of a random draw from a pool of $N \geq n$ elements without replacement. $(X_1, ..., X_n)$ is indeed a finite exchangeable sequence, but we explain why $(X_1, ..., X_n)$ is not infinitely extendible. Drawing all N elements from the pool, the sequence $(X'_1, ..., X'_N)$ is such that $(X_1, ..., X_N) \overset{\mathcal{D}}{=} (X'_1, ..., X'_N)$. Therefore, we say that $(X_1, ..., X_n)$ is N-extendible. However, we cannot find an infinite exchangeable sequence $X'' = (X''_1, X''_2, ...)$ such that $(X_1, ..., X_n) \overset{\mathcal{D}}{=} (X''_1, ..., X''_n)$. As a consequence, $(X_1, ..., X_n)$ is not infinitely extendible. Infinite extendibility is the necessary condition for finite exchangeable sequences to be represented with i.i.d. variables via the de Finetti's theorem (Konstantopoulos and Yuan, 2019; Mai, 2020). In later sections, for finite sequences, we will refer to infinite exchangeability, or more simply *exchangeability*, when they are exchangeable and infinitely extendible. We will explicitly write *finite exchangeability* when the sequences are not infinitely extendible.

Exchangeable arrays

Arrays of random variables can describe more complex data, such as network data or relationship between more than two entities. Let $X = (X_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathbb{N}^k}$ be an infinite array of random variables indexed by k-tuples $\boldsymbol{i} = (i_1, i_2, ..., i_k)$. Typically, if X is a network adjacency matrix, then X is a two-dimensional array ($k = 2$). For this type of more complex data, the notion of exchangeability is ambiguous. Indeed, there are different ways to apply permutations on the indices of multidimensional arrays.

The equivalent notion to the unidimensional case is when all the entries of the arrays are fully exchangeable (Adamczak et al., 2016), i.e. for all permutations σ over $\mathbb{N}^k$, the set of all k-tuples,

$$(X_{\sigma(\boldsymbol{i})})_{\boldsymbol{i} \in \mathbb{N}^k} \overset{\mathcal{D}}{=} X.$$

If X is a network adjacency matrix, this property means that the edges of the network are exchangeable.

Full exchangeability is not always suitable for relational data. In network data, it means for example that $(X_{12}, X_{13}) \overset{\mathcal{D}}{=} (X_{12}, X_{34})$. This is sometimes too strong an assumption, as one might expect the couple (X_{12}, X_{13}) to behave differently due to having one node in common, as opposed to (X_{12}, X_{34}). Many exchangeable network models often assume that not the edges, but only the nodes are exchangeable. Node-exchangeability can be translated to two different mathematical properties : joint and separate exchangeability.

Definition 1.4.5. *Let $X = (X_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathbb{N}^k}$ be a k-dimensional array of random variables.*

- *X is jointly exchangeable if for any permutation $\sigma \in \mathbb{S}_\infty$,*

$$(X_{\sigma(i_1)\dots\sigma(i_k)})_{(i_1,\dots,i_k)\in\mathbb{N}^k} \overset{\mathcal{D}}{=} X,$$

- *X is separately exchangeable if for any set of k permutations $(\sigma_1,\dots,\sigma_k) \in \mathbb{S}_\infty^k$,*

$$(X_{\sigma_1(i_1)\dots\sigma_k(i_k)})_{(i_1,\dots,i_k)\in\mathbb{N}^k} \overset{\mathcal{D}}{=} X.$$

The two properties are remarkably related. A separately exchangeable infinite array is also jointly exchangeable. Separate exchangeability implements additional symmetries that are not found in joint exchangeability. Although mathematically close, in practice, there is, however, only little ambiguity to distinguish the two properties. The type of exchangeability depends on the allowed permutations on the indices and this is usually known. In a unipartite networks, shuffling the interacting entities with a permutation σ yields the adjacency matrix $(X_{\sigma(i)\sigma(j)})_{i\geq1,j\geq1}$. In a bipartite network adjacency matrix, applying a permutation σ_1 to the first set of nodes and another permutation σ_2 to the second set yields the matrix $(X_{\sigma_1(i)\sigma(2)(j)})_{i\geq1,j\geq1}$. Since the networks of interest in my work represent mutualistic interactions, I will mainly consider the latter case. Because the two indices of the adjacency matrix represent its rows and columns, separate exchangeability of the matrix can be interpreted as if the rows and the columns of the matrix are separately exchangeable. For that reason, we will say that the separately exchangeable matrix is row-column exchangeable (RCE).

Definition 1.4.6. *An infinite matrix Y is row-column exchangeable (RCE) if and only if for any couple of permutations $(\sigma_1,\sigma_2) \in \mathbb{S}_\infty^2$,*

$$(Y_{\sigma_1(i)\sigma_2(j)})_{i\geq1,j\geq1} \overset{\mathcal{D}}{=} Y.$$

Infinite exchangeable arrays admit analogous representations as the de Finetti's theorem. A similar representation first arose from independent studies of Hoover (1979) and Aldous (1981). Kallenberg's colossal work has tidied up this framework and extended the theory to different types of distributional symmetries, most of which is covered in Kallenberg (2005). For these reasons, this theory is often referred to as the Aldous-Hoover-Kallenberg (AHK) framework. Here, we only give the representation theorems for jointly exchangeable matrices and for RCE matrices.

Theorem 1.4.7 (AHK for jointly exchangeable matrices)**.** *$X = (X_{ij})_{i\geq1,j\geq1}$ is jointly exchangeable if and only if there exists a function $f : [0,1]^4 \to E$ and α, $(\xi_i)_{i\geq1}$ and $(\zeta_{ij})_{i\geq1,j\geq1}$ i.i.d. random variables with distribution $\mathcal{U}[0,1]$ such that for all $i \geq 1$, $j \geq 1$,*

$$X_{ij} \overset{a.s.}{=} f(\alpha,\xi_i,\xi_j,\zeta_{ij}). \tag{1.3}$$

Theorem 1.4.8 (AHK for RCE matrices)**.** *$X = (X_{ij})_{i\geq 1, j\geq 1}$ is RCE if and only if there exists a function $f : [0,1]^4 \to E$ and α, $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ and $(\zeta_{ij})_{i\geq 1, j\geq 1}$ i.i.d. random variables with distribution $\mathcal{U}[0,1]$ such that for all $i \geq 1$, $j \geq 1$,*

$$X_{ij} \overset{a.s.}{=} f(\alpha, \xi_i, \eta_j, \zeta_{ij}). \tag{1.4}$$

Like in the uni-dimensional case, this representation has been given for infinite matrices. To apply this result to networks of finite size, their finite adjacency matrices must be infinitely extendible. In this framework, we will consider that an observed adjacency matrix is the sub-matrix consisting of the leading rows and columns of an infinite RCE matrix. This approach is common in random network analysis (Orbanz and Roy, 2014; Veitch and Roy, 2015). This ensures that the observed network is infinitely extendible and can be represented by the Aldous-Hoover-Kallenberg (AHK) representation.

1.4.2. Ecological implications of exchangeability

Exchangeability introduces probabilistic symmetries in the random graph models which are useful mathematical properties. One can question whether a strong assumption such as node exchangeability is relevant for applications. Our goal is to study the organization of ecological ecosystems, i.e. the general structure of species interaction networks. With this objective in mind, exchangeability is an acceptable assumption.

From a probabilistic point of view, exchangeability means that the network distribution remains the same when the nodes are shuffled. Since our questions of interest only revolve around the general structure of the network, we can consider that two adjacency matrices only differing by a permutation of the rows or the columns represent the same network. Common network metrics such as connectance, modularity, nestedness or centrality do not depend on the order of the nodes. In many ecological studies, exchangeability would be a transparent assumption.

As a consequence of exchangeability, the labels of the nodes do not play a role in the network models. Exchangeability means "working without species names". If the property of the network does not depend on the particular ordered list of species, it should persist when rows and columns of the adjacency matrix are exchanged. Therefore, this does not matter when studying the general structure of the networks, but any analysis involving a particular set of species invalidates the assumption of exchangeability. Also, any network analysis method involving node correspondence is rendered pointless. However, this does not mean that species names cannot be used for interpretation of the results because of exchangeability. For example, one might fit a model with exchangeable latent variables on an observed network and then interpret

the relative position of some particular species in the latent space. Taxonomy is only omitted in the network model used for the mathematical analysis of the structure.

Exchangeability means "working without species names" – if the property of the network does not depend on the particular ordered list of species, it should persist when rows and columns of the adjacency matrix are exchanged.

In fact, many null model analyses already make an underlying exchangeability assumption. This applies to most null models shuffling the interactions but keeping node properties, e.g. the configuration model keeping the exact degrees or the Chung-Lu model keeping the expectation of the degrees. The goal of these models is to generate networks with a similar structure (for example, determined by the constraints of the null model) while severing the relation between a particular observed node and its edges. Some null models are even explicitly exchangeable, for example Bassetti et al. (2007) or Dormann et al. (2009).

Using exchangeable models, it is possible to jointly study networks independently of the collection of species involved in each network. This is of great interest in ecology. For instance, when studying the evolution of networks with respect to some environmental gradient, data consist in several networks observed under different conditions (geographical location, time, weather, etc.). Species composition often differ between the networks, which may have different sizes.

1.4.3. Exchangeable models

Exchangeability in random graph models

For networks, exchangeability can refer to the nodes but also to the edges. From a mathematical point of view, edge-exchangeability means that the edges can be viewed as an exchangeable sequence, while node-exchangeability implies that the adjacency matrix is a jointly exchangeable (or RCE) array. Recent research has been carried on edge-exchangeable models by Cai et al. (2016), Williamson (2016) and Crane and Dempsey (2018). However, node-exchangeability is an assumption more fitted to the ecological context. Node-exchangeable models have been studied for more than two decades and is backed by strong theoretical developments. From now, network and graph exchangeability refer to node-exchangeability. Among the random graph models presented in the previous section, some are exchangeable or may only require minor modifications to be exchangeable.

The $G_{n,M}$ model is not exchangeable but only finitely exchangeable due to the fact that the number of edges M is constant even if n grows. The BA and WS are procedure-based

models that are not even finitely exchangeable per se. Normally, the BA starts with a fixed network and randomly connected nodes are sequentially added. The probabilities of interaction of a node depend on which nodes come first. The BA is not exchangeable by design. However, Bollobás et al. (2007) suggested an exchangeable model reproducing some properties of the BA, including edge probabilities and degree distribution. The WS starts with a fixed lattice circular network and sequentially modifies the edges connected to each node, in a particular fixed order. The WS becomes finitely exchangeable if the order of the modified nodes is randomly and uniformly chosen, but not exchangeable. The CM can be finitely exchangeable if one consider that the nodes draw their degree without replacement from the specified degree sequence. If the node degrees are drawn with replacement, then this is a degree distribution model, which is exchangeable. Finally, the ERGMs are generally not finitely exchangeable, but all finitely exchangeable models can be written as ERGMs (Lauritzen et al., 2018).

In constrast, in the $G_{n,p}$ model, the edges are i.i.d., therefore the nodes are exchangeable. The infinite extendibility comes from the fact that the $G_{n,p}$ model is invariant when the number of nodes n grows, i.e. for two integers $n_1 < n_2$, matrices generated by $G_{n_1,p}$ have the same distribution than submatrices of size n_2 of matrices generated by $G_{n_2,p}$ networks. The CM variations using degree or expected degree distributions are also infinitely exchangeable because the node degrees are i.i.d. Similarly, the latent space models formalized earlier become exchangeable when the latent variables associated to the nodes are i.i.d. With respect to the AHK representation theorem of exchangeable arrays, this is not surprising how the "i.i.d.-ness" may form exchangeable models. A class of latent space models encompassing all the exchangeable models will be made explicit in the next paragraph.

The graphon

Exchangeable random graphs can be directly linked to dense graph limit theory via an object called the graphon. This object is a general tool to build and to study exchangeable graph models. Here, I introduce some notions of dense graph limit theory and, defining the concept of graphon, I explain how it relates to the limit of exchangeable random graph models and the AHK representation of exchangeable arrays. For convenience, we consider binary graphs. The extension to weighted graphs will be done later. The material in this section is essentially adapted from Lovász and Szegedy (2006).

We need to define a few nations before defining the graphon. Let $G = (V(G), E(G))$ be a binary graph. The *number of automorphisms* $\mathrm{Aut}(G)$ of G is the number of permutations on its vertices which leave the graph unchanged. Let H be another finite graph. The *number of*

homomorphisms, i.e. copies, of H in G is given by

$$\mathrm{Hom}(H,G) = \frac{1}{|\mathrm{Aut}(H)|} \sum_{\mathbf{i}\in\mathcal{P}_{v(H)}(V(G))} \prod_{(k,\ell)\in E(H)} A(G)_{i_k i_\ell},$$

where $v(\cdot)$ is the number of vertices of a graph and $A(G)$ is the adjacency matrix of G. If $v(G) \leq v(H)$, then $\mathrm{Hom}(H,G) = 0$. The *homomorphism density* $t(H,G)$ is obtained by normalizing $\mathrm{Hom}(H,G)$ as follows

$$t_H(G) = \frac{\mathrm{Hom}(H,G)}{v(G)^{v(H)}},$$

so that $t_H(G) \in [0,1]$ is the probability of obtaining H when sampling $v(H)$ nodes from $V(G)$ and returning the induced subgraph. $\mathrm{Hom}(H,G)$ (resp. $t_H(G)$) is also sometimes called a subgraph count (resp. density) or a motif count (resp. density).

Let $(G_n)_{n\geq 1}$ be a finite graph sequence. $(G_n)_{n\geq 1}$ is said to be *dense* if the number of edges $e(G_n) \asymp v(G_n)^2$. Lovász and Szegedy (2006) proved the following theorem, which is the keystone of graphon theory, applying to dense graph sequences.

Theorem 1.4.9. *Let $(G_n)_{n\geq 1}$ be a dense graph sequence. If $(G_n)_{n\geq 1}$ converges, i.e. $t_H(G_n)$ converges for all finite graphs H, then there exists a graphon, i.e. a symmetric measurable function $w : [0,1]^2 \to [0,1]$, such that*

$$t_H(G_n) \xrightarrow[n\to\infty]{} t_H(w) := \int_{[0,1]^{|v(H)|}} \prod_{(k,\ell)\in e(H)} w(x_k, x_\ell) dx_1 ... dx_{|v(H)|}.$$

There are several takeaways from this theorem. First, graph convergence here means convergence of the subgraph densities. The fact that the $(G_n)_{n\geq 1}$ should be dense is essential and we will see how it to relates to exchangeability later. Second, the graphon is simply a symmetric measurable function $w : [0,1]^2 \to [0,1]$ and $t_H(w)$ can be defined for all graph H. The expression $t_H(w)$ strongly resembles the formula of the homomorphism density for H in a graph G. In fact, w seems to play the role of the adjacency matrix $A(G)$ but continuous and rescaled in a square $[0,1]^2$. This intuition is actually spot on, as we will see the role of w in the graphon model. The key takeaway from this theorem is that any graph sequence limit can be represented by a graphon. Next, we extend this notion to random graphs.

Consider the $G(n,p)$ model. For all n, let G_n be one realization of the $G(n,p)$ model. We may abusively write $A(G_n) \sim G(n,p)$, for some fixed $p \in]0,1]$. Since $A(G_n)_{ij} \overset{i.i.d.}{\sim} \mathcal{B}(p)$, then for any graph H, we have $t_H(G_n) \xrightarrow[n\to\infty]{a.s.} p^{|e(H)|}$. This means that not only the limit of G_n can be represented by the constant graphon $w \equiv p$ but this is a hint that $w \equiv p$ can also represent the limit of the sequence of random graph models, here the $G(n,p)$ models.

More generally, for some graphon w, define the $G(n,w)$ models such that for all $n \geq 1$, $A(G_n) \sim G(n,w)$ is generated by

$$\begin{aligned} \xi_i &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq i \leq n, \\ A(G_n)_{ij} \mid \xi_i, \xi_j &\sim \mathcal{B}(w(\xi_i, \xi_j)), && \forall 1 \leq i, j \leq n. \end{aligned}$$

The $G(n,w)$ is a latent space model where the latent variables $(\xi_i)_{1\leq i\leq n}$ are i.i.d. Obviously, this is a node-exchangeable model. Again, we can show that for any graph H, we have $t_H(G_n) \xrightarrow[n\to\infty]{a.s.} t_H(w)$. Therefore, w can be considered as the limit of the exchangeable random graph models $G(n,w)$ (Fig. 1.11). This model is called the W-graph model or the graphon model. Every W-graph or graphon model can be represented by its limit, which is no other than its own graphon.

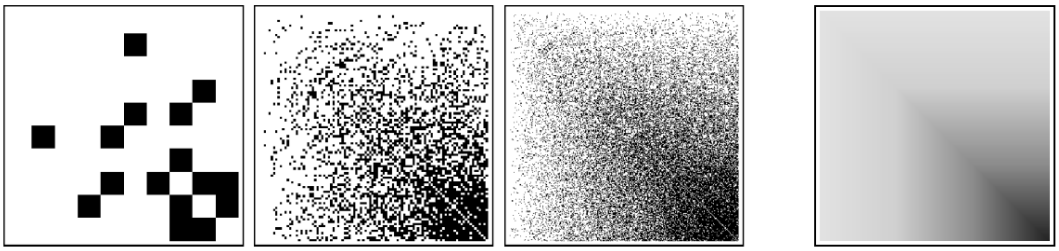

Figure 1.11 – The graphon as the limit of the adjacency matrices of $G(n,w)$ graphs. Left: Three adjacency matrices of $G(n,w)$ binary graphs for increasing values of n, unique w. Right: The corresponding graphon w. Figure taken from Orbanz and Roy (2014)

The graphon model bridges the gap between the exchangeable random graph models and the graph limit theory. Not only is the graphon model is exchangeable but, more interestingly, the converse is true as well, i.e. exchangeable models can be written as a graphon model, under an additional assumption. From the AHK representation theorem, if α is constant, then a jointly exchangeable array can be written as a graphon model. The property implying that α is constant is called dissociatedness. A jointly exchangeable matrix Y is said to be *dissociated* if for all $n \geq 1$,

$$(Y_{ij})_{1\leq i,j\leq n} \text{ is independent of } (Y_{ij})_{i>n,j>n}.$$

Indeed, after ensuring α is constant, setting

$$f(\xi_i, \xi_j, \zeta_{ij}) = \mathbb{1}\{\zeta_{ij} \leq w(\xi_i, \xi_j)\}$$

in equation (1.3) gives the graphon model. Therefore, every dissociated exchangeable random graph model can be represented by a graphon. The graphon is a useful object to study complex models. Although the graphon might be complex, it is usually described by a unique formula.

Even though I have only introduced the theory for unipartite binary graphs. The theory has been extended for bipartite graphs (Diaconis and Janson, 2008) and for weighted graphs

(Lovász and Szegedy, 2010). For bipartite graphs, a "bipartite" graphon can be defined, which only differs from a classic graphon because it is not required to be symmetric. The graphon model is written with two families $(\xi_i)_{1\leq i\leq m}$ and $(\eta_j)_{1\leq j\leq n}$ of i.i.d. variables instead of one and each conditional edge probability is given by $w(\xi_i, \eta_j)$. For weighted graphs, instead of a classic graphon, a "colored" graphon is considered. More details will be given in Section 1.6.1. Because all the types of dissociated exchangeable graphs models can be represented by a graphon, this unique object can be used to study these models without loss of generality.

Now, I have given a general introduction to the family of exchangeable models. Next, I will present the second major object used in my framework: U-statistics.

1.5. U-statistics

My framework aims to perform statistical inference on random graph models. Since I have not tackled the subject of inference yet, it is maybe still unclear where the exchangeability assumption might be helpful. The class of estimators used to perform statistical inference on these models in my framework is the class of the U-statistics for network data. U-statistics have a long history in the literature, especially for independent and identically distributed data. In this chapter, I provide a gentle introduction to U-statistics, mostly focusing on i.i.d. data. I put emphasis on limit theorems and a few techniques and tools that are commonly used to obtain them. I keep in mind that I am interested in extending U-statistics to dependent data, especially exchangeable network data, but in this section, I only give some hints foreshadowing how exchangeability will come into play in my framework. This section is built upon a compilation of elements from the textbooks of Hall and Heyde (1980), Lee (1990), Korolyuk and Borovskich (1993) and Van der Vaart (2000). Proofs for most results can be found in at least one of these references.

1.5.1. The basics

Let $\mathcal{P}$ be a family of probability distribution functions on a set E. Let $(X_1, ..., X_n)$ be a sequence of independent and identically distributed (i.i.d.) random variables with distribution function $P \in \mathcal{P}$. Let $h^0 : E^k \to \mathbb{R}$ be a function of k variables, $k \leq n$, and $h : E^k \to \mathbb{R}$ be its symmetrized version, i.e. h is defined following

$$h(x_1, ..., x_k) = \frac{1}{k!} \sum_{\pi \in \mathbb{S}_k} h^0(x_{\pi(1)}, ..., x_{\pi(k)}),$$

where $\mathbb{S}_k$ is the symmetric group of order k. The U-statistic on $(X_1, ..., X_n)$ with kernel h is defined as

$$U_n^h(X_1, ..., X_n) = \binom{n}{k}^{-1} \sum_{\mathbf{i} \in \mathcal{P}_k([\![n]\!])} h(X_{\mathbf{i}}),$$

where $\mathbf{i} := \{i_1, ..., i_k\}$, and by symmetry of h, we can write $h(X_{\mathbf{i}}) := h(X_{i_1}, X_{i_2}, ..., X_{i_k})$ because it does not depend on the order of the elements $\mathbf{i}$. We assume that $\mathbb{E}[h(X_{\mathbf{i}})^2] < \infty$. We will solely focus on the case $E = \mathbb{R}$, unless when explicitly specified. When there is no ambiguity, we will denote $U_n := U_n^h(X_1, ..., X_n)$.

Let $\theta(P)$ be a real-valued functional defined on $\mathcal{P}$ such that $\mathbb{E}_P[h^0(X_1, ..., X_k)] = \theta(P)$. It is straightforward that $\mathbb{E}_P[U_n^h] = \theta(P)$. Halmos (1946) proved that U_n^h is the unique symmetric function of n variables that is an unbiased estimator of $\theta(P)$ and in addition, it is the unbiased estimator of $\theta(P)$ with the least variance. When there is no ambiguity, we will denote $\theta := \theta(P)$.

If $k = 1$, then

$$U_n = \frac{1}{n} \sum_{1 \le i \le n} h(X_i).$$

Therefore, U-statistics are in fact the generalization of the empirical mean to functions of several variables. Numerous commonly used estimators belong to the class of U-statistics. For example, the unbiased empirical variance estimator of the sample $(X_1, ..., X_n)$ is given by the U-statistic with kernel $h(x_1, x_2) = (x_1 - x_2)^2/2$. Other centered moments $\mathbb{E}[(X - \mathbb{E}[X])^p]$ can be estimated by the U-statistic with $h^0(x_1, ..., x_p) = (-1)^p \prod_{i=1}^p x_i + \sum_{\ell=1}^p (-1)^{p-\ell} \binom{p}{\ell} x_1^{\ell-1} \prod_{i=1}^{p-k+1} x_i$ (symmetrize to obtain h).

Many classic statistics can also be expressed as U-statistics. For instance, Wilcoxon's one-sample statistic is yielded with the kernel $h(x_1, x_2) = \mathbb{1}(x_1 + x_2 \ge 0)$. An example where $E = \mathbb{R}^2$ is Kendall's τ which is $U_n((X_1, Y_1), ..., (X_n, Y_n))$ with the kernel $h((x_1, y_1), (x_2, y_2)) = \mathbb{1}((y_2 - y_1)(x_2 - x_1) \ge 0) - 1$.

The key strength of the theory of U-statistics is how it provides a unified approach to the properties of all these different estimators. For example, the development of the variance of a U-statistic is $\mathbb{V}[U_n] = \binom{n}{k}^{-2} \sum_{\mathbf{i} \in \mathcal{P}_k([\![n]\!])} \sum_{\mathbf{i}' \in \mathcal{P}_k([\![n]\!])} \mathrm{Cov}(h(X_{\mathbf{i}}), h(X_{\mathbf{i}'}))$. It follows that it can be expressed for any kernel h with the following formula.

Proposition 1.5.1. *The variance of a U-statistic with kernel h is*

$$\mathbb{V}[U_n] = \binom{n}{k}^{-1} \sum_{c=1}^{k} \binom{k}{c} \binom{n-k}{k-c} v_c,$$

where for all $1 \le c \le k$, $v_c = \mathrm{Cov}(h(X_{\mathbf{i}}), h(X_{\mathbf{i}'}))$ where $\mathbf{i}$ and $\mathbf{i}'$ have c common elements.

This expression of the variance of U_n is remarkable. Not only it appears that $\mathbb{V}[U_n]$ consists

in a combination of only k covariance terms, but it also gives its asymptotic behavior:

$$\mathbb{V}[U_n] = \sum_{c=1}^{k} \frac{V^{(c)}}{n^c} + o\left(\frac{1}{n^k}\right), \tag{1.5}$$

where the $V^{(c)}$ are linear combinations of the terms $(v_{c'})_{1\leq c'\leq c}$. For $c = 1$, we have $V^{(1)} = k^2 v_1$. This expansion foreshadows the Central Limit Theorem (CLT) for U-statistics and their behavior in the degenerate case, i.e. when $\sqrt{n}(U_n - \theta) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. It is apparent this case occurs when $V^{(1)} = 0$. Before commenting further, let introduce the definition of a degenerate U-statistic.

Definition 1.5.2. *Let $2 \leq d \leq k$. We say that the U-statistic U_n with kernel h is degenerate of order $d-1$ when $0 = v_1 = ... = v_{d-1} < v_d$, where for all $1 \leq c \leq k$, $v_c = \mathrm{Cov}(h(X_{\mathbf{i}}), h(X_{\mathbf{i}'}))$ where $\mathbf{i}$ and $\mathbf{i}'$ have c common elements.*

We now apply this relation to the above expression of variance. We see that a non-degenerate U-statistic has $v_1 > 0$. In this case, $\mathbb{V}[\sqrt{n}U_n] \xrightarrow[n\to\infty]{} k^2 v_1 > 0$. This situation leads to the convergence of $\sqrt{n}(U_n - \theta)$ towards a centered Gaussian distribution with variance $k^2 v_1$, as we will later state the CLT for non-degenerate U-statistic. Else, if the U-statistic is degenerate of some order $d-1$, then we see that $V^{(1)} = ... = V^{(d-1)} = 0$ and $V^{(d)} \neq 0$. In that case, we would have $\mathbb{V}[n^{\frac{c}{2}} U_n] \xrightarrow[n\to\infty]{} 0$ for all $1 \leq c \leq d-1$, so we see that $\sqrt{n}(U_n - \theta) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. On the other hand, we have $\mathbb{V}[n^{\frac{d}{2}} U_n] \xrightarrow[n\to\infty]{} V^{(d)}$. This gives a hint about what happens in the degenerate case, where the right normalization for a non-trivial limit distribution is $n^{\frac{d}{2}}$ rather than $\sqrt{n}$. We will discuss the degenerate case in a later section. Before investigating the asymptotics of U-statistics, we introduce two useful tools: the Hoeffding decomposition and martingales.

1.5.2. Hoeffding decomposition

The Hoeffding decomposition has been formalized for the first time by Hoeffding (1961), even though the leading terms of the decomposition were already used by Hoeffding (1948). The concept of this decomposition is to use orthogonal projections to break down the U-statistics into orthogonal components. These components are U-statistics themselves but they display different asymptotic behaviors and can be studied separately due to their orthogonality.

For $1 \leq c \leq k$, define the function ψ^c as

$$\psi^c : (x_1, ..., x_c) \longmapsto \mathbb{E}[h(x_1, ..., x_c, X_{c+1}, ..., X_k)].$$

By symmetry of h, for some set $\mathbf{i} \in \mathcal{P}_c([\![n]\!])$, we can denote $\psi^c(X_{\mathbf{i}}) := \psi^c(X_{i_1}, ..., X_{i_c})$ since the order of the elements of $\mathbf{i}$ does not matter, by symmetry. Set $p^0 = \theta = \mathbb{E}[h(X_1, ..., X_k)]$ and define recursively

$$p^c(X_{\mathbf{i}}) = \psi^c(X_{\mathbf{i}}) - \sum_{c'=0}^{c-1} \sum_{\mathbf{i}' \in \mathcal{P}_{c'}(\mathbf{i})} p^{c'}(X_{\mathbf{i}'}). \tag{1.6}$$

for all subsets $\mathbf{i} \in \mathcal{P}_c(\llbracket n \rrbracket)$, for all $1 \leq c \leq k$. Then $h(X_{i_1}, ..., X_{i_k})$ can be written

$$h(X_{\mathbf{i}}) = \sum_{0 \leq c \leq k} \sum_{\mathbf{i}' \in \mathcal{P}_c(\mathbf{i})} p^c(X_{\mathbf{i}'}). \tag{1.7}$$

From a geometric perspective, this decomposition of $h(X_{\mathbf{i}})$ is a decomposition on orthogonal subspaces. For all $\mathbf{i} \subseteq \llbracket n \rrbracket$, denote $L_2(\mathbf{i})$ the set of all square-integrable random variables of the form $f(X_{\mathbf{i}})$, equipped with the scalar product $\langle f, g \rangle = \mathbb{E}[f(X_{\mathbf{i}})g(X_{\mathbf{i}})]$. Let $L_2^*(\mathbf{i}) \subset L_2(\mathbf{i})$ the subspace defined as

$$L_2^*(\mathbf{i}) = \{A \in L_2(\mathbf{i}) : \mathbb{E}[A|X_{\mathbf{i}'}] = 0, \forall \mathbf{i}' \subset \mathbf{i}\}.$$

These spaces have the following properties :

1. Let $\mathbf{i}_1$ and $\mathbf{i}_2$ two disjoint subsets of $\llbracket n \rrbracket$, then $L_2^*(\mathbf{i}_1) \perp L_2^*(\mathbf{i}_2)$.
2. Let $\mathbf{i} \subseteq \llbracket n \rrbracket$, then
$$L_2(\mathbf{i}) = \bigoplus_{\mathbf{i}' \subseteq \mathbf{i}}^{\perp} L_2^*(\mathbf{i}').$$

Therefore, for a given $\mathbf{i}$, the $(L_2^*(\mathbf{i}'))_{\mathbf{i}' \subseteq \mathbf{i}}$ form a decomposition of $L_2(\mathbf{i})$ in orthogonal subspaces. Now, note that for all $\mathbf{i}' \subset \llbracket n \rrbracket$, we have $p^c(X_{\mathbf{i}'}) \in L_2^*(\mathbf{i}')$. So if $\mathbf{i}' \subseteq \mathbf{i}$, then $p^c(X_{\mathbf{i}'})$ is the projection of $h(X_{\mathbf{i}})$ on the space $L_2^*(\mathbf{i}')$. Therefore, the decomposition (1.7) is the decomposition of $h(X_{\mathbf{i}}) \in L_2(\mathbf{i})$ on these orthogonal subspaces.

Finally, notice that the U-statistic U_n can be written as

$$U_n = \theta + \sum_{c=1}^{k} \binom{k}{c} P_n^c,$$

where for $1 \leq c \leq k$, $P_n^c = \binom{n}{c}^{-1} \sum_{\mathbf{i} \in \mathcal{P}_c(\llbracket n \rrbracket)} p^c(X_{\mathbf{i}})$. This decomposition is useful as the U-statistics P_n^c hold many interesting properties. First, all the quantities $p^c(X_{\mathbf{i}})$ are orthogonal except those arising from the permutations of a given set $\mathbf{i}$, which are identical quantities. This ensures that, like the kernel decomposition, the decomposition of U-statistics is also on orthogonal subspaces. Define, for $0 \leq c \leq k$, the orthogonal spaces $L_2^c(\llbracket n \rrbracket) = \bigoplus_{\mathbf{i} \in \mathcal{P}_c(\llbracket n \rrbracket)}^{\perp} L_2^*(\mathbf{i})$, then we have $U_n \in L_2(\llbracket n \rrbracket) = \bigoplus_{c=0}^{k} L_2^c(\llbracket n \rrbracket)$. For $1 \leq c \leq k$, $P_n^c \in L_2^c(\llbracket n \rrbracket)$ and $\theta \in L_2^0(\llbracket n \rrbracket)$ so they are the orthogonal projections of U_n on $L_2^c(\llbracket n \rrbracket)$. Second, the orthogonality properties also ensures that the U-statistics P_n^c are degenerate of order $c - 1$. Therefore, the variance of each P_n^c can be reduced to only one $O(n^{-c})$ term (as $v^{c'} = 0$, for $1 \leq c' < c$ in Proposition 1.5.1).

Proposition 1.5.3. *1. For $1 \leq c_1 \neq c_2 \leq k$, we have $\mathrm{Cov}(P_n^{c_1}, P_n^{c_2}) = 0$,*

2. For $1 \leq c \leq k$, we have $\mathbb{V}[P_n^c] = \binom{n}{c}^{-1} \mathbb{V}[p^c(X_{\llbracket c \rrbracket})]$.

As a consequence, we can derive an alternative expression for the variance of U_n. Very interestingly, it is now apparent that from the expression of their respective variances, the P_n^c are $O_P(n^{\frac{c}{2}})$, $1 \leq c \leq k$ hence their different asymptotic behaviors.

Corollary 1.5.4. *1. The variance of U_n can be written $\mathbb{V}[U_n] = \sum_{c=1}^{k} \binom{k}{c}^2 \binom{n}{c}^{-1} \mathbb{V}[p^c(X_{[\![c]\!]})]$,*

2. For $1 \leq c \leq k$, $0 \leq c' \leq c-1$, we have $\mathbb{V}[n^{\frac{c'}{2}} P_n^c] \xrightarrow[n\to\infty]{} 0$.

We see that the first part of the corollary gives a different variance decomposition than Proposition 1.5.1, but it still (unsurprisingly) implies (1.5). The leading terms of this decomposition has been used by Hoeffding (1948) to prove the asymptotic normality of $\sqrt{n}(U_n - \theta)$. The decomposition has also been notably useful to study degenerate U-statistics (Bretagnolle, 1983; Arcones and Gine, 1992).

1.5.3. Martingale properties

Background

Let $(X_1, ..., X_n)$ be i.i.d. random variables with mean $\mathbb{E}[X_1] = 0$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. It is known that four classical limit theorems apply to the sequence $(\bar{X}_n)_{n\geq 1}$: the weak and strong laws of large numbers (WLLN, SLLN), the Central Limit Theorem (CLT) and the law of iterated logarithm (LIL). Martingale theory has been motivated to generalize theory for sums of i.i.d. variables to a larger class of stochastic processes that are "constant in average".

Definition 1.5.5. *Let $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$ be a filtration, i.e. a sequence of σ-fields $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$ such that for all $n \geq 1$, $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, and $M = (M_n)_{n\geq 1}$ a sequence of integrable random variables adapted to $\mathcal{F}$. $(M_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale if and only if for all $n \geq 1$, $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$.*

See how $(n\bar{X}_n)_{n\geq 1}$ can also be written as a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 1}$ where $\mathcal{F}_n = \sigma(X_1, ..., X_n)$. The WLLN for i.i.d. variables can be considered as a consequence of martingale inequalities (except if it is not required that $\mathbb{E}[X_1^2] < \infty$, but in this case, the independence criteria is not necessary and the proof can be extended to martingales). The SLLN for i.i.d. variables is a consequence of the martingale convergence theorem (which is why some authors refer to the latter as the martingale SLLN) and there exists a martingale version of the CLT and the LIL.

In the case of U-statistics, we are most interested in backward martingales.

Definition 1.5.6. *Let $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$ be a decreasing filtration, i.e. a sequence of σ-fields $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$, i.e. such that for all $n \geq 1$, $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n$, and $M = (M_n)_{n\geq 1}$ a sequence of integrable random variables adapted to $\mathcal{F}$. $(M_n, \mathcal{F}_n)_{n\geq 1}$ is a backward martingale if and only if for all $n \geq 1$, $\mathbb{E}[M_n \mid \mathcal{F}_{n+1}] = M_{n+1}$.*

Backward martingales can be considered as martingales in reversed time, e.g. $(M_n, \mathcal{F}_n)_{1\leq n\leq N}$ is a martingale if and only if $(M_{N-n+1}, \mathcal{F}_{N-n+1})_{1\leq n\leq N}$ is a backward martingale. For this reason, many martingale limit theorems admit a backward martingale version. These versions also generalize the corresponding limit theorems for i.i.d. variables as $(\bar{X}_n)_{n\geq 1}$ can be written as a backward martingale, with respect to the decreasing filtration defined by $\mathcal{F}_n = \sigma(\bar{X}_n, \bar{X}_{n+1}, ...)$. Actually, the U-statistics U_n have a "forward" martingale structure since the Hoeffding projections $(\binom{n}{c}P_n^c)_{n\geq 1}$ are "forward" martingales, but more interestingly, the $(U_n)_{n\geq 1}$ themselves are backward martingales. For this reason, it is more direct to investigate U_n under the light of backward martingales. We will mainly be interested in the backward martingale convergence theorem and the backward martingale CLT.

Theorem 1.5.7. *Let $(M_n, \mathcal{F}_n)_{n\geq 1}$ be a backward martingale. Then, $(M_n)_{n\geq 1}$ is uniformly integrable, and, denoting $M_\infty = \mathbb{E}[M_1 \mid \mathcal{F}_\infty]$ where $\mathcal{F}_\infty = \bigcap_{n=1}^\infty \mathcal{F}_n$, we have*

$$M_n \xrightarrow[n\to\infty]{a.s., L_1} M_\infty.$$

Furthermore, if $(M_n)_{n\geq 1}$ is square-integrable, then $M_n \xrightarrow[n\to\infty]{L_2} M_\infty$.

Theorem 1.5.8 (Eagleson and Weber, 1978)**.** *Let $(M_n, \mathcal{F}_n)_{n\geq 1}$ be a square-integrable reverse martingale, V a $\mathcal{F}$-measurable, a.s. finite, positive random variable. Denote $M_\infty := \mathbb{E}[M_1 \mid \mathcal{F}_\infty]$ where $\mathcal{F}_\infty := \bigcap_{n=1}^\infty \mathcal{F}_n$. Set $Z_{nk} := \sqrt{n}(M_k - M_{k+1})$. If:*

1. *$\sum_{k=n}^\infty \mathbb{E}[Z_{nk}^2 \mid \mathcal{F}_{k+1}] \xrightarrow[n\to\infty]{\mathbb{P}} V$,*
2. *for all $\epsilon > 0$, $\sum_{k=n}^\infty \mathbb{E}[Z_{nk}^2 \mathbb{1}_{\{|Z_{nk}|>\epsilon\}} \mid \mathcal{F}_{k+1}] \xrightarrow[n\to\infty]{\mathbb{P}} 0$,*

then $\sum_{k=n}^\infty Z_{nk} = \sqrt{n}(M_n - M_\infty) \xrightarrow[n\to\infty]{\mathcal{D}} W$, where W is a random variable with characteristic function $\phi(t) = \mathbb{E}[\exp(-\frac{1}{2}t^2 V)]$.

The a.s. limit of M_n is $M_\infty = \mathbb{E}[M_1 \mid \mathcal{F}_\infty]$, a $\mathcal{F}_\infty$-measurable random variable. In the CLT, the limiting distribution of $\sqrt{n}(M_n - M_\infty)$ is a mixture of centered Gaussians with variances determined by the values taken by V, also a $\mathcal{F}_\infty$-measurable random variable. Both the SLLN and the CLT for sums of i.i.d. random variables are straightforward corollaries of these theorems where M_∞ and V are constants. Indeed, when $M_n = \bar{X}_n$ and $\mathcal{F}_n = \sigma(\bar{X}_n, \bar{X}_{n+1}, ...)$, then $\mathcal{F}_\infty$ is trivial. It can be hinted how dropping the independence assumption affects the results.

Example 1. Let $(Y_n)_{n\geq 1}$ be a sequence of random variables such that $Y_i := X_i + Z$, where the $(X_n)_{n\geq 1}$ are i.i.d. random variables and Z is an integrable random variable. Then the $(Y_n)_{n\geq 1}$ are not independent, but $(\bar{Y}_n)_{n\geq 1}$ is still a backward martingale with respect to $\mathcal{G}_n := \sigma(\bar{Y}_n, \bar{Y}_{n+1}, ...) = \sigma(\bar{X}_n + Z, \bar{X}_{n+1} + Z, ...)$. Here, $\mathcal{G}_\infty = \sigma(Z)$, so in some way, $\mathcal{G}_\infty$ contains information on the variability of the $(Y_n)_{n\geq 1}$ that cannot be averaged out.

Link with U-statistics

Here we showcase the martingale structure of U_n.

Proposition 1.5.9. *Let $(\mathcal{F}_n)_{n\geq 1}$ be the filtration defined by $\mathcal{F}_n = \sigma(X_1, ..., X_n)$. If $\mathbb{E}[|h(X_1, ..., X_k)|] < \infty$, then for all $1 \leq c \leq k$, $(\binom{n}{c} P_n^c, \mathcal{F}_n)_{n\geq 1}$ is a martingale.*

Proposition 1.5.9 applies to the U-statistics P_n^c. Earlier, it has been mentioned that in the case of i.i.d. variables with mean 0, $(n\bar{X}_n)_{n\geq 1}$ is a martingale. This result is included in Proposition 1.5.9. If $k = 1$ and h is the identity, then $\theta = 0$ so that $(n\bar{X}_n)_{n\geq 1} = (n(U_n - \theta))_{n\geq 1} = (nP_n^1)_{n\geq 1}$ is a martingale. One would be tempted to assume that $(\binom{n}{k} U_n)_{n\geq 1}$ is also a martingale. However, this is wrong in general as $n(U_n - \theta)$ is a linear combination of the $(\binom{n}{c} P_n^c)_{1\leq c\leq k}$ where the coefficients depend on n. For this reason, the application of (forward) martingale results to U-statistics require some additional work. On the other hand, we have the following property directly on U_n.

Proposition 1.5.10. *Let $(\mathcal{F}_n)_{n\geq 1}$ be the decreasing filtration defined by $\mathcal{F}_n = \sigma(U_n, U_{n+1}, ...)$. If $\mathbb{E}[|h(X_1, ..., X_k)|] < \infty$, then for all $1 \leq c \leq k$, $(U_n, \mathcal{F}_n)_{n\geq 1}$ is a backward martingale.*

The backward martingale structure of U_n is more straightforward. Hence, backward martingale inequalities and limit theorems can extend to U-statistics of i.i.d. variables.

1.5.4. Asymptotic behavior of U-statistics

We have mentioned the three classic limit theorems for sums of i.i.d. random variables (WLLN/SLLN, CLT, LIL). Obviously, for U-statistics with degree $k = 1$, all these asymptotic results apply. In general, we have to exploit the Hoeffding decomposition and/or the forward/backward martingale structure of the U-statistics to obtain an equivalent for each. In the non-degenerate cases, U-statistics hold similar properties to sums of i.i.d. variables due to the dominant term P_n^1 of the Hoeffding decomposition being a sum of i.i.d. variable itself. If U_n is degenerate of order $d - 1$, $2 \leq d \leq k$, the dominant term is P_n^d, which is not a sum of i.i.d. variables but still a forward and backward martingale.

The SLLN for U-statistics of i.i.d. variables is given by the following theorem.

Theorem 1.5.11 (Hoeffding, 1961)**.** *If $\mathbb{E}[|h(X_1, ..., X_k)|] < \infty$, then $U_n - \theta \xrightarrow[n\to\infty]{a.s., L_1} 0$.*

Although Hoeffding (1961) proved this result using both the Hoeffding decomposition and the "forward" martingale structure of the U-statistic, a more concise proof would be to simply apply Theorem 1.5.7 using the backward martingale structure of U_n (Berk, 1966).

This SLLN obviously implies the WLLN for U-statistics. The next theorem is the CLT for non-degenerate U-statistics.

Theorem 1.5.12 (Hoeffding, 1948)**.** *If* $\mathbb{E}[h(X_1,...,X_k)^2] < \infty$ *and* $v_1 > 0$*, then* $\sqrt{n}(U_n - \theta) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, k^2 v_1)$.

Again, several strategies of proof exist. Hoeffding (1948) used the Hoeffding decomposition and one could have applied the backward martingale CLT. As a remark, note that if U_n is degenerate, then by extension, the result would be still valid with $v_1 = 0$, but it simply yields $\sqrt{n}(U_n - \theta) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. The distribution of $n^{\frac{d}{2}}(U_n - \theta)$ when U_n is degenerate of order $d-1$, $2 \le d \le k$, has been investigated by Rubin and Vitale (1980).

Theorem 1.5.13 (Rubin and Vitale, 1980)**.** *If* $\mathbb{E}[h(X_1,...,X_k)^2] < \infty$ *and* $0 = v_1 = ... = v_{d-1} < v_d$ *for some* $2 \le d \le k$*, then* $n^{\frac{d}{2}}(U_n - \theta)$ *and* $n^{\frac{d}{2}}\binom{k}{d}P_n^d$ *both converge in distribution to the same limit. This limit is*

$$\binom{k}{d} \sum_{\mathbf{j} \in \mathbb{N}^d} \langle p^d, e_{j_1} \otimes ... \otimes e_{j_d} \rangle \prod_{\ell=1}^{\infty} H_{\kappa_\ell(\mathbf{j})}(W_\ell),$$

where

- $(e_i)_{i\ge1}$ *is an orthonormal basis of the Hilbert space* $L_2(\{1\})$,
- $e_{j_1} \otimes ... \otimes e_{j_d}$ *is the function defined by* $(e_{j_1} \otimes ... \otimes e_{j_d})(x_1,...,x_d) = e_{j_1}(x_1)...e_{j_d}(x_d)$,
- p^d *is the kernel of* P_n^d*, i.e. the function defined by equation* (1.6),
- $(W_i)_{i\ge1}$ *are independent standard normal variables,*
- H_u *is the* u*-th Hermite polynomial, for* $u \ge 0$,
- $\kappa_\ell(\mathbf{j})$ *is the number of times* ℓ *appears as an element in the* d*-tuple* $\mathbf{j}$.

Albeit the expression of the limit is complex and does not always admit a close form, this result is still of great interest for degenerate U-statistics. For degenerate U-statistics of order 1, it has been known that the limit distribution is that of an infinite sum of independent χ^2 distributions $\sum_{\ell=1}^{\infty} \lambda_\ell (W_\ell^2 - 1)$, where the λ_ℓ are constants (Gregory, 1977; Eagleson, 1979). Theorem 1.5.13 extends to degenerate U-statistics of order $d - 1 > 1$, telling us that the limit distribution is an (infinite) linear combination of polynomials of the W_ℓ of degree d, i.e. it is an (infinite) sum of monomials $\prod_{c=1}^{d} W_c^{a_c}$ where $a_c \in \{0, 1, ..., d\}$ for all $1 \le c \le d$ and $\sum_{c=1}^{d} a_c \le d$.

Not always would one need the expression of this limit, as it is possible to retrieve its cumulative distribution function by bootstrap. Although the bootstrap of U_n when U_n is degenerate is bound to fail (Bretagnolle, 1983), Theorem 1.5.13 states that the target distribution is also the limit distribution of $n^{\frac{d}{2}}\binom{k}{d}P_n^d$. This opens up the possibility to bootstrap $U_n - \sum_{c=1}^{d-1}\binom{k}{c}P_n^c$ which solves the issue (see Section 5.2.1).

Compared to the LLN and the CLT, the LIL has more limited applications in statistics. A LIL for U-statistics can still be written.

Theorem 1.5.14. *If* $\mathbb{E}[h(X_1,...,X_k)^2] < \infty$ *and* $v_1 > 0$, *then*

$$\limsup_{n\to\infty} \frac{n(U_n - \theta)}{\sqrt{k^2 v_1 n \log\log n}} \overset{a.s.}{=} \sqrt{2}. \tag{1.8}$$

This result is obtained applying the classic LIL to the dominant term of the Hoeffding decomposition. A similar result for degenerate U-statistics can be deduced from Dehling et al. (1984). The rate of convergence is $\frac{n}{(n\log\log n)^{\frac{d}{2}}}$ where $d-1$ is the order of degeneracy.

1.5.5. Dependent data

The results enunciated in the previous section are now well known, but they do not easily transpose to dependent data. In fact, the dependency structure of the $(X_1, X_2, ...)$ determines that of the associated U-statistics. Literature on the asymptotics of U-statistics are closely related to sums of dependent data. Classical literature on U-statistics assumes the data $(X_1, X_2, ...)$ to be i.i.d. variables. In this case, the $h(X_{i_1},...,X_{i_k})$ are dependent. Denote $Y_{i_1...i_k} = h(X_{i_1},...,X_{i_k})$. Then we easily see that $Y = (Y_{\mathbf{i}})_{\mathbf{i}\in\mathbb{N}^k}$ is in fact jointly exchangeable, so that a U-statistic of i.i.d. variables is a sum of jointly exchangeable variables.

We have described two tools to prove limit theorems of U-statistics of i.i.d. variables: martingales and the Hoeffding decomposition. Most developments around U-statistics of dependent data were built upon these two tools. Even if the $(X_1, X_2, ...)$ are not independent, one could hope to obtain similar results if U_n retains its martingale structure or if an Hoeffding-type decomposition can be derived. For example, investigating U-statistics of ergodic stationary processes, Yoshihara (1976) and Denker and Keller (1983) used the Hoeffding decomposition alongside with coupling techniques to derive a series of CLTs with different conditions on the sequences and Aaronson et al. (1996) discussed the conditions for a LLN.

U-statistics of jointly exchangeable arrays are still sums of jointly exchangeable variables, so it is not surprising that results for U-statistics of i.i.d. data translate well to them. Effectively, Silverman (1976) proved a limit theorem for jointly exchangeable arrays using Hoeffding-type arguments and Eagleson and Weber (1978) revisited it with martingales. Many other asymptotic results for U-statistics of exchangeable sequences were derived afterwards, such as a Berry-Esseen bound (van Zwet, 1984) and a law of iterated logarithm (Scott and Huggins, 1985). Eagleson and Weber (1978) also proved a SLLN for these sums. More generally, the SLLN of Kallenberg (1999) applies to sums of π-exchangeable variables, i.e. arrays X indexed by a k-tuples such that there exists a partition π of $[\![k]\!]$ and a vector of permutations $(\sigma_I)_{I\in\pi}$ such that

$$X \overset{\mathcal{D}}{=} \left(X_{\sigma_{I_1}(i_1)...\sigma_{I_k}(i_k)}\right)_{\mathbf{i}\in\mathbb{N}^k},$$

where for $\ell \in [\![k]\!]$, I_ℓ denotes the subset $I \in \pi$ containing ℓ. This family of variables includes both U-statistics of jointly exchangeable arrays and of separately exchangeable arrays. The case of finite exchangeability has also been considered, notably for the problem of sampling in a finite population, for example by Nandi and Sen (1963) and Zhao and Chen (1990).

Concurrently, many results have been obtained using Stein's method (Stein, 1972). Stein's method uses the properties of the solutions of a differential equation (Stein's equation) to bound the approximation error of a random variable by a standard normal random variable, most often in Wasserstein distance. For a sequence of random variables, the goal is to find a bound that decreases asymptotically to 0, which guarantees the convergence in distribution to a standard normal variable. Stein's method comes up with the advantage that it automatically provides a Berry-Esseen bound, informing on the rate of convergence to the limit distribution. In general, one makes use of the structure of the investigated random variable to find this bound, so this is very problem-dependent and some non trivial work is required to adapt Stein's method. In relation to the U-statistics considered in this thesis, Austern and Orbanz (2022) applied Stein's method to prove the asymptotic normality of U-statistics of exchangeable variables, therefore obtaining a Berry-Esseen bound. Other examples of application to U-statistics of independent or dependent variables can be found in the literature (Barbour and Eagleson, 1985; Maesono, 1991; Chatterjee, 2008; Reinert and Röllin, 2010; Reitzner and Schulte, 2013).

1.6. Contributions

In the first sections, I have presented the context in which network analysis intervenes in ecological studies. The contributions of this thesis exclusively focus on the development of a new methodology for network analysis in this particular context. In the previous sections, I have introduced exchangeable random graph models and U-statistics. Basically, these are the class of network models and the tool for statistical inference that will be used. In this section, I will first clearly define the framework in which these objects are used. Next, I will compile the diverse contributions of this thesis and link them together. Finally, I give the general organization of this thesis, showing how the different contributions are spread out in the different chapters of this thesis.

1.6.1. General framework

Bipartite exchangeable network models

We have previously seen how graphon theory can be used to represent dissociated exchangeable graph models. Graphon models are latent space models with a set of i.i.d. latent variables. We generalize it to bipartite and weighted graphs:

$$\begin{aligned} \xi_i &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq i \leq m, \\ \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq j \leq n, \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{W}(\xi_i, \eta_j), && \forall 1 \leq i \leq m, 1 \leq j \leq n, \end{aligned}$$

where $\mathcal{W} : [0,1]^2 \to \Pi(E)$ and $\Pi(E)$ is the set of all probability distributions on E.

This representation is usually called a colored (bipartite) graphon. The graphon is classically a symmetric function $w : [0,1]^2 \to [0,1]$. A bipartite graphon drops the symmetry assumption, since the sets of random variables $(\xi_i)_{1 \leq i \leq m}$ and $(\eta_j)_{1 \leq j \leq n}$ represent the two groups of nodes of the bipartite graph. The colored graphon extends the graphon model to weighted graphs, where the conditional distribution of Y_{ij} now takes values in E instead of $\{0,1\}$. In this thesis, a simpler family of colored graphon models will be considered. These colored graphons can be written as $\mathcal{W}(\cdot,\cdot) = \mathcal{L}(w(\cdot,\cdot))$, where $w : [0,1]^2 \to \mathbb{R}$ and $(\mathcal{L}(\mu))_{\mu \in \mathbb{R}}$ is a family of probability distribution with a unique parameter.

$$\begin{aligned} \xi_i &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq i \leq m, \\ \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq j \leq n, \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{L}(w(\xi_i, \eta_j)), && \forall 1 \leq i \leq m, 1 \leq j \leq n, \end{aligned}$$

The classic graphon model fits in this family when w only takes values in $[0,1]$ and setting $\mathcal{L} := \mathcal{B}$. By analogy with the classic definition, we will simply refer to w as the graphon, despite dropping the symmetry assumption and being real-valued and $\mathcal{L}$ is the link distribution.

Latent block model The Latent Block Model (LBM) (Govaert and Nadif, 2003) is the bipartite version of the SBM. The LBM assumes that the two sets of nodes of the bipartite graphs can be separately partitioned into groups. The distribution of the interaction between two nodes is determined by the involved groups. Usually, this distribution is parametrized by a unique parameter depending on the groups. The hierarchical form of the LBM is given by

$$\begin{aligned} Z_i &\overset{iid}{\sim} \mathcal{M}(1; \boldsymbol{\alpha}), && \forall 1 \leq i \leq m, \\ W_j &\overset{iid}{\sim} \mathcal{M}(1; \boldsymbol{\beta}), && \forall 1 \leq j \leq n, \\ Y_{ij} \mid Z_i = k, W_j = \ell &\sim \mathcal{L}(\pi_{k\ell}), && \forall 1 \leq i \leq m, 1 \leq j \leq n, \end{aligned}$$

where α and β are probability vectors of sizes K_1 and K_2, the respective numbers of groups for nodes of type 1 and type 2, $(\mathcal{L}(\theta))_{\theta\in\Theta}$ is a family of probability distribution and $\boldsymbol{\pi} \in \mathcal{M}_{K_1,K_2}(\Theta)$ is a matrix of $K_1 \times K_2$ parameters (Fig. 1.12).

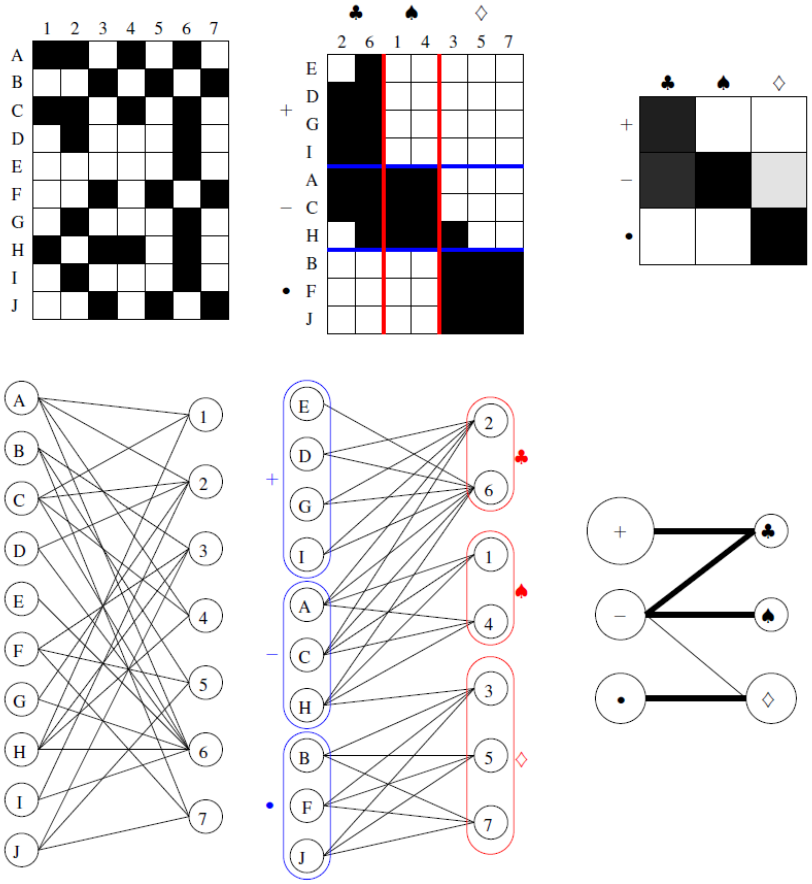

Figure 1.12 – The latent block model (LBM). Left: a bipartite network and its adjacency matrix. Middle: the ordered network and adjacency matrix according to the membership of the nodes to groups. Right: a "summary" network, representing the probability matrix $\boldsymbol{\pi}$ as a network between groups. Figure taken from Brault (2014).

The LBM can be written as a block-constant graphon, as

$$w(\xi_i, \eta_j) = \sum_{k,\ell} \pi_{k\ell} \mathbb{1}\{s(\xi_i) = k\} \mathbb{1}\{t(\eta_j) = \ell\},$$

with

$$s(\xi_i) = 1 + \sum_{k=1}^{K} \mathbb{1}\left\{\xi_i > \sum_{k'=1}^{k} \alpha_{k'}\right\}, \qquad \text{and} \qquad t(\eta_j) = 1 + \sum_{\ell=1}^{L} \mathbb{1}\left\{\eta_j > \sum_{\ell'=1}^{\ell} \beta_{\ell'}\right\}.$$

Sometimes, to specify the distribution $\mathcal{L}$ of the model, we might write $\mathcal{L}$-LBM, e.g. Bernoulli-LBM or Poisson-LBM.

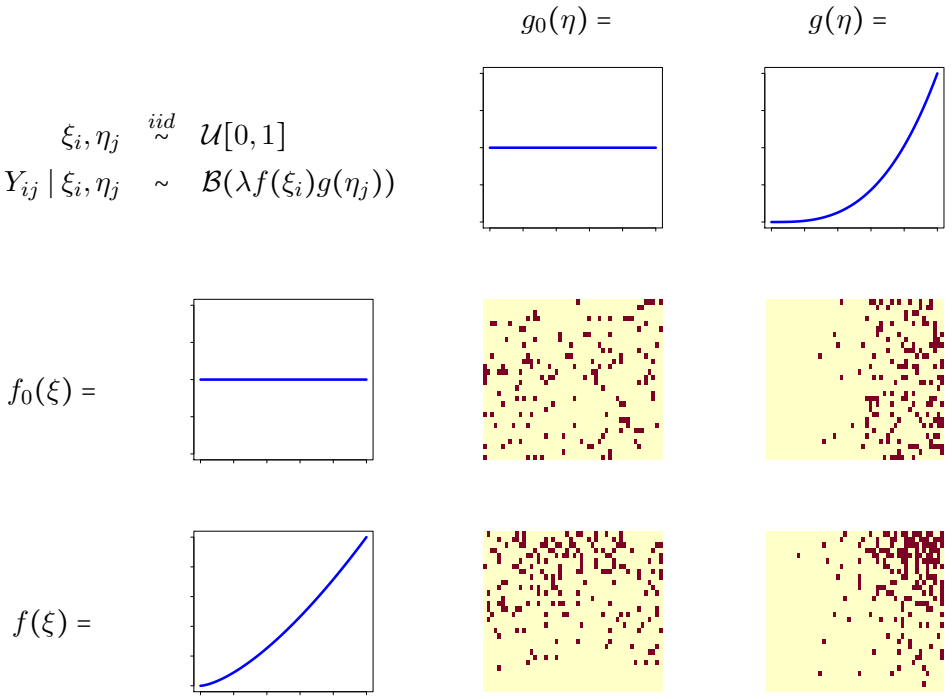

Figure 1.13 – The bipartite expected degree distribution model (BEDD). The figure represents several (reordered) adjacency matrices generated by binary BEDDs with constant (f_0, g_0) or some other (f, g) degree distributions.

Bipartite expected degree distribution model The bipartite expected degree distribution model (BEDD) (Ouadah et al., 2022) is the bipartite version of the EDD. Instead of one degree distribution, the BEDD considers two degree distributions, one for each set of nodes. It can be formulated as follows

$$\begin{aligned} &\xi_1, ..., \xi_m \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \le i \le m, \\ &\eta_1, ..., \eta_n \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \le j \le n, \\ &Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{L}(\lambda f(\xi_i) g(\eta_j)), && \forall 1 \le i \le m, 1 \le j \le n, \end{aligned}$$

where $(\mathcal{L}(\mu))_{\mu \in \mathbb{R}}$ is a family of probability distributions parametrized by their mean μ, $\lambda \in \mathbb{R}$ is the density of the graph, f and g are functions $[0,1] \to \mathbb{R}$ such that $\int f = \int g = 1$.

In this formulation, the distributions of degrees are characterized by the functions f and g (Fig. 1.13). More specifically, f and g can be viewed as the inverse of the cumulative distribution function of the renormalized degree distribution, since $\mathbb{E}[n^{-1} \sum_{j=1}^{n} Y_{ij} \mid \xi_i] = \lambda f(\xi_i)$ and $\mathbb{E}[m^{-1} \sum_{i=1}^{m} Y_{ij} \mid \eta_j] = \lambda g(\eta_j)$. The BEDD is a graphon model where the graphon has a product form $w(\xi, \eta) = \lambda f(\xi) g(\eta)$. Sometimes, to specify the distribution $\mathcal{L}$ of the model, we might write $\mathcal{L}$-BEDD, e.g. Bernoulli-BEDD or Poisson-BEDD.

Asymptotics

An asymptotic framework defines what happens when more data become available. There are two types of asymptotics for networks. In the first one, new data brings more networks that are considered as replicates from a unique model. In the other one, new data brings more nodes and edges to a unique yet incomplete network, which is a subsample of the complete, possibly infinite, network drawn from the model of interest.

Interaction data from ecosystems are collected by ecologists at different places, at different times, under different conditions. Therefore, as opposed to many other types of network data, ecological networks are by essence the result of an aggregation of individually sampled interaction data. Because ecological interaction sampling is costly and most of the time incomplete, one wishes to use as much data as possible to construct the networks. Even so, networks obtained by aggregation are likely to be incomplete. The sampling effort needed to obtain a complete network is not easily estimated. For these reasons, it is less common to use repeated measurements of networks that can be considered as replicates drawn from the same model. Even if replicates are available, they are often aggregated to grow a more complete network. Therefore, it is reasonable to work with an asymptotic framework where the growing quantity is the number of nodes, instead of a growing quantity of replicates.

This is especially well adapted to the exchangeability assumption, as each network can be represented by a unique model, independent of the size. Indeed the extendibility property means that every observed adjacency matrix is a submatrix of an infinite RCE matrix. Networks can be studied in a space of models. This can be considered as a kind of "embedding". Embedding a network means mapping it to some lower dimension space. In the framework of this thesis, each network is mapped to the space of all RCE network models, possibly restricted to some family of models. Therefore, the networks can be investigated in the space of models through their representative using probabilistic and statistical tools. This makes sense for ecological data since, to compare two ecosystems, one usually likes to compare the two networks associated to the two ecosystems. A model-based analysis would compare the two models fitted to these networks, which means the representatives of the two networks in the space of models. If the space is restricted to some family of parametric models, then this can be done by comparing their parameters. Since graphons characterize exchangeable graph models, they can also be used to construct a comparison metric.

In this thesis, an observed adjacency matrix with growing dimensions $m_N \times n_N$ are considered, with $m_N \to \infty$ and $n_N \to \infty$. In most results, it is assumed that the numbers of rows and columns grow at the same speed, with $N = m_N + n_N$ and $m_N/N \to \rho \in]0, 1[$, but it is actually easy to generalize the results to other behaviours of m_N and n_N, as shown in Chapter 4.

One advantage of working with exchangeable models and this asymptotic framework is how easy it is to jointly study networks of different sizes. Whereas statistical estimation of parameters is one technique to investigate the models, many other tools can be used. As models are characterized by their probability distributions, many approaches using tools from parametric and nonparametric statistics, probability theory and even information theory can be combined to analyze networks. The approach consists in identifying quantities of interests which can be estimated with U-statistics.

U-statistics on bipartite networks

U-statistics are defined as the average of a function of a subsample of the data. When the data is a bipartite adjacency matrix of size $m \times n$, the kernel function $h : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ is a function of a submatrix of size $p \times q$, with $1 \leq p \leq m$ and $1 \leq q \leq n$. The symmetry assumption on the kernel translates to: for all $(\sigma_1, \sigma_2) \in \mathbb{S}_p \times \mathbb{S}_q$,

$$h(Y_{(i_{\sigma_1(1)},\dots,i_{\sigma_1(p)};j_{\sigma_2(1)},\dots,j_{\sigma_2(q)})}) = h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}),$$

where $Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}$ is the $p \times q$ submatrix consisting of the rows and columns of Y indexed by $i_1, \dots, i_p$ and $j_1, \dots, j_q$ respectively.

With this symmetry assumption, since the order of the elements of $\mathbf{i} = \{i_1, \dots, i_p\}$ and $\mathbf{j} = \{j_1, \dots, j_q\}$ does not matter, we use the notation $h_{\mathbf{i},\mathbf{j}} := h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)})$. Then the associated U-statistic is

$$U_{m,n} = \binom{m}{p}^{-1} \binom{n}{q}^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j} \in \mathcal{P}_q([\![n]\!])}} h_{\mathbf{i},\mathbf{j}}. \tag{1.9}$$

Note that the assumption on the symmetry of h can be made without loss of generality. Indeed, if $h^0 : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ is not symmetric, then $h : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ defined by

$$h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}) = (p!q!)^{-1} \sum_{(\sigma_1,\sigma_2) \in \mathbb{S}_p \times \mathbb{S}_q} h^0(Y_{(i_{\sigma_1(1)},\dots,i_{\sigma_1(p)};j_{\sigma_2(1)},\dots,j_{\sigma_2(q)})}) \tag{1.10}$$

has the correct symmetry property and the same average as h^0.

The properties of $U_{m,n}$ depend on the dependency structure of the summed elements. Let $X_{(i_1,\dots,i_p;j_1,\dots,j_q)} := h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)})$, so that $U_{m,n}(Y)$ is the average of the first entries of the infinite array X. In our framework, Y is an RCE matrix. As a consequence, X is a π-exchangeable array such that for any two permutations σ_1 and σ_2 of $\mathbb{S}_\infty$, we have

$$X \overset{\mathcal{D}}{=} (X_{(\sigma_1(i_1),\dots,\sigma_1(i_p);\sigma_2(j_1),\dots,\sigma_2(j_q))})_{\substack{1 \leq i_1 \neq \dots \neq i_p \leq \infty \\ 1 \leq j_1 \neq \dots \neq j_q \leq \infty}}.$$

This also means

$$\mathbb{E}[U_{m,n}] = \mathbb{E}[X_{(1,\dots,p;1,\dots,q)}] = \mathbb{E}[h_{[\![p]\!],[\![q]\!]}].$$

Therefore, $U_{m,n}$ is an unbiased estimator of $X_{(1,\dots,p;1,\dots,q)} = h_{[\![p]\!],[\![q]\!]}$. The methodology developed makes use of U-statistics to perform inference on the RCE network model. Thus, it does not require a parametric model. It applies to all quantities that can be estimated by a function of a subnetwork.

General method in a nutshell

Let θ be a quantity of interest that needs to be estimated on an observed network. Usually, θ is a function of the parameters of a random network model. The goal is to find an estimator $\widehat{\theta}_N$, as well as a confidence interval for θ. The idea is to use estimators that are either U-statistics, or functions of U-statistics, i.e. of the form $\widehat{\theta}_N := U_N$ or $\widehat{\theta}_N := g(U_N^{h_1}, \dots, U_N^{h_D})$, where g is a differentiable function, $(h_1, \dots, h_D)$ a vector of kernel functions, potentially of different sizes, and U_N^h the U-statistic with kernel h.

Then, if the network model is RCE, we hope to be able to use some theoretical results to identify the limit distribution for $\widehat{\theta}_N$ when $N \to \infty$. These theoretical results are often of the form

$$\sqrt{\frac{\gamma(N)}{v}}\left(\widehat{\theta}_N - \theta\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{Q},$$

where $\gamma(N)$ is some function of N such that $\gamma(N) \to \infty$, v is the asymptotic variance and $\mathcal{Q}$ is some distribution with variance 1. $\gamma(N)$ and v are related to the variance of $\widehat{\theta}_N$ by the relation $\mathbb{V}[\theta_N] = v/\gamma(N) + o(\gamma(N)^{-1})$. However, v usually must be estimated from the data. If $\widehat{v}_N$ is a consistent estimator for v, then Slutsky's theorem yields

$$\sqrt{\frac{\gamma(N)}{\widehat{v}_N}}\left(\widehat{\theta}_N - \theta\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{Q}. \tag{1.11}$$

For $x \in]0,1[$, denote q_x the quantile of order x for $\mathcal{Q}$, i.e. if W is a random variable with distribution $\mathcal{Q}$, then $\mathbb{P}(W \le q_x) = x$. Asymptotic confidence intervals for θ can be built from (1.11). For example,

$$I_N(\alpha) = \left[\widehat{\theta}_N - q_{1-\alpha/2}\sqrt{\frac{\widehat{v}_N}{\gamma(N)}}, \widehat{\theta}_N + q_{\alpha/2}\sqrt{\frac{\widehat{v}_N}{\gamma(N)}}\right]$$

is an asymptotic confidence interval at level α for θ, i.e. $\mathbb{P}(\theta \in I_N(\alpha)) \xrightarrow[N\to\infty]{} 1-\alpha$.

With confidence intervals as such, we are able to obtain statistical guarantees for $\widehat{\theta}_N$. Therefore, we are able to perform most statistical inference tasks on network models, including estimation of parameters, hypothesis testing and comparison of networks.

1.6.2. Results

The list of ingredients needed to apply this method to observed networks consists of:

1. an RCE network model, from which the quantity of interest can be identified,
2. an estimator for the quantity of interest, which can be written as a function of U-statistics,
3. a weak convergence result on this function of U-statistics,
4. an estimator of the asymptotic variance of this function of U-statistics,
5. a way to efficiently compute the two estimators of the quantity of interest and the asymptotic variance.

The theoretical part of this thesis mainly deals with establishing limit theorems for functions of U-statistics and estimators for their asymptotic variance. Through many examples illustrating these theoretical results, I will show how to choose suitable network models and estimators according to the network analysis question. Finally, the computation cost of these estimators is usually problem-dependent. Although there is a high computational cost in general for these estimators, in many examples, I give some insights about how one might mitigate this problem, for example using matrix products. The code is available on an online repository and implements some of these examples, using efficient computation.

Recoverability of BEDD models

The first result of this thesis is a bit different from the rest. It does not directly deal with U-statistics, but it is one of the motivating reasons for using them in my methodology. It establishes the recoverability of the BEDD model by the joint distribution of what we call a *quadruplet*, that is, a submatrix of size 2×2.

Before anything, note that there is a problem of identifiabilty which is inherent to exchangeable network models, but this can be easily circumvented. Indeed, any RCE network model can be written as a graphon model. However, the graphon model is not technically identifiable. Let w be a graphon. For any pair $\pi = (\pi_1, \pi_2) \in \Pi([0,1])^2$ where $\Pi([0,1])$ is the set of the measure-preserving transformations on $[0,1]$, denote $w_\pi(\xi, \eta) := w(\pi_1(\xi), \pi_2(\eta))$.

The $G(n, w)$ model yields the same distribution as the $G(n, w_\pi)$ model where $w_\pi(\xi, \eta) = w(\pi_1(\xi), \pi_2(\eta))$. Conversely, if two graphon models $G(n, w)$ and $G(n, w')$ yield the same distribution, then necessarily, there exists $\pi \in \Pi([0,1])^2$ and $\pi' \in \Pi([0,1])^2$ such that $w'_{\pi'} = w_\pi$ (Diaconis and Janson, 2008). Therefore, the family of graphons

$$\left\{w' : w'_{\pi'} = w_\pi, \pi \in \Pi([0,1])^2, \pi' \in \Pi([0,1])^2, w' : [0,1]^2 \to [0,1]\right\}$$

is a class of equivalence for w, and we say that the graphon model is identifiable up to its class of equivalence.

In the BEDD, the same phenomenon occurs. The functions f and g are identifiable up to their class of equivalence, generated by all the measure-preserving transformations of $[0,1]$. As a side note, it is possible to set constraints on the functions f and g in the definition of the BEDD such that there is only a unique eligible f and g from their respective class of equivalence, for example, see Yang et al. (2014) for a similar argument on graphons.

Now, the following theorem states that, f and g are uniquely characterized, up to their class of equivalence, by some families $(F_k)_{k\geq 1}$ and $(G_k)_{k\geq 1}$.

Theorem 2.3.2 (Chp. 2, Le Minh, 2023)**.** *Let $\Theta = (\lambda, f, g)$ be BEDD parameters and $Y \sim BEDD(\Theta)$. The distribution of Y is uniquely determined by λ, $(F_k)_{k\geq 1}$ and $(G_k)_{k\geq 1}$, where $F_k := \int f^k$ and $G_k := \int g^k$ for all $k \geq 1$.*

Furthermore, to show the recoverability of the BEDD model by a quadruplet, it is enough to show that the two distinct sets of parameters (which are not equivalent) for the BEDD lead to necessarily distinct distributions for a quadruplet. With the previous theorem, we only need to prove it for distinct λ, $(F_k)_{k\geq 1}$ and $(G_k)_{k\geq 1}$. We have managed to show this result for a class of BEDD models in which the distribution $\mathcal{L}$ verifies an assumption.

Theorem 2.3.4 (Chp. 2, Le Minh, 2023)**.** *Suppose that for the family of distributions $\mathcal{L}(\mu)$, there exists a sequence of functions $(\Psi_k)_{k\geq 1}$ such that if a random variable $X \sim \mathcal{L}(\mu)$, then for every $k \geq 1$,*

$$\mathbb{E}[\Psi_k(X)] = \mu^k.$$

Then, in this case, for all $k \in \mathbb{N}$, F_k and G_k are uniquely determined by the joint distribution of a quadruplet.

In reality, this result, even though it has only been proved later in the thesis, was intuited before starting to devise my method. It had been the motivation behind my use of U-statistics: since a quadruplet is enough to contain all the information of the model, there is hope that a U-statistic averaging a function of quadruplets is enough to estimate anything of interest. This is also the reason why my first article is mainly focused on U-statistics of quadruplets, instead of submatrices of any size $p \times q$.

Hoeffding-type decompositions

Decomposition on AHK variables As I have explained earlier, the Hoeffding decomposition is a convenient tool to investigate U-statistics of i.i.d. variables. However, generalizing the

decomposition to U-statistics of RCE matrices is not straightforward. Remember that in the i.i.d. case, the Hoeffding decomposition consists of orthogonal projections of the kernel h on functional subspaces generated by the observations $(X_1, X_2, ...)$, see equation (1.6). For a RCE matrix Y, a decomposition based on subspaces generated by the entries of Y would not have verified the orthogonality conditions, because the latter are not i.i.d. The main idea behind finding a Hoeffding-type decomposition is to use the Aldous-Hoover-Kallenberg representation of RCE matrices. This representation enables us to write Y, and therefore the U-statistic, as a function of the AHK variables $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ and $(\zeta_{ij})_{i\geq 1, j\geq 1}$. Let $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ and $(\zeta_{ij})_{i\geq 1, j\geq 1}$ be families of AHK variables associated to Y, e.g. there exists a function ϕ such that for all $i \geq 1$, $j \geq 1$, we have

$$Y_{ij} \overset{a.s.}{=} \phi(\xi_i, \eta_j, \zeta_{ij}). \tag{1.12}$$

Instead of projecting on the subspaces generated by the observations, we can project on the subspaces generated by AHK variables. Because the AHK variables are i.i.d., the orthogonality can be obtained by choosing the right sets of AHK variables to generate the subspaces.

Like in the unidimensional case, the decomposition of the U-statistic is directly derived from the decomposition of the kernel function. I ended up deriving two distinct systems of projections, which can qualify as Hoeffding-type decompositions for U-statistics of RCE matrices.

First decomposition The first decomposition has been inspired by the fact that the Hoeffding decomposition in the unidimensional case is a decomposition subspaces generated by observations, i.e. sets of the form $(X_{\mathbf{i}})$. Since we cannot directly use the entries of Y without breaking the orthogonality, we use the sets of AHK variables $A_{\mathbf{i},\mathbf{j}}$ defined as follows. For any $\mathbf{i} \in \mathcal{P}(\llbracket m \rrbracket)$ and $\mathbf{j} \in \mathcal{P}(\llbracket n \rrbracket)$, $h(Y_{\mathbf{i},\mathbf{j}})$ is measurable by the set of AHK variables

$$A_{\mathbf{i},\mathbf{j}} := ((\xi_i)_{i\in\mathbf{i}}, (\eta_j)_{j\in\mathbf{j}}, (\zeta_{ij})_{\substack{i\in\mathbf{i}\\ j\in\mathbf{j}}}).$$

The projections are defined as follows. Let $\mathbf{i}' \in \mathcal{P}_{r'}(\mathbb{N})$ and $\mathbf{j}' \in \mathcal{P}_{c'}(\mathbb{N})$.

$$p^{\mathbf{i}',\mathbf{j}'}(h_{\mathbf{i},\mathbf{j}}) = \mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid A_{\mathbf{i}',\mathbf{j}'}] - \sum_{(0,0)\leq(r',c')<(r,c)} \sum_{\substack{\mathbf{i}''\subset\mathbf{i}'\\ \mathbf{i}''\subset\mathbf{i}'}} p^{\mathbf{i}'',\mathbf{j}''}(h_{\mathbf{i},\mathbf{j}}).$$

Then, the decomposition is given by

$$h_{\mathbf{i},\mathbf{j}} = \sum_{\substack{\mathbf{i}'\subseteq\mathbf{i}\\ \mathbf{i}'\subseteq\mathbf{i}}} p^{\mathbf{i}',\mathbf{j}'}(h_{\mathbf{i},\mathbf{j}}).$$

Second decomposition The second decomposition emerged later in the thesis, while adapting the theory of generalized U-statistics by Janson and Nowicki (1991) to my case. Let G be a

bipartite graph. We can define the set $H(G)$ of AHK variables associated to G as

$$H(G) = ((\xi_i)_{i\in V_1(G)}, (\eta_j)_{j\in V_2(G)}, (\zeta_{ij})_{(i,j)\in E(G)})$$

and $\mathcal{H}(G) = \sigma(H(G))$, the σ-field generated by the variables of $H(G)$. The projections are then defined similarly than in the previous system

$$p^G(h_{\mathbf{i},\mathbf{j}}) = \mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid H(G)] - \sum_{F\subset G} p^F(h_{\mathbf{i},\mathbf{j}}),$$

and the decomposition for h follows

$$h_{\mathbf{i},\mathbf{j}} = \sum_{G\subseteq K_{\mathbf{i},\mathbf{j}}} p^G(h_{\mathbf{i},\mathbf{j}}).$$

Link between the two decompositions Both decompositions are proved to be orthogonal decompositions, but they have different uses. The first projection is the minimal Hoeffding-type decomposition, insofar as the U-statistic is, in consequence, decomposed into the smallest number of terms capturing all the order of variation of the variance of $U_{m,n}$. Indeed, we will show that

$$\mathbb{V}[U_{m,n}] = \sum_{(0,0)<(r,c)\leq(p,q)} \binom{p}{r}^2 \binom{q}{c}^2 \binom{m}{r}^{-1} \binom{n}{c}^{-1} \mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})], \tag{1.13}$$

so for some (r,c) such that $(0,0) < (r,c) \leq (p,q)$, the contribution of $\mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})]$ in $\mathbb{V}[U_{m,n}]$ is $O(m^{-r}n^{-c})$.

The second decomposition uses projections on more subspaces than the first decomposition. Indeed, the sets used in the first decomposition can be written with bipartite graphs like in the second decomposition, since $A_{\mathbf{i},\mathbf{j}} = H(K_{\mathbf{i},\mathbf{j}})$, where $K_{\mathbf{i},\mathbf{j}} = (V_1(K_{\mathbf{i},\mathbf{j}}), V_2(K_{\mathbf{i},\mathbf{j}}), E(K_{\mathbf{i},\mathbf{j}})) := (\mathbf{i},\mathbf{j},\mathbf{i}\times\mathbf{j})$ is the fully connected bipartite graph with row nodes $\mathbf{i}$ and column nodes $\mathbf{j}$. In constrast, the second decomposition also considers the subspaces generated by $H(G)$ for all $G \subseteq K_{\mathbf{i},\mathbf{j}}$. Although it is much more complex than the first decomposition, it can be used to identify all the limit distributions for degenerate U-statistics, which is not possible with the first decomposition.

Limit theorems

The key theoretical result for my methodology is the identification of the limit distribution of the U-statistics $U_N := U_{m_N,n_N}$ of row-column exchangeable matrices of size $m_N \times n_N$ when $N \to \infty$, where $N = m_N + n_N$ and $m_N/N \to \rho \in]0,1[$. All of these results can be extended to the multivariate case, i.e. the joint convergence of a vector of U-statistics via the Cramér-Wold device. Provided all the involved U-statistics have the same rate of convergence, e.g. they are all non-degenerate or degenerate of the same order, then these results can be further extended to functions of U-statistics, using the *delta-method*, the details of which are left aside for now.

Quadruplets kernels The first convergence results apply to U-statistics with kernels of quadruplets, i.e. submatrices of size 2×2.

Theorem 2.2.5 (Chp. 2, Le Minh, 2023)**.** *Let Y be a RCE matrix. Let h be a quadruplet kernel such that $\mathbb{E}[h^2_{\{1,2\},\{1,2\}}] < \infty$. Let $\mathcal{F}_N = \sigma\big((U^h_{k,l}, k \geq m_N, l \geq n_N)\big)$ and $\mathcal{F}_\infty := \bigcap_{N=1}^\infty \mathcal{F}_N$. Set $U_\infty = \mathbb{E}[h_{\{1,2\};\{1,2\}} \mid \mathcal{F}_\infty]$ and*

$$V = \frac{4}{\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{1,3\},\{3,4\}} \mid \mathcal{F}_\infty\big) + \frac{4}{1-\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{3,4\},\{1,3\}} \mid \mathcal{F}_\infty\big).$$

If $\mathbb{P}(V > 0) > 0$, then

$$\sqrt{N}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} W,$$

where W is a random variable with characteristic function $\phi(t) = \mathbb{E}[\exp(-\frac{1}{2}t^2 V)]$.

Theorem 2.2.7 (Chp. 2, Le Minh, 2023)**.** *In addition to the hypotheses of Theorem 2.2.5, if Y is dissociated, then U^h_∞ and V^h are constant and*

$$\sqrt{N}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V),$$

More precisely,

1. $U_\infty = \mathbb{E}[h_{\{1,2\},\{1,2\}}]$,
2. $V = \frac{4}{\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{1,3\},\{3,4\}}\big) + \frac{4}{1-\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{3,4\},\{1,3\}}\big)$.

The first theorem applies to both non-dissociated and dissociated U-statistics. It has been proved using the backward martingale convergence theorem of Eagleson and Weber (1978) (Thm. 1.5.8). The hypothesis $\mathbb{P}(V > 0) > 0$ ensures that the U-statistic is non-degenerate. This result shows that in the general non-degenerate case, where Y may not be dissociated, the distribution of a U-statistic converges to a mixture of gaussian variables. Because it is a complex distribution, this result cannot be easily exploited. The second theorem is a consequence in the dissociated case. It is deduced from a Hewitt-Savage-type argument. For statistical applications, the second part is the one being exploited.

Non-degenerate kernels of size $p \times q$ Although the previous theorem has been proved for kernels of size 2×2, the proof can actually be extended to kernels of any size $p \times q$, with the burden of heavier notation and combinatorial complexity. Later in the thesis, I have found out that the Hoeffding-type decompositions give simpler proofs for this result, even for kernels of size $p \times q$. The following theorem is the extension of the previous theorem in the dissociated case, proved with a Hoeffding-type decomposition.

Theorem 3.3.1 (Chp. 3, Le Minh et al., 2023)**.** *Let Y be a dissociated RCE matrix. Let h be a $p \times q$ kernel function such that $\mathbb{E}[h^2_{[\![p]\!],[\![q]\!]}] < \infty$. Set $U_\infty = \mathbb{E}[h_{[\![p]\!],[\![q]\!]}]$ and*

$$V = \frac{p^2}{\rho}\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})] + \frac{q^2}{1-\rho}\mathbb{V}[p^{\varnothing,\{1\}}(h_{[\![p]\!],[\![q]\!]})].$$

If $V > 0$, then

$$\sqrt{N}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V).$$

Note that despite having seemingly different expressions, the asymptotic variance V is the same in the two theorems. If $p = q = 2$, then

$$\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![2]\!],[\![2]\!]})] = \mathbb{V}[\mathbb{E}[h_{\{1,2\},\{1,2\}} \mid \xi_1]] = \mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{1,3\},\{3,4\}}\big) \tag{1.14}$$

and

$$\mathbb{V}[p^{\varnothing,\{1\}}(h_{[\![2]\!],[\![2]\!]})] = \mathbb{V}[\mathbb{E}[h_{\{1,2\},\{1,2\}} \mid \eta_1]] = \mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{3,4\},\{1,3\}}\big). \tag{1.15}$$

Degenerate kernels of size $p \times q$ Finally, in the degenerate case, the problem is much more complex, even in the dissociated case. To understand degeneracy, one could use the developed expression of the variance of $\mathbb{V}[U_N]$ given by equation (1.13). The degenerate case arises when V, which would have been the dominant term of the $\mathbb{V}[U_N]$, is equal to 0. In this case, the dominant part corresponds to the positive variance terms $\mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})]$ with the smallest value $d = r + c$. In this case, their contribution is $O(N^{-d})$. The right normalization for the limit theorem is therefore not $\sqrt{N}$, but $N^{d/2}$.

As for the limit distribution, it is not necessarily Gaussian. It can be identified with the second Hoeffding-type decomposition. Again, the variance can be decomposed according to this projection system similarly to (1.13). Let $\Gamma_{r,c}$ be the set of bipartite graphs G with $V_1(G) = [\![r]\!]$ and $V_2(G) = [\![c]\!]$ such that any bipartite graph with r row nodes and c column nodes is isomorphic to exactly one element of $\Gamma_{r,c}$. Then

$$\mathbb{V}[U_{m,n}] = \sum_{(0,0)<(r,c)\leq(p,q)} \binom{p}{r}^2 \binom{q}{c}^2 \binom{m}{r}^{-1} \binom{n}{c}^{-1} r!c! \sum_{G\in\Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1} \mathbb{V}[p^G].$$

The dominating part consists of positive variance terms corresponding to projections characterized by graphs $G \in \bigcup_{(r,c):r+c=d} \Gamma_{r,c}$, for some d. d is called the principal degree of h and these graphs G are called the principal support graphs. The limit distribution of U_N depends on the form of its principal support graphs. The following theorem identifies the gaussian case.

Theorem 4.2.8 (Chp. 4)**.** *If all principal support graphs of U_N are connected, then*

$$N^{d/2}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \sum_{\substack{(0,0)<(r,c)\leq(p,q)\\ r+c=d}} \frac{p!^2 q!^2}{(p-r)!^2(q-c)!^2} \sum_{G\in\Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1}\mathbb{V}[p^G].$$

I have not proved the other cases in this thesis. However, one can assume that, as in Janson and Nowicki (1991), the type of the limit distribution depends on the highest number of connected components, here denoted b, found in the principal support graphs. In Janson and Nowicki (1991), if $b = 1$, then the limit distribution is a Gaussian, if $b = 2$, then it is a sum of chi-squared distributions. For any b, it is a polynomial function of independent Gaussian distributions with degree b. There is a striking analogy with the asymptotics of degenerate U-statistics of i.i.d. observations (Thm. 1.5.13), but the form of the limit distribution is given by the number b instead of the principal degree d. Despite still unproven, I give several examples that seem to corroborate this assumption.

A variance estimator

In this thesis, I have used two approaches to estimate the asymptotic variance of U-statistics. Note that, just as limit theorems for U-statistics can be generalized to functions of U-statistics through the delta-method, so can the estimators of the asymptotic variance.

First variance estimator The asymptotic variances given by the limit theorems depend on the distribution of Y, i.e. the network model, and the kernel h. An analytic expression can be calculated for the asymptotic variance. Then, one hopes that all the quantities appearing in the analytic expression can be estimated to build a consistent estimator for the asymptotic variance. This approach works for both non-degenerate and degenerate U-statistics, but there are two downsides. First, the analytic calculation of the asymptotic variance might be tedious, especially for degenerate cases. Second, there is no general technique to estimate the quantities appearing in the analytic expression. In my examples, all these quantities can be estimated with other U-statistics. However, this might be due to chance.

Second variance estimator Later, I have devised a general estimator for V of Theorem 3.3.1. Notice that the first term appearing in V is $\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})] = \mathbb{V}[\mathbb{E}[h_{[\![p]\!],[\![q]\!]} \mid \xi_1]]$. Because the $(\xi_i)_{i\geq 1}$ are i.i.d., the conditional expectations $\mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid \xi_i]$ are also i.i.d. for all $i \geq 1$ as long as

$i \in \mathbf{i}$. Denote

$$\mu^{(i)} = \mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid \xi_i]$$

for any $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p(\mathbb{N}) \times \mathcal{P}_q(\mathbb{N})$ such that $i \in \mathbf{i}$. By exchangeability, this quantity does not depend on the elements of $(\mathbf{i},\mathbf{j})$ other than i. Therefore, the unbiased variance estimator for $\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})]$ using m_N rows is

$$\frac{1}{m_N(m_N-1)} \sum_{1 \le i_1 < i_2 \le m_N} (\mu^{(i_1)} - \mu^{(i_2)})^2.$$

However, the $(\mu^{(i)})_{i \ge 1}$ are unknown, so we also need to estimate them. This can be done with the following estimator

$$\widehat{\mu}_N^{(i)} := \binom{m_N - 1}{p-1}^{-1} \binom{n_N}{q}^{-1} \sum_{\substack{(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p(\mathbb{N}) \times \mathcal{P}_q(\mathbb{N}) \\ i \in \mathbf{i}}} h_{\mathbf{i},\mathbf{j}}. \tag{1.16}$$

Plugging in the estimators for the conditional expectations in the expression of the unbiased variance estimator, we obtain

$$\widehat{v}_N^{1,0} := \binom{m_N}{2}^{-1} \sum_{1 \le i_1 < i_2 \le m_N} \frac{(\widehat{\mu}_N^{(i_1)} - \widehat{\mu}_N^{(i_2)})^2}{2}.$$

By symmetry, define for all $j \ge 1$,

$$\widehat{\nu}_N^{(j)} := \binom{m_N}{p}^{-1} \binom{n_N - 1}{q-1}^{-1} \sum_{\substack{(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p(\mathbb{N}) \times \mathcal{P}_q(\mathbb{N}) \\ j \in \mathbf{j}}} h_{\mathbf{i},\mathbf{j}}. \tag{1.17}$$

and

$$\widehat{v}_N^{0,1} := \binom{n_N}{2}^{-1} \sum_{1 \le j_1 < j_2 \le n_N} \frac{(\widehat{\nu}_N^{(j_1)} - \widehat{\nu}_N^{(j_2)})^2}{2}.$$

Then, the following theorem enables us to use $\widehat{v}_N^{1,0}$ and $\widehat{v}_N^{0,1}$ to build a consistent estimator for V.

Theorem 3.4.4 (Chp. 3, Le Minh et al., 2023)**.** *We have* $\widehat{v}_N^{1,0} \xrightarrow[N\to\infty]{\mathbb{P}} \mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})]$ *and* $\widehat{v}_N^{0,1} \xrightarrow[N\to\infty]{\mathbb{P}} \mathbb{V}[p^{\varnothing,\{1\}}(h_{[\![p]\!],[\![q]\!]})]$. *As a consequence,*

$$\widehat{V}_N := \frac{p^2}{\rho}\widehat{v}_N^{1,0} + \frac{q^2}{1-\rho}\widehat{v}_N^{0,1} \xrightarrow[N\to\infty]{\mathbb{P}} V.$$

$\widehat{V}_N$ enables us to consistently estimate V without actually doing any analytic calculation. In this respect, it is similar to resampling methods such as bootstrap or jackknife methods. In my simulation studies, $\widehat{V}_N$ is at least as precise as the variance estimated through the first method. Therefore, this estimator works well in the non-degenerate case. For degenerate cases,

it is actually possible to derive analogous consistent estimators for $\mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})]$ for all $(0,0) \leq (r,c) \leq (p,q)$ in the same way, but as r and c respectively grow closer to p and q, they are less precise. This is due to the fact that the number of terms in the estimator of the type (1.16) or (1.17) is the average of $O(m_N^{p-r} n_N^{q-c})$ terms.

1.6.3. Outline

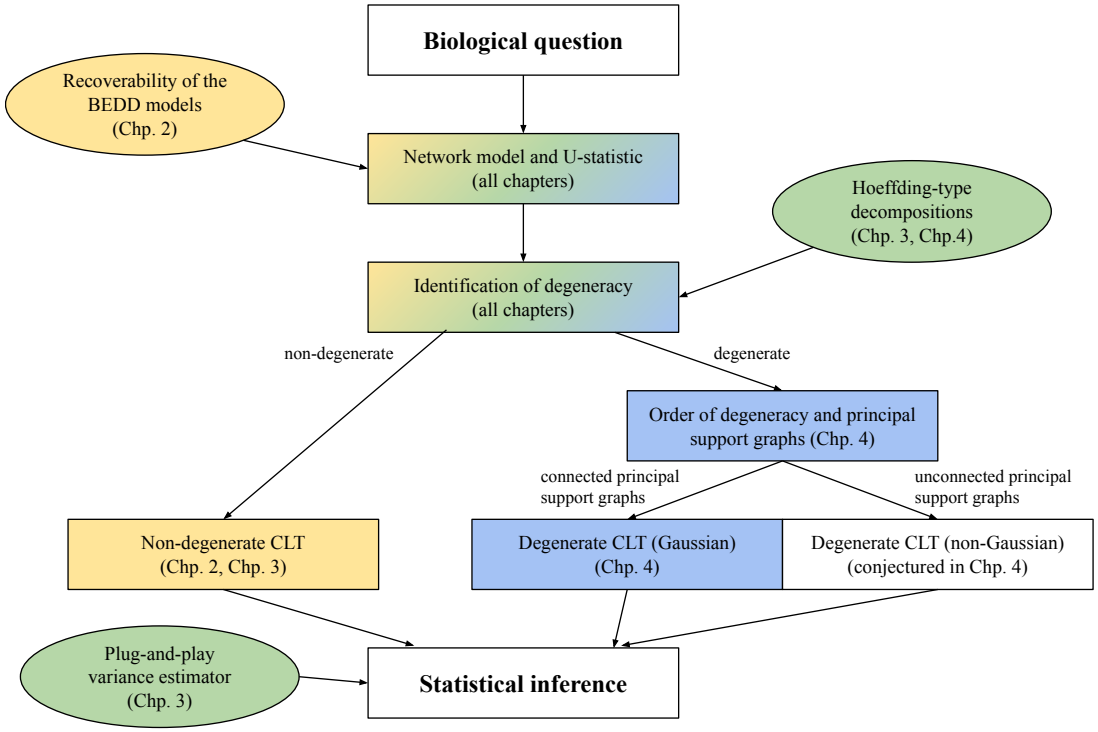

Figure 1.14 – The theoretical results and the methodology developed in this thesis.

Chapter 2 focuses on limit theorems for U-statistics of both non-dissociated and dissociated RCE matrices, with quadruplet kernels. These results are obtained using backward martingale arguments. In the non-dissociated case, the U-statistic converges to a mixture of Gaussians, which becomes a simple Gaussian in the dissociated case. The recoverability of the BEDD models is also proven in this chapter. Examples include the estimation of network row degree heterogeneity, network comparison and motif counting in networks.

Chapter 3 defines a first Hoeffding-type decomposition of U-statistics on dissociated RCE matrices. This decomposition is used to prove the limit theorem for U-statistics, with kernels of any size. A consistent estimator for their asymptotic variance is also built. Examples include motif counting in networks, estimation of graphon distances and the estimation of network row degree heterogeneity. The analysis of a legislature network dataset illustrates the method.

Chapter 4 deals with degenerate U-statistics of dissociated RCE matrices. A second Hoeffding-type decomposition is exhibited. Using this new decomposition, a limit theorem is derived for degenerate U-statistics when the limit is Gaussian. In the general case (non-Gaussian limit), a conjecture is formulated and commented. Examples of degenerate statistics are given, including the testing of network row degree heterogeneity.

Chapter 5 suggests avenues for future research, building on the work presented in this thesis. Some of these ideas consist in completing the methodology, providing it with the missing elements. Other ideas aim to improve the methodology, by dealing with the approximation error when using asymptotic results. Finally, the last ideas are about extending the current framework using RCE matrices to network models better suited to real networks, in particular sparse network models.

Figure 1.14 summarizes the previously listed results, mentioning in which chapter they appear.

Chapter 2

U-statistics on bipartite exchangeable networks

This chapter corresponds to the following article:

Le Minh, T. (2023). U-statistics on bipartite exchangeable networks. *ESAIM: Probability and Statistics*, 27:576–620. https://doi.org/10.1051/ps/2023010

Abstract Bipartite networks with exchangeable nodes can be represented by row-column exchangeable matrices. A quadruplet is a submatrix of size 2×2. A quadruplet U-statistic is the average of a function on a quadruplet over all the quadruplets of a matrix. We prove several asymptotic results for quadruplet U-statistics on row-column exchangeable matrices, including a weak convergence result in the general case and a central limit theorem when the matrix is also dissociated. These results are applied to statistical inference in network analysis. We suggest a method to perform parameter estimation, network comparison and motifs count for a particular family of row-column exchangeable network models: the bipartite expected degree distribution (BEDD) models. These applications are illustrated by simulations.

Contents

2.1. Introduction

RCE matrices Networks arise naturally when considering interaction data. The nodes of a network represent the entities of a system and an edge between two nodes represents the interaction between the associated entities. The network is bipartite when there are two different sets of nodes, and edges only link nodes of different types. A natural representation for a bipartite network is its rectangular adjacency matrix. The rows and columns of an adjacency matrix Y represent the two different types of nodes and each entry Y_{ij} encodes the interaction between the nodes associated to row i and column j, e.g. for binary networks, $Y_{ij} = 1$ if i and j interact and $Y_{ij} = 0$ else, or for weighted networks, Y_{ij} is the weight of the edge linking i and j.

Many probabilistic network models assume that the network units, either edges or nodes, are exchangeable, i.e. are invariant by permutation. In the adjacency matrix of a bipartite network, the edge-exchangeability corresponds to the exchangeability of all its entries (full exchangeability), such as in Adamczak et al. (2016), while the node-exchangeability refers to the exchangeability of its rows and columns, such as in Aldous (1981).

Recent developments have been made for edge-exchangeable models (Cai et al., 2016; Williamson, 2016), but node-exchangeable models have a longer history for both unipartite and bipartite networks and encompasses families of models such as the stochastic block model (Holland et al., 1983; Snijders and Nowicki, 1997), the latent block model (Govaert and Nadif, 2003), the latent position model (Hoff et al., 2002) or the random dot product graph model (Young and Scheinerman, 2007). Implicitly, the exchangeability of the network units is associated with a sampling assumption. The choice of whether considering the exchangeability for edges or nodes depends on what is assumed to be sampled to observe the networks, whether it be edges or nodes (Crane and Dempsey, 2018).

Let us observe a bipartite network represented by a finite submatrix of size $m \times n$. We assume that the nodes of the same type are infinitely exchangeable, which means that the row elements and the column elements of the adjacency matrix are separately invariant under infinite permutation. The infinite exchangeability assumption is equivalent to considering that this observed network is made of the first m rows and n columns of an infinite adjacency matrix, whose rows and columns are exchangeable. This assumption is similar to Orbanz and Roy (2014) for unipartite networks and provides a consistent framework to analyze networks of different sizes. It can be used with many random network models, including the ones listed above (stochastic block model, latent block model, latent position model and random dot product graph model).

However, it has to be distinguished from the finitely exchangeable case. Finite exchangeability does not imply infinite exchangeability, for example, if the observed network consists of the first m rows and n columns of a larger adjacency matrix but of finite size. In that case, we say that the (finitely) exchangeable sequences of nodes are not infinitely extendible (Konstantopoulos and Yuan, 2019; Mai, 2020). The finitely exchangeable case for networks has been notably studied by Lauritzen et al. (2018), but is out of scope of our paper. From here, the concept of exchangeability will always refer to infinite exchangeability, unless explicitely specified.

Thus, the exchangeability property of our infinite adjacency matrices is called row-column exchangeability. Let $\mathbb{S}_\infty$ be the group of finite permutations over $\mathbb{N}$. An infinite matrix Y is said to be row-column exchangeable (RCE) if for any couple $\Phi = (\sigma_1, \sigma_2) \in \mathbb{S}_\infty^2$,

$$\Phi Y \overset{\mathcal{D}}{=} Y,$$

where $\Phi Y := (Y_{\sigma_1(i)\sigma_2(j)})_{i\geq 1, j\geq 1}$.

U-statistics U-statistics form a large class of statistics with interesting properties for many purposes such as estimation and hypothesis testing. We are interested in using them to analyze RCE networks.

Given a sequence of random variables $(Y_1, Y_2, ..., Y_n)$ numbered with a unique index, a U-statistic is defined as the following average

$$U_n^h = \binom{n}{k}^{-1} \sum_{1 \le i_1 < i_2 < ... < i_k \le n} h(Y_{i_1}, Y_{i_2}, ..., Y_{i_k}), \tag{2.1}$$

where h is a symmetric function of size k referred to as the kernel.

The case where the $(Y_i)_{i \ge 1}$ are i.i.d. is well-studied. Halmos (1946) established the optimality of U-statistics as unbiased estimators and Hoeffding (1948) derived a central limit theorem (CLT), which ensures their asymptotic normality provided $\mathbb{E}[h(Y_{i_1}, Y_{i_2}, ..., Y_{i_k})^2] < \infty$. For dependent cases, results exist for several dependency strucures, for example Nandi and Sen (1963); Zhao and Chen (1990) for finitely exchangeable variables, Reitzner and Schulte (2013) for Poisson point processes or Duchemin et al. (2020, 2022) for Markov chains.

In the infinitely exchangeable case, it is particularly convenient to view $h(Y_{i_1}, Y_{i_2}, ..., Y_{i_k})$ as an array of random variables $(X_{\boldsymbol{i}})_{\boldsymbol{i}}$ indexed by k-tuples $\boldsymbol{i} = (i_1, i_2, ..., i_k)$ where $X_{\boldsymbol{i}} = X_{(i_1, i_2, ..., i_k)}$. With this notation, it becomes clear that the U-statistic defined by (2.1) is the sum of the corresponding entries of the k-dimensional array X. But $(Y_1, Y_2, ...)$ being exchangeable implies that the array X is jointly exchangeable, i.e. it is invariant by the action of joint permutations on each of its indices, for any sequence of k-tuples $(\boldsymbol{i}, \boldsymbol{j}, ...) \in \mathbb{N}^k$ and for any finite permutation $\sigma \in \mathbb{S}_\infty$,

$$(X_{\boldsymbol{i}}, X_{\boldsymbol{j}}, ...) \overset{\mathcal{D}}{=} (X_{(\sigma(i_1), \sigma(i_2), ..., \sigma(i_k))}, X_{(\sigma(j_1), \sigma(j_2), ..., \sigma(j_k))}, ...). \tag{2.2}$$

Eagleson and Weber (1978) proved a CLT for sums jointly exchangeable arrays, which applies to U-statistics of exchangeable sequences. Many other asymptotic results for U-statistics of exchangeable sequences were derived afterwards, such as a Berry-Esseen bound (van Zwet, 1984) and a law of iterated logarithm (Scott and Huggins, 1985).

In relation with the existing literature, we add here the definition of separate exchangeability. $(X_{\boldsymbol{i}})_{\boldsymbol{i}}$ is said to be separately exchangeable if for any sequence of k-tuples $(\boldsymbol{i}, \boldsymbol{j}, ...) \in \mathbb{N}^k$ and for any permutations $\sigma_1, \sigma_2, ..., \sigma_k$ of $\mathbb{S}_\infty$,

$$(X_{\boldsymbol{i}}, X_{\boldsymbol{j}}, ...) \overset{\mathcal{D}}{=} (X_{(\sigma_1(i_1), \sigma_2(i_2), ..., \sigma_k(i_k))}, X_{(\sigma_1(j_1), \sigma_2(j_2), ..., \sigma_k(j_k))}, ...). \tag{2.3}$$

U-statistics for RCE matrices Our contribution applies to U-statistics based on submatrices of size 2×2, that we call quadruplets, of an (infinite) RCE matrix Y

$$Y_{(i_1, i_2; j_1, j_2)} := \begin{pmatrix} Y_{i_1 j_1} & Y_{i_1 j_2} \\ Y_{i_2 j_1} & Y_{i_2 j_2} \end{pmatrix}.$$

Their kernels are real functions h over quadruplets. To mimic the kernel symmetry in the unidimensional case, we assume that they present the following symmetry property: for any

matrix Y,

$$h(Y_{(1,2;1,2)}) = h(Y_{(2,1;1,2)}) = h(Y_{(1,2;2,1)}) = h(Y_{(2,1;2,1)}). \tag{2.4}$$

This assumption can be made without loss of generality for U-statistics, since any quadruplet function k can be made symmetric considering

$$h(Y_{(1,2;1,2)}) = \frac{1}{4}\left(k(Y_{(1,2;1,2)}) + k(Y_{(2,1;1,2)}) + k(Y_{(1,2;2,1)}) + k(Y_{(2,1;2,1)})\right)$$

and $\mathbb{E}[h(Y_{(1,2;1,2)})] = \mathbb{E}[k(Y_{(1,2;1,2)})]$.

Applied to an observed network represented by the first m rows and n columns of Y, a quadruplet U-statistic is then defined by

$$U^h_{m,n} = \binom{m}{2}^{-1}\binom{n}{2}^{-1} \sum_{\substack{1\le i_1<i_2\le m\\ 1\le j_1<j_2\le n}} h(Y_{(i_1,i_2;j_1,j_2)}), \tag{2.5}$$

where $\binom{m}{2}$ is the number of 2-combinations from m elements. For clarity, we define the 4-dimensional array X using the following notation

$$X_{\{i_1,i_2;j_1,j_2\}} := h(Y_{(i_1,i_2;j_1,j_2)})$$

which means that a U-statistic is the mean of the first $\binom{m}{2} \times \binom{n}{2}$ entries of X. However, U-statistics of jointly exchangeable arrays deal with the mean of the first $\binom{m}{4}$ entries of the array X. Therefore, Theorem 4 of Eagleson and Weber (1978) applies to U-statistics of square matrices, but not generally to the case of bipartite networks, where row and column nodes are distinct by nature. In particular, U-statistics for RCE matrices allow row and column indices to overlap and most importantly, m to be different from n. Instead, the invariance structure of X is a special case of π-exchangeability (Kallenberg, 1999), where for any two permutations σ_1 and σ_2 of $\mathbb{S}_\infty$, we have

$$(X_{\{i_1,i_2;j_1,j_2\}})_{\substack{\{i_1,i_2\}\subset\mathbb{N}\\ \{j_1,j_2\}\subset\mathbb{N}}} \overset{\mathcal{D}}{=} (X_{\{\sigma_1(i_1),\sigma_1(i_2);\sigma_2(j_1),\sigma_2(j_2)\}})_{\substack{\{i_1,i_2\}\subset\mathbb{N}\\ \{j_1,j_2\}\subset\mathbb{N}}}. \tag{2.6}$$

Therefore, X is not separately exchangeable because, compared to (2.3), the same permutation σ_1 has to be applied on both row indices i_1 and i_2 and the same permutation σ_2 on both column indices j_1 and j_2.

Lemma 12 of Kallenberg (1999) establishes a strong law of large numbers for π-exchangeable variables, which applies to our U-statistics. Our aim is to establish a weak convergence theorem similar to Theorem 4 of Eagleson and Weber (1978). In the recent literature, two related results were obtained by Austern and Orbanz (2022) and Davezies et al. (2021). Austern and Orbanz (2022) explains how a result from Lindenstrauss (1999) can be translated to a strong law of large numbers for sums of exchangeable arrays and their Theorem 17 is analogous to Theorem

4 of Eagleson and Weber (1978) but it is obtained using Stein's method. Davezies et al. (2021) adopted a functional point of view. Their Theorem 2.1 is a Donsker-type version of the same result on jointly exchangeable arrays and Theorem 3.4 is an extension to separately exchangeable arrays. Because U-statistics of jointly exchangeable arrays are not suited to bipartite networks and because our arrays are not separately exchangeable, these results do not apply to our U-statistics of RCE matrices, as defined in (2.5) where (2.6) is satisfied.

To generalize these results to our case, our proof relies on the convergence of sums of backward martingales (Thm. 1 of Eagleson and Weber, 1978). We derive a CLT result in the case where the RCE matrix Y is dissociated, i.e. any of its submatrices with disjoint indexing sets are independent. This CLT excludes the degenerate case (i.e. when the convergence rate to the limiting distribution is greater than $\sqrt{N}$, see Sect. 2.4) through a clear assumption on the asymptotic variance. In the degenerate case, the convergence result of Austern and Orbanz (2022) does not lead to a CLT neither, and Davezies et al. (2021) proved a different convergence theorem. We offer a discussion on the degenerate case and its implications in Section 2.4. Finally, we recall that the backward martingale approach also yields Kallenberg's strong law of large numbers (Kallenberg, 1999).

In the last part of this work, we will put a special emphasis on the statistical analysis of bipartite networks. We introduce two versions of a RCE matrix model, the Bipartite Expected Degree Distribution (BEDD) model and we explain how our theorems apply to both of them. We suggest a method to perform statistical inference on these models using quadruplet U-statistics through several examples and we discuss how one can extend it.

Outline The weak convergence theorem in the general case and the CLT in the dissociated case are presented and proven in Section 2.2. We shed further light on the difference between the dissociated and the non-dissociated cases using the Aldous-Hoover representation theorem. Section 2.3 illustrates our results with applications to statistical network analysis using a RCE model and several examples of inference tasks.

2.2. Main result

2.2.1. Asymptotic framework

Our results apply in an asymptotic framework where the numbers of rows and columns of the submatrix used in the calculation of the U-statistic grow at the same rate, i.e. $m/(m+n) \to c \in]0,1[$. To simplify the proofs, we allow only one row or one column to be added to the

submatrix. Now, we build a sequence of dimensions $(m_N, n_N)_{N\geq 1}$ for the submatrix satisfying these conditions.

Definition 2.2.1 (Sequence of dimensions)**.** *Let c be an irrational number such that $0 < c < 1$. For all $N \in \mathbb{N}$, we define $m_N = 2 + \lfloor c(N+1) \rfloor$ and $n_N = 2 + \lfloor (1-c)(N+1) \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function.*

Proposition 2.2.2. *m_N and n_N satisfy:*

1. $\frac{m_N}{m_N+n_N} \xrightarrow[N\to\infty]{} c$,
2. $m_N + n_N = 4 + N$, *for all* $N \in \mathbb{N}$.

Corollary 2.2.3. *At each iteration $N \in \mathbb{N}^*$, one and only one of these two propositions is true:*

1. $m_N = m_{N-1} + 1$ *and* $n_N = n_{N-1}$,
2. $n_N = n_{N-1} + 1$ *and* $m_N = m_{N-1}$.

Throughout the paper, only sequences satisfying Definition 2.2.1 are considered. Such sequences m_N and n_N satisfy the desired growth conditions (proof given in Appendix 2.A). We investigate the asymptotic behaviour of $U^h_{m,n}$ through U^h_N defined as follows.

Definition 2.2.4. *With $(m_N, n_N)_{N\geq 1}$ introduced by Definition 2.2.1 and $U^h_{m,n}$ specified by equation (2.5), we set the sequence of U-statistics $(U^h_N)_{N\geq 1}$ such that for all $n \in \mathbb{N}$, $U^h_N = U^h_{m_N,n_N}$.*

2.2.2. Theorems

We establish the following results on the asymptotic behaviour of U-statistics over RCE matrices.

Theorem 2.2.5 (Main theorem)**.** *Let Y be a RCE matrix. Let h be a quadruplet kernel such that $\mathbb{E}[h(Y_{(1,2;1,2)})^2] < \infty$. Let $(U^h_N)_{N\geq 1}$ be the sequence of U-statistics associated with h defined in Definition 2.2.4. Let $\mathcal{F}_N = \sigma\big((U^h_{k,l}, k \geq m_N, l \geq n_N)\big)$ and $\mathcal{F}_\infty := \bigcap_{N=1}^{\infty} \mathcal{F}_N$. Set $U^h_\infty = \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty]$ and*

$$V = \frac{4}{c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(1,3;3,4)})\big|\mathcal{F}_\infty\big) + \frac{4}{1-c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(3,4;1,3)})\big|\mathcal{F}_\infty\big).$$

If $\mathbb{P}(V > 0) > 0$, then

$$\sqrt{N}(U^h_N - U^h_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} W,$$

where W is a random variable with characteristic function $\phi(t) = \mathbb{E}[\exp(-\frac{1}{2}t^2 V)]$.

Theorem 2.2.5 states that the limiting distribution of $\sqrt{N}(U_N^h - U_\infty^h)$ is a mixture of Gaussians. V consists of two terms corresponding to the covariance of the kernel taken on two quadruplets sharing one row or one column, conditional on $\mathcal{F}_\infty$. The condition $\mathbb{P}(V > 0) > 0$ is used to avoid the case $V = 0$ almost surely, which is a degenerate case discussed in Section 2.4. Since V is not constant in general, the limit distribution is an infinite mixture of Gaussians. This expression is analogous to η^2 in Theorem 4 of Eagleson and Weber (1978) for jointly exchangeable arrays and the covariance kernel in Theorems 2.1 and 3.4 of Davezies et al. (2021) for jointly and separately exchangeable arrays. We see that if V is constant, then the asymptotic distribution is a simple Gaussian. One may observe that U_N^h and $\sum_{\phi \in \mathbb{S}_N^2} f(\phi Y)$ studied in Corollary 19 of Austern and Orbanz (2022) are related as $U_N^h = (m_N!)^{-2} \sum_{\phi \in \mathbb{S}_{m_N}^2} f(\phi Y)$ when $m_N = n_N$ and $f(Y) = h(Y_{(1,2;1,2)})$. Still the convergence rates given in Theorem 2.2.5 and Corollary 19 of Austern and Orbanz (2022) (the proof of which is not given in the paper) are inconsistent. From what we understand, Corollary 19 actually corresponds to a degenerate case ($V = 0$) in Theorem 2.2.5. Next we identify a class of models where the limiting distribution of $\sqrt{N}(U_N^h - U_\infty^h)$ is a simple Gaussian.

Definition 2.2.6. *Y is a dissociated matrix if and only if $(Y_{ij})_{1 \le i \le m, 1 \le j \le n}$ is independent of $(Y_{ij})_{i > m, j > n}$, for all m and n.*

In other words, Y is dissociated if submatrices that are not sharing any row or column are independent. Now we claim the following extension to Theorem 2.2.5 for dissociated RCE matrices.

Theorem 2.2.7. *In addition to the hypotheses of Theorem 2.2.5, if Y is dissociated, then U_∞^h and V are constant and*

$$\sqrt{N}(U_N^h - U_\infty^h) \xrightarrow[N \to \infty]{\mathcal{D}} \mathcal{N}(0, V),$$

More precisely,

1. $U_\infty^h = \mathbb{E}[h(Y_{(1,2;1,2)})]$,
2. $V = \frac{4}{c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(1,3;3,4)})\big) + \frac{4}{1-c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(3,4;1,3)})\big)$.

This result can be more directly exploited for statistical applications, as the limiting distribution is much more simple. Another important result is the joint asymptotic normality of U-statistics, which holds as long as Theorem 2.2.7 applies to each kernel separately and they are linearly independent.

Theorem 2.2.8. *Let Y be a RCE dissociated matrix. Let $(h_1, h_2, ..., h_n)$ be a vector of quadruplet kernels such that*

1. *Theorem 2.2.7 applies for each kernel, i.e. $\mathbb{E}[h_k(Y_{(1,2;1,2)})^2] < \infty$ and $U_\infty^{h_k}$ and V^{h_k} are as defined in Theorem 2.2.7 for each kernel h_k, $1 \le k \le n$,*

2. *for* $t \in \mathbb{R}^n$, $\sum_{k=1}^n t_k h_k \equiv 0$ *if and only if* $t = (0, ..., 0)$.

Then

$$\sqrt{N}\left(\begin{pmatrix} U_N^{h_1} \\ U_N^{h_2} \\ \dots \\ U_N^{h_n} \end{pmatrix} - \begin{pmatrix} U_\infty^{h_1} \\ U_\infty^{h_2} \\ \dots \\ U_\infty^{h_n} \end{pmatrix}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

with

$$\Sigma = \left(C^{h_i,h_j}\right)_{1\le i,j\le n},$$

where $C^{h_k,h_\ell} = \lim_{N\to+\infty} N\mathrm{Cov}(U_N^{h_k}, U_N^{h_\ell})$ *for all* $1 \le k, \ell \le n$ *(and* $C^{h_k,h_k} = V^{h_k}$*).*

This theorem allows us to obtain the asymptotic normality of linear combinations of U-statistics and more interestingly, the asymptotic normality of differentiable functions of U-statistics (see Sect. 2.3.2).

Remark. Lemma 12 of Kallenberg (1999) already provided a strong law of large numbers for π-exchangeable arrays, of which quadruplet kernels are a subcase. The following theorem rephrases Kallenberg's law of large numbers for quadruplet U-statistics and gives an additional precision in the dissociated case for which we provide an alternative proof, as it is a natural consequence of our proof of Theorems 2.2.5 and 2.2.7.

Theorem 2.2.9. *Let* Y *be a RCE matrix. Let* h *be a quadruplet kernel. Let* $(U_N^h)_{N\ge1}$ *the sequence of* U*-statistics associated with* h *defined in Definition 2.2.4. Let* $\mathcal{F}_N = \sigma\big((U_{k,l}^h, k \ge m_N, l \ge n_N)\big)$ *and* $\mathcal{F}_\infty := \bigcap_{N=1}^\infty \mathcal{F}_N$*. We have*

$$U_N^h \xrightarrow[N\to\infty]{a.s.} \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty].$$

Furthermore, if Y *is dissociated, then* $\mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty] = \mathbb{E}[h(Y_{(1,2;1,2)})]$.

2.2.3. The Aldous-Hoover theorem

We shall explain Theorems 2.2.5 and 2.2.7 in the light of the Aldous-Hoover representation theorem. Theorem 1.4 of Aldous (1981) states that for any RCE matrix Y, there exists a real function f such that if we denote $Y_{ij}^* = f(\alpha, \xi_i, \eta_j, \zeta_{ij})$, for $1 \le i, j < \infty$, where the α, ξ_i, η_j and ζ_{ij} are i.i.d. random variables with uniform distribution over $[0, 1]$, then

$$Y \overset{\mathcal{D}}{=} Y^*. \tag{2.7}$$

It is possible to identify the role of each of the random variables involved in the representation theorem. We notice that each Y_{ij} is determined by α, ξ_i, η_j and ζ_{ij}. ζ_{ij} is entry-specific while

ξ_i is shared by all the entries involving the row i and η_j by the ones involving the column j. Therefore, the ξ_i and η_j represent the contribution of each individual of type 1 and type 2 of the network, i.e. each row and column of the matrix. These contributions are i.i.d., which makes the network exchangeable. Finally, α is global to the whole network and shared by all entries.

Proposition 3.3 of Aldous (1981) states that if Y is dissociated, then Y^* can be written without α, i.e. it is of the form $Y^*_{ij} = f(\xi_i, \eta_j, \zeta_{ij})$, for $1 \leq i, j < \infty$. In this case, because the ξ_i, η_j and ζ_{ij} are i.i.d., averaging with the U-statistic over an increasing number of nodes nullifies the contribution of each individual interaction (ζ_{ij}) and node (ξ_i and η_j). In the general case, i.e. when Y is not dissociated, then conditionally on α, Y is dissociated. It is easy to see that the mixture of Gaussians from Theorem 2.2.5 results from this conditioning.

In practice, dissociated exchangeable random graph models are widely spread. Notably, a RCE model is dissociated if and only if it can be written as a W-graph (or graphon), i.e. it is defined by a distribution $\mathcal{W}$ depending on two parameters in $[0,1]$ such that for $1 \leq i, j < \infty$:

$$\xi_i, \eta_j \overset{i.i.d.}{\sim} \mathcal{U}[0,1]$$
$$Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{W}(\xi_i, \eta_j),$$

see Diaconis and Janson (2008) for binary bipartite graphs, Lovász and Szegedy (2010) for an extension to weighted graphs but in a unipartite setup. In this definition, it is easy to recognize the variables from the representation theorem of Aldous-Hoover. We simply identify the ξ_i and η_j, then it suffices to take $\phi^{-1}_{\xi_i,\eta_j}$ the inverse distribution function of $\mathcal{W}(\xi_i, \eta_j)$ to see that defining the dissociated RCE matrix Y^* such that $Y^*_{ij} = f(\xi_i, \eta_j, \zeta_{ij}) := \phi^{-1}_{\xi_i,\eta_j}(\zeta_{ij})$ fulfills $Y^* \overset{\mathcal{D}}{=} Y$.

2.2.4. Proof of Theorem 2.2.5

To prove Theorem 2.2.5, we adapt the proof of Eagleson and Weber (1978) establishing the asymptotic normality of sums of backward martingale differences. The definition of a backward martingale is reminded in Appendix 2.B.

Theorem 2.2.10 (Eagleson and Weber, 1978)**.** *Let $(M_n, \mathcal{F}_n)_{n\geq 1}$ be a square-integrable reverse martingale, V a $\mathcal{F}$-measurable, a.s. finite, positive random variable. Denote $M_\infty := \mathbb{E}[M_1 \mid \mathcal{F}_\infty]$ where $\mathcal{F}_\infty := \bigcap_{n=1}^\infty \mathcal{F}_n$. Set $Z_{nk} := \sqrt{n}(M_k - M_{k+1})$. If:*

1. *$\sum_{k=n}^\infty \mathbb{E}[Z^2_{nk} \mid \mathcal{F}_{k+1}] \xrightarrow[n\to\infty]{\mathbb{P}} V$ (asymptotic variance),*
2. *for all $\epsilon > 0$, $\sum_{k=n}^\infty \mathbb{E}[Z^2_{nk} \mathbb{1}_{\{|Z_{nk}|>\epsilon\}} \mid \mathcal{F}_{k+1}] \xrightarrow[n\to\infty]{\mathbb{P}} 0$ (conditional Lindeberg condition),*

then $\sum_{k=n}^\infty Z_{nk} = \sqrt{n}(M_n - M_\infty) \xrightarrow[n\to\infty]{\mathcal{D}} W$, where W is a random variable with characteristic function $\phi(t) = \mathbb{E}[\exp(-\frac{1}{2}t^2 V)]$.

Proof of Theorem 2.2.5. The three steps to apply Theorem 2.2.10 to $(M_N)_{N\geq 1} = (U_N^h)_{N\geq 1}$ are to show that it is a backward martingale for a well chosen filtration and that it fulfills conditions 1 and 2. The expression of V is made explicit along the way. More precisely,

1. first, defining $\mathcal{F}_N = \sigma\big((U_{k,l}^h, k \geq m_N, l \geq n_N)\big)$, Proposition 2.C.1 states that $(U_N^h, \mathcal{F}_N)_{N\geq 1}$ is indeed a square-integrable reverse martingale ;
2. then, Proposition 2.D.1 implies that $\sum_{K=N}^{\infty} \mathbb{E}[Z_{NK}^2 \mid \mathcal{F}_{K+1}]$, where $Z_{NK} := \sqrt{N}(U_K - U_{K+1})$, does converge to a random variable V with the desired expression ;
3. finally, the conditional Lindeberg condition is ensured by Proposition 2.E.1, since from it, we deduce that for all $\epsilon > 0$, $\sum_{K=N}^{\infty} \mathbb{E}[Z_{NK}^2 \mathbb{1}_{\{|Z_{NK}|>\epsilon\}} \mid \mathcal{F}_{K+1}] \xrightarrow[N\to\infty]{\mathbb{P}} 0$.

Hence, if V is positive, Theorem 2.2.10 can be applied to U_N^h and we obtain that $\sqrt{N}(U_N^h - U_\infty^h) \xrightarrow[N\to\infty]{\mathcal{D}} W$, where W is a random variable with characteristic function $\phi(t) = \mathbb{E}[\exp(-\frac{1}{2}t^2 V)]$. The proofs of Propositions 2.C.1, 2.D.1 and 2.E.1 are provided in Appendices 2.C, 2.D and 2.E respectively. □

2.2.5. Proof of Theorem 2.2.7

The proof of Theorem 2.2.7 relies on a Hewitt-Savage type zero-one law for events that are permutable in our row-column setup. Therefore, it is useful to define first what a row-column permutable event is. We remind the Aldous-Hoover representation theorem for dissociated RCE matrices as stated earlier: if Y is a dissociated RCE matrix, then its distribution can be written with $(\xi_i)_{1\leq i<m_N}$, $(\eta_j)_{1\leq j<n_N}$ and $(\zeta_{ij})_{1\leq i\leq m_N, 1\leq j\leq n_N}$ arrays of i.i.d. random variables.

Then let us consider such arrays of i.i.d. random variables $(\xi_i)_{1\leq i<m_N}$, $(\eta_j)_{1\leq j<n_N}$ and $(\zeta_{ij})_{1\leq i\leq m_N, 1\leq j\leq n_N}$. If we were to consider events depending only on them, there is no loss of generality in using the product probability space $(\Omega_N, \mathcal{A}_N, \mathbb{P}_N)$, where

$$\begin{aligned}
\Omega_N &= \left\{(\omega^\xi, \omega^\eta, \omega^\zeta) : \omega^\xi \in \mathbb{R}^{m_N}, \omega^\eta \in \mathbb{R}^{n_N}, \omega^\zeta \in \mathbb{R}^{m_N n_N}\right\} = \mathbb{R}^{m_N + n_N + m_N n_N},\\
\mathcal{A}_N &= \mathcal{B}(\mathbb{R})^{m_N + n_N + m_N n_N},\\
\mathbb{P}_N &= \mu^{m_N + n_N + m_N n_N}.
\end{aligned}$$

We then define the action of a row-column permutation on an element of Ω_N.

Definition 2.2.11. *Let $\Phi = (\sigma_1, \sigma_2) \in \mathbb{S}_{m_N} \times \mathbb{S}_{n_N}$. The action of Φ on $\omega \in \Omega_N$ is defined by*

$$\Phi\omega = \left(\sigma_1\omega^\xi, \sigma_2\omega^\eta, (\sigma_1, \sigma_2)\omega^\zeta\right)$$

where $\sigma_1\omega^\xi = (\omega^\xi_{\sigma_1(i)})_{1\leq i<m_N}$, $\sigma_2\omega^\eta = (\omega^\eta_{\sigma_2(j)})_{1\leq j<n_N}$ and $(\sigma_1, \sigma_2)\omega^\zeta = (\omega^\zeta_{\sigma_1(i)\sigma_2(j)})_{1\leq i<m_N, 1\leq j<n_N}$

Definition 2.2.12. *Let $A \in \mathcal{A}_N$. A is invariant by the action of $\mathbb{S}_{m_N} \times \mathbb{S}_{n_N}$ if and only if for all $\Phi \in \mathbb{S}_{m_N} \times \mathbb{S}_{n_N}$, $\Phi^{-1}A = A$, i.e.*

$$\{\omega : \Phi\omega \in A\} = \{\omega : \omega \in A\}.$$

Notation. In this section, we denote by $\mathcal{E}_N$ the collection of events of $\mathcal{A}_N$ that are invariant by row-column permutations of size $m_N \times n_N$, i.e. $\Phi \in \mathbb{S}_{m_N} \times \mathbb{S}_{n_N}$. We denote $\mathcal{E}_\infty := \bigcap_{n=1}^{\infty} \mathcal{E}_N$, which is the collection of events that are invariant by permutations of size $m_N \times n_N$, for all N.

The following theorem is an extension of the Hewitt-Savage zero-one law to the row-column setup.

Theorem 2.2.13. *For all $A \in \mathcal{E}_\infty$, $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.*

The proof of Theorem 2.2.13 is given in Appendix 2.F. Now we use this result to derive Theorem 2.2.7 from Theorem 2.2.5.

Proof of Theorem 2.2.7. In this proof, we specify the matrices over which the U-statistics are taken, i.e. for a RCE matrix Y, we denote $U^h_{k,l}(Y)$ instead of $U^h_{k,l}$ the U-statistic of size $k \times l$ with kernel h taken on Y, given by formula (2.5), and analogously $U^h_N(Y) := U^h_{m_N,n_N}(Y)$ and $U^h_\infty(Y) := \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty(Y)]$. We denote also $\mathcal{F}_N(Y) := \sigma\big((U^h_{k,l}(Y), k \geq m_N, l \geq n_N)\big)$ which are sets of events depending on Y, and $\mathcal{F}_\infty(Y) := \bigcap_{n=1}^{\infty} \mathcal{F}_N(Y)$.

Since Y is RCE and dissociated, Proposition 3.3 of Aldous (1981) states the existence of a real function f such that for $1 \leq i, j < \infty$, $Y^*_{ij} = f(\xi_i, \eta_j, \zeta_{ij})$ and $Y^* \overset{\mathcal{D}}{=} Y$, where ξ_i, η_j and ζ_{ij}, for $1 \leq i, j < \infty$ are i.i.d. random variables with uniform distribution on $[0,1]$. Therefore we can consider such function f and these random variables, the product spaces $(\Omega_N, \mathcal{A}_N, \mathbb{P}_N)$ and the sets $\mathcal{E}_N$ of invariant events defined earlier.

But $\mathcal{F}_N(Y^*) = \sigma\big((U^h_{k,l}(Y^*), k \geq m_N, l \geq n_N)\big) \subset \sigma(U^h_N(Y^*), \xi_i, \eta_j, \zeta_{ij}, i > m_N, j > n_N)$, so for all N, $\mathcal{F}_N(Y^*) \subset \mathcal{E}_N$. It follows that $\mathcal{F}_\infty(Y^*) \subset \mathcal{E}_\infty$, so $U_\infty(Y^*)$ is $\mathcal{E}_\infty$-measurable. Theorem 2.2.13 states that all the events in $\mathcal{E}_\infty$ happen with probability 0 or 1, so it ensures that $U^h_\infty(Y^*) = \mathbb{E}[h(Y^*_{\{1,2;1,2\}}) \mid \mathcal{F}_\infty(Y^*)] = \mathbb{E}[h(Y^*_{\{1,2;1,2\}})]$ is constant. Moreover, since the distribution of $U^h_N(Y)$ is the same as this of $U^h_N(Y^*)$, we can conclude that $U^h_\infty(Y) = \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty(Y)] = \mathbb{E}[h(Y_{(1,2;1,2)})]$.

Likewise, we deduce that $\mathbb{E}[h(Y_{(1,2;1,2)})h(Y_{(1,3;3,4)}) \mid \mathcal{F}_\infty(Y)] = \mathbb{E}[h(Y_{(1,2;1,2)})h(Y_{(1,3;3,4)})]$ and $\mathbb{E}[h(Y_{(1,2;1,2)})h(Y_{(3,4;1,3)}) \mid \mathcal{F}_\infty(Y)] = \mathbb{E}[h(Y_{(1,2;1,2)})h(Y_{(3,4;1,3)})]$ which gives the desired result for V. Thus we conclude that W of Theorem 2.2.5 follows a Gaussian distribution of variance V.

□

2.2.6. Proof of Theorem 2.2.8

The proof of Theorem 2.2.8 relies on the Cramér-Wold theorem (see Thm. 29.4 of Billingsley, 1995). If is enough to show that any linear combination of U-statistics converges to the corresponding linear combination of their limits.

Proof of Theorem 2.2.8. Let $(Z^{h_k})_{1\leq k\leq n}$ be a vector of random variables following a centered multivariate Gaussian distribution with covariance matrix Σ defined in the theorem. Then $Z^{h_k} \sim \mathcal{N}(0, V^{h_k})$ for all $1 \leq k \leq n$ and $\mathrm{Cov}(Z^{h_i}, Z^{h_j}) = C^{h_i,h_j}$ for all $1 \leq i \leq n$ and $1 \leq j \leq n$.

For some $t = (t_1, t_2, ..., t_n) \in \mathbb{R}^n$, we set $h_t := t_1h_1 + t_2h_2 + ... + t_nh_n$. First, assume that $t \neq (0, ..., 0)$. Then by hypothesis, $h_t \not\equiv 0$, therefore $\sum_{k=1}^n t_k U_N^{h_k} = U_N^{h_t}$ is a U-statistic with quadruplet kernel h_t. Using Cauchy-Schwarz inequality and the fact that $\mathbb{E}[h_k(Y_{(1,2;1,2)})^2] < \infty$ for all $1 \leq k \leq n$, we have furthermore

$$\begin{aligned}
\mathbb{E}[h_t(Y_{(1,2;1,2)})^2] &= \sum_{k=1}^n t_k^2 \mathbb{E}[h_k(Y_{(1,2;1,2)})^2] + 2 \sum_{1\leq k\neq \ell\leq n} t_k t_\ell \mathbb{E}[h_k(Y_{(1,2;1,2)}) h_\ell(Y_{(1,2;1,2)})], \\
&\leq \sum_{k=1}^n t_k^2 \mathbb{E}[h_k(Y_{(1,2;1,2)})^2] + 2 \sum_{1\leq k\neq \ell\leq n} t_k t_\ell \sqrt{\mathbb{E}[h_k(Y_{(1,2;1,2)})^2]\mathbb{E}[h_\ell(Y_{(1,2;1,2)})^2]}, \\
&< \infty.
\end{aligned}$$

Therefore, Theorem 2.2.7 also applies for $U_N^{h_t}$ and $\sqrt{N}(U_N^{h_t} - U_\infty^{h_t}) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V^{h_t})$, where $U_\infty^{h_t} = \sum_{k=1}^n t_k U_\infty^{h_k}$ and $V^{h_t} = \sum_{k=1}^n \sum_{\ell=1}^n t_k t_\ell C^{h_k,h_\ell} = t^T \Sigma t$ with $C^{h_k,h_k} = V^{h_k} > 0$ since Theorem 2.2.7 applies. This means that $\sqrt{N}(U_N^{h_t} - U_\infty^{h_t}) = \sqrt{N} \sum_{k=1}^n t_k (U_N^{h_k} - U_\infty^{h_k}) \xrightarrow[N\to\infty]{\mathcal{D}} \sum_{k=1}^n t_k Z^{h_k}$.

Now assume that $t = (0, ..., 0)$. Then $h_t \equiv 0$ so $U_N^{h_t} = 0 = \sum_{k=1}^n t_k Z^{h_k}$. Therefore, $\sqrt{N}(U_N^{h_t} - U_\infty^{h_t}) = \sqrt{N} \sum_{k=1}^n t_k (U_N^{h_k} - U_\infty^{h_k}) \xrightarrow[N\to\infty]{\mathcal{D}} \sum_{k=1}^n t_k Z^{h_k}$ is still true.

We have proven that $\sqrt{N}(U_N^{h_t} - U_\infty^{h_t}) \xrightarrow[N\to\infty]{\mathcal{D}} \sum_{k=1}^n t_k Z^{h_k}$ for all $t \in \mathbb{R}^n$, so we can finally apply the Cramér-Wold theorem (Thm. 29.4 of Billingsley, 1995) which states that $\sqrt{N}\left(U_N^{h_k} - U_\infty^{h_k}\right)_{1\leq k\leq n}$ converges jointy in distribution to $(Z^{h_k})_{1\leq k\leq n}$, which is a centered multivariate Gaussian with covariance matrix Σ, so this concludes the proof. □

2.2.7. Proof of Theorem 2.2.9

The proof for the first part of the theorem can be derived from the proof of Proposition 2.C.1, without needing the hypothesis $\mathbb{E}[h(Y_{(1,2;1,2)})^2] < \infty$. Indeed, it is enough to show that $(U_N^h, \mathcal{F}_N)_{N\geq 1}$ is a (not necessarily square-integrable) backward martingale and to apply Theorem 2.B.3.

As for the dissociated case, $\mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty] = \mathbb{E}[h(Y_{(1,2;1,2)})]$ is ensured by the proof of Theorem 2.2.7.

2.3. Applications

In this section, we illustrate how the results from the previous section can be used for statistical inference for network data through different examples. First, we introduce the Bipartite Expected Degree Distribution (BEDD) models, a family of RCE models and we show how Theorems 2.2.5 and 2.2.7 apply. Then, we detail three examples to show how one might exploit the U-statistics properties to analyze networks. In the first example, we use different quadruplet kernels to estimate the row heterogeneity of a network, with the help of the delta-method. Next, we extend this example to build a statistical test to compare the row heterogeneity of two networks. In the last example, we use U-statistics to estimate the frequency of network motifs.

2.3.1. The BEDD model

General model As examples of models for RCE matrices, we consider the family of the BEDD models, which are weighted, bipartite and exchangeable extensions of the Expected Degree Sequence model (Chung and Lu, 2002; Ouadah et al., 2022). For binary graphs, the degree of a node is the number of edges that stem from it. For weighted graphs, the equivalent notion is the sum of the weights of these edges. It is sometimes called node strength (Barrat et al., 2004), but we will simply refer to it as node weight. A BEDD model draws the node weights from two distributions, characterised by càdlàg, non-decreasing and bounded real functions f and g of $[0,1] \to \mathbb{R}_+$. The expected edge weights Y_{ij} are then proportional to the expected weights of the involved nodes. A BEDD model can be written in a hierarchical form

$$\begin{aligned} \xi_i, \eta_j &\overset{iid}{\sim} \mathcal{U}[0,1] \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{L}(\lambda f(\xi_i) g(\eta_j)). \end{aligned} \tag{2.8}$$

where given any real number $\mu \geq 0$, we denote by $\mathcal{L}(\mu)$ a family of probability distributions with expectation μ and finite variance, λ is a positive real number and f and g are normalized by the condition $\int f = \int g = 1$. For a graph of size $m \times n$, conditionally to λ, ξ_i and η_j, the expected weight of the i-th row is $n\lambda f(\xi_i)$ and the expected weight of the j-th column is $m\lambda g(\eta_j)$. Consequently, λ is the mean intensity of the network. One can define different BEDD models by specifying different families of distributions $\mathcal{L}$. For a family $\mathcal{L}$, we refer to the $\mathcal{L}$-BEDD model (e.g. Poisson-BEDD or Bernoulli-BEDD).

Furthermore, we define two versions of BEDD models:

Version 1 λ is constant,

Version 2 λ is a random variable.

By construction, the BEDD models are RCE, so Theorem 2.2.5 can be applied to matrices Y generated by these two versions of BEDD models. Theorem 2.2.7 only applies to Version 1, where the matrix is dissociated. Indeed, we see that in both models conditionally on λ, the expected mean of the interactions of any submatrix is λ. Therefore any 2 submatrices are independent if λ is constant. We could also have noticed that λ is determined by the α from the representation theorem of Aldous-Hoover, see equation (2.7). As a remark, it is straightforward that unlike Version 2, Version 1 of BEDD models can be written as a W-graph model as in Formula (2.2.3), setting $\mathcal{W}(\xi_i, \eta_j) \coloneqq \mathcal{L}(\lambda f(\xi_i) g(\eta_j))$.

Since Theorem 2.2.7 only applies to Version 1, we will be only considering this version in the rest of the article.

Definition 2.3.1. *Given a family of distributions $\mathcal{L}(\mu)$ a family of probability distributions with expectation μ and finite variance, a $\mathcal{L}$-BEDD model is a semi-parametric model described by the triplet $\Theta = (\lambda, f, g)$ where*

1. *$\lambda \in \mathbb{R}$,*
2. *f and g are real functions f and g of $[0,1] \to \mathbb{R}_+$ which are bounded, càdlàg, non-decreasing and normalized with $\int f = \int g = 1$.*

We call Θ the BEDD parameters and the matrix Y generated by a $\mathcal{L}$-BEDD model with these parameters is written $Y \sim \mathcal{L}\text{-BEDD}(\Theta)$ and is described by (2.8), for all $(i,j) \in \mathbb{N}^2$.

In this definition, the normalizing constraint on $\int f = \int g = 1$ ensures that $\mathbb{E}[Y_{ij}] = \lambda$ for all $(i,j) \in \mathbb{N}^2$. The boundedness of f and g ensures that the variables $f(\xi_i)$ and $g(\eta_j)$ are bounded and their moments exist. In the binary case (Bernoulli-BEDD), it also puts a condition on λ. Since $\mathbb{P}(Y_{ij} = 1 \mid \xi_i, \eta_j) = \lambda f(\xi_i) g(\eta_j)$, the condition $\lambda \leq \|f\|_\infty^{-1} \|g\|_\infty^{-1}$ must hold. The non-decreasing and càdlàg conditions are similar to the condition of Bickel and Chen (2009) for their random graph model and ensures the identifiability of the model since otherwise, f and g could be replaced with any $f \circ \pi_1$ and $g \circ \pi_2$, where π_1 and π_2 are measure-preserving transformations.

Identifiability by a quadruplet In addition to being a dissociated RCE model, the BEDD models are particularly well adapted to use of quadruplet U-statistics. In this paragraph, we show that for some choices of $\mathcal{L}$ such as the Poisson distribution, the model can be recovered by a single quadruplet. We use the following two theorems of which proofs are given in Appendix 2.G. The first theorem implies that the functions f and g of the BEDD models are characterised by their moments $F_k \coloneqq \int f^k$ and $G_k \coloneqq \int g^k$.

Theorem 2.3.2. *Let $\Theta = (\lambda, f, g)$ be BEDD parameters and $Y \sim \mathcal{L}\text{-BEDD}(\Theta)$ for some family of distributions $\mathcal{L}$. The distribution of Y is uniquely determined by λ, $(F_k)_{k\geq 1}$ and $(G_k)_{k\geq 1}$, where $F_k := \int f^k$ and $G_k := \int g^k$ for all $k \geq 1$.*

Now we specify an assumption on the family of distributions $\mathcal{L}(\mu)$, under which a quadruplet identifies the parameters $\Theta = (\lambda, f, g)$ of a BEDD model.

Assumption 2.3.3. *For the family of distributions $\mathcal{L}(\mu)$, there exists a sequence of functions $(\Psi_k)_{k\geq 1}$ such that if a random variable $X \sim \mathcal{L}(\mu)$, then for every $k \geq 1$,*

$$\mathbb{E}[\Psi_k(X)] = \mu^k.$$

This assumption holds for many usual distributions families such as the Poisson or the Binomial distributions. As an example, if X follows a Poisson distribution, we have $\Psi_k(X) = X(X-1)...(X-k+1)$. This assumption does not hold for the Bernoulli distribution, as if $X \sim \text{Bernoulli}(\mu)$, for any function φ, $\mathbb{E}[\varphi(X)] = \mu\varphi(1)$. This assumption is a sufficient condition to be able to recover the BEDD parameters from the joint distribution of a quadruplet.

Theorem 2.3.4. *If Assumption 2.3.3 holds for the family of distributions $\mathcal{L}(\mu)$, then for all $k \in \mathbb{N}$, F_k and G_k are uniquely determined by the joint distribution of a quadruplet.*

This theorem suggests that all the BEDD information is contained in the distribution of a quadruplet, therefore it is possible to extract any information only with quadruplet kernels. Quadruplet U-statistics are then especially of interest.

2.3.2. Heterogeneity in the row weights of a network

In this first example, we are interested in evaluating the heterogeneity of the row weights of a network. In the BEDD models, conditional to the latent variables $(\xi_i)_{1\leq i\leq m}$ following a uniform distribution on $[0,1]$, the expected weights of the row nodes are given by the $(f(\xi_i))_{1\leq i\leq m}$, as seen in Section 2.3.1. Therefore, the heterogeneity of the rows, i.e. the variance of the expected weight of a row node can be quantified by $F_2 := \int_0^1 f^2(u)du$. In the example of an interaction network, if f is constant, i.e. $f \equiv 1$ and $F_2 = 1$, then the expected weight is constant for all rows and the row weight distribution is homogeneous, with all the row individuals having around the same number of interactions. Besides, the higher F_2 is, the more this distribution is unbalanced. In ecology, a large value of F_2 indicates a strong distinction between generalist (with high degree) and specialists (with low degree) species.

F_2 can be estimated using $\widehat{\theta}_N := U_N^{h_1}/U_N^{h_2}$ where $U_N^{h_1}$ and $U_N^{h_2}$ are the U-statistics based on the quadruplet kernels h_1 and h_2 defined as

$$h_1(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_1j_2} + Y_{i_2j_1}Y_{i_2j_2}),$$

and

$$h_2(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_2j_2} + Y_{i_2j_1}Y_{i_1j_2}).$$

Proposition 2.3.5. *Let $\widehat{\theta}_N := U_N^{h_1}/U_N^{h_2}$ be defined as above. Then*

$$\sqrt{\frac{N}{V^\delta}}\left(\widehat{\theta}_N - F_2\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1), \tag{2.9}$$

where

$$V^\delta = \frac{1}{c}\left(F_4 + F_2(4F_2^2 - F_2 - 4F_3)\right) \tag{2.10}$$

and for all $k > 0$, $F_k := \int f^k$ and $G_k := \int g^k$.

This result comes from the composition of the asymptotic normality of two U-statistics. In the following, we show how to obtain equations (2.9) and (2.10). First, we see that $\mathbb{E}[h_1(Y_{(i_1,i_2;j_1,j_2)})] = \lambda^2 F_2$ and $\mathbb{E}[h_2(Y_{(i_1,i_2;j_1,j_2)})] = \lambda^2$. So applying Theorem 2.2.7 successively to $U_N^{h_1}$ and $U_N^{h_2}$ gives the following results (V^{h_1} and V^{h_2} are derived in Lemmas 2.H.3 and 2.H.4):

$$\sqrt{\frac{N}{V^{h_1}}}(U_N^{h_1} - \lambda^2 F_2) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1), \tag{2.11}$$

and

$$\sqrt{\frac{N}{V^{h_2}}}(U_N^{h_2} - \lambda^2) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1), \tag{2.12}$$

where

$$V^{h_1} = \frac{\lambda^4}{c}(F_4 - F_2^2) + \frac{4\lambda^4}{1-c}F_2^2(G_2 - 1), \tag{2.13}$$

and

$$V^{h_2} = \frac{4\lambda^4}{c}(F_2 - 1) + \frac{4\lambda^4}{1-c}(G_2 - 1). \tag{2.14}$$

To combine the results (2.11) and (2.12), we apply Theorem 2.2.8 to (h_1, h_2). We find that

$$\sqrt{N}\left(\begin{pmatrix} U_N^{h_1} \\ U_N^{h_2} \end{pmatrix} - \begin{pmatrix} \lambda^2 F_2 \\ \lambda^2 \end{pmatrix}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,\Sigma), \tag{2.15}$$

with

$$\Sigma = \begin{pmatrix} V^{h_1} & C^{h_1,h_2} \\ C^{h_1,h_2} & V^{h_2} \end{pmatrix},$$

where $C^{h_1,h_2} = 2\lambda^4 c^{-1}(F_3 - F_2) + 4\lambda^4(1-c)^{-1}F_2(G_2 - 1)$. The derivation of C^{h_1,h_2} is given by Lemma 2.H.5. We suggest two methods to derive the weak convergence result for $\widehat{\theta}_N = U_N^{h_1}/U_N^{h_2}$.

First method: the delta-method The first-order Taylor expansion of $\phi(U_N^{h_1}, U_N^{h_2}) = U_N^{h_1}/U_N^{h_2} = \widehat{\theta}_N$ at the point $(U_N^{h_1}, U_N^{h_2}) = (\lambda^2 F_2, \lambda^2)$ is

$$\widehat{\theta}_N - F_2 = \nabla\phi(\lambda^2 F_2, \lambda^2)^T \left(\begin{pmatrix} U_N^{h_1} \\ U_N^{h_2} \end{pmatrix} - \begin{pmatrix} \lambda^2 F_2 \\ \lambda^2 \end{pmatrix} \right) + o_P\left(\left\| \begin{pmatrix} U_N^{h_1} \\ U_N^{h_2} \end{pmatrix} - \begin{pmatrix} \lambda^2 F_2 \\ \lambda^2 \end{pmatrix} \right\| \right)$$

where $\nabla\phi$ is the gradient of ϕ and $\nabla\phi(U_N^{h_1}, U_N^{h_2})^T = (1/U_N^{h_2}, -U_N^{h_1}/(U_N^{h_2})^2)$.

As the result of (2.15), the delta-method (see Chapter 3 of Van der Vaart, 2000) gives equation (2.9) with

$$V^\delta = \nabla\phi(\lambda^2 F_2, \lambda^2)\Sigma\nabla\phi(\lambda^2 F_2, \lambda^2)^T = \frac{1}{\lambda^4} V^{h_1} - \frac{2F_2}{\lambda^4} C^{h_1,h_2} + \frac{F_2^2}{\lambda^4} V^{h_2},$$

which is equation (2.10).

Second method The delta-method is a generic method that applies to all differentiable functions ϕ. However, for our particular case, there is another way to find the same confidence intervals without using the delta-method. Let $t := (1, -F_2)^T$. One could have noticed that (2.15) also implies

$$\sqrt{\frac{N}{V^t}} t^T \left(\begin{pmatrix} U_N^{h_1} \\ U_N^{h_2} \end{pmatrix} - \begin{pmatrix} \lambda^2 F_2 \\ \lambda^2 \end{pmatrix} \right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1),$$

where $V^t := t^T \Sigma t = V^{h_1} - 2F_2 C^{h_1,h_2} + F_2^2 V^{h_2} = \lambda^4 V^\delta$. This can be rewritten

$$\sqrt{\frac{N}{V^\delta}} \frac{U_N^{h_2}}{\lambda^2} \left(\widehat{\theta}_N - F_2 \right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{2.16}$$

Since $U_N^{h_2}/\lambda^2 \xrightarrow[N\to\infty]{\mathbb{P}} 1$, then Slutsky's theorem yields equations (2.9) and (2.10).

Confidence intervals In order to exploit equations (2.9) and (2.10), we have to estimate the remaining unknown quantities λ^2, G_2, F_3 and F_4. We use the kernels h_3, h_4, h_5 and h_6 listed in Table 2.1. Using equation (2.10) and two additional U-statistics $U_N^{h_5}$ and $U_N^{h_6}$ based on the kernels h_5 and h_6 defined in Table 2.1, we build $\widehat{V}_N^\delta$ a consistent estimator for V^δ, defined as

$$\widehat{V}_N^\delta = \frac{1}{c} \left(\frac{U_N^{h_4}}{(U_N^{h_3})^2} + \frac{U_N^{h_1}}{U_N^{h_2}} \left(4\frac{(U_N^{h_1})^2}{(U_N^{h_2})^2} - \frac{U_N^{h_1}}{U_N^{h_2}} - 4\frac{U_N^{h_6}}{U_N^{h_5} U_N^{h_3}} \right) \right). \tag{2.17}$$

Finally, it follows from Slutsky's theorem that

$$\sqrt{\frac{N}{\widehat{V}_N^\delta}} \left(\widehat{\theta}_N - F_2 \right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1), \tag{2.18}$$

From this result, one can derive the following asymptotic confidence interval at level $\alpha \in]0,1[$ for F_2 using the $(1-\alpha/2)$-th percentile $q_{1-\alpha/2}$ of the standard normal distribution: for $N \geq 1$,

$$CI^{\delta}_{F_2}(\alpha, N) = \left[\widehat{\theta}_N - q_{1-\alpha/2}\sqrt{\frac{\widehat{V}^{\delta}_N}{N}}, \widehat{\theta}_N + q_{1-\alpha/2}\sqrt{\frac{\widehat{V}^{\delta}_N}{N}}\right]. \tag{2.19}$$

h	$h(Y_{(i_1,i_2;j_1,j_2)})$	$\mathbb{E}[h(Y_{(i_1,i_2;j_1,j_2)})]$
h_1	$\frac{1}{2}(Y_{i_1j_1}Y_{i_1j_2} + Y_{i_2j_1}Y_{i_2j_2})$	$\lambda^2 F_2$
h_2	$\frac{1}{2}(Y_{i_1j_1}Y_{i_2j_2} + Y_{i_1j_2}Y_{i_2j_1})$	λ^2
h_3	$\frac{1}{2}(Y_{i_1j_1}Y_{i_2j_1} + Y_{i_1j_2}Y_{i_2j_2})$	$\lambda^2 G_2$
h_4	$\frac{1}{2}\left((Y^2_{i_1j_1} - Y_{i_1j_1})(Y^2_{i_1j_2} - Y_{i_1j_2}) + (Y^2_{i_2j_1} - Y_{i_2j_1})(Y^2_{i_2j_2} - Y_{i_2j_2})\right)$	$\lambda^4 F_4 G_2^2$
h_5	$\frac{1}{4}(Y_{i_1j_1} + Y_{i_1j_2} + Y_{i_2j_1} + Y_{i_2j_2})$	λ
h_6	$\frac{1}{4}(Y_{i_1j_1}Y_{i_1j_2}(Y_{i_1j_1} + Y_{i_1j_2} - 2) + Y_{i_2j_1}Y_{i_2j_2}(Y_{i_2j_1} + Y_{i_2j_2} - 2))$	$\lambda^3 F_3 G_2$

Table 2.1 – Kernels h_1 to h_6 and their expectations.

Simulations To illustrate this example, we have simulated networks with the Poisson-BEDD model. We have chosen power functions for f and g, i.e. we have set α_f and α_g in $[0, +\infty[$ and $f(u) = (\alpha_f + 1)u^{\alpha_f}$ and $g(v) = (\alpha_g + 1)v^{\alpha_g}$. Therefore, the values of F_2 and G_2 can be set by α_f and α_g. The constant c is set at 0.5, so we have considered square matrices ($m = n$). Figure 2.1 represents the frequency with which 2 confidence intervals, built with respectively equations (2.9) and (2.18) for $\alpha = 0.05$, contain the true value of F_2. The curve associated with V^{δ} suggests that $\sqrt{N}(\widehat{\theta}_N - F_2)$ becomes close to its limiting distribution for $N \gtrsim 250$. For smaller values of N, the frequencies are significantly higher than 0.95, so the confidence intervals are slightly larger than they should. The curve associated with $\widehat{V}^{\delta}_N$ suggests that $\widehat{V}^{\delta}_N$ underestimates V^{δ}, but using Slutsky to plug in $\widehat{V}^{\delta}_N$ for V^{δ} in (2.18) still leads to acceptable frequencies that converge when N grows, especially for $N \gtrsim 250$. Figure 2.2 represents the empirical distribution of $\widehat{\theta}_N$ for different sizes N. It confirms that $\sqrt{N}(\widehat{\theta}_N - F_2)$ converges quickly to a normal distribution with variance V^{δ}.

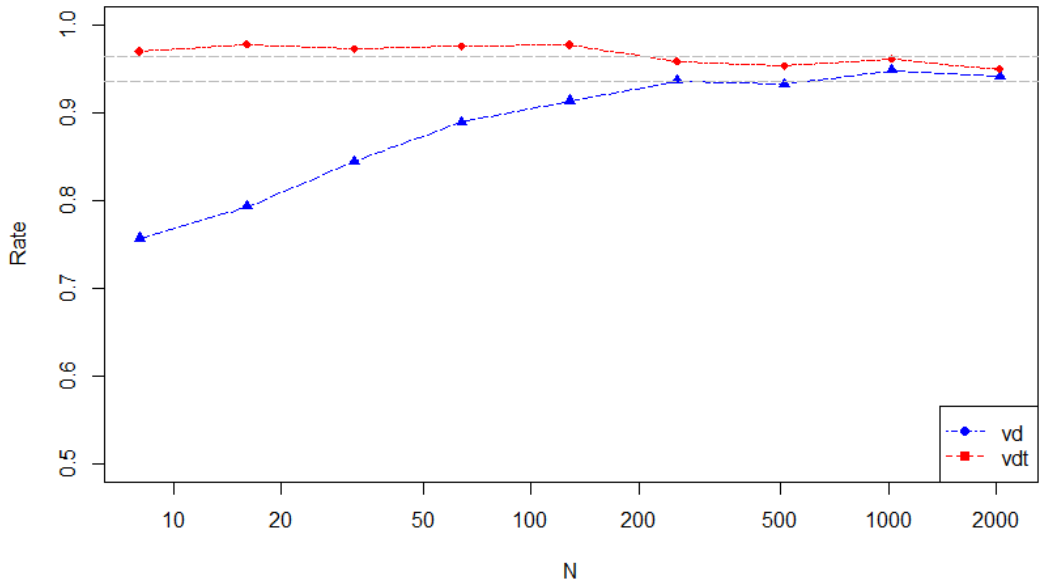

Figure 2.1 – Estimation of F_2: Frequency of the confidence intervals that contain the true value of F_2 for different values of N (on a logarithmic scale). For each $N \in \{8, 16, 32, 64, 128, 256, 512, 1024, 2048\}$, we simulate $K = 1000$ networks with $\lambda = 1$, $F_2 = 3$, $G_2 = 2$. For each simulated network, we estimate F_2 with the estimator $\widehat{\theta}_N$ and at level $1 - \alpha = 0.95$, we build the asymptotic confidence intervals from the weak convergence results: [vdt] built from (2.9) (true value of V^δ) and [vd] built from (2.18) (estimated value of V^δ by $\widehat{V}_N^\delta$). The horizontal dashed lines represent the confidence interval at level 0.95 of the frequency $Z = X/K$, if X follows the binomial distribution with parameters K and α.

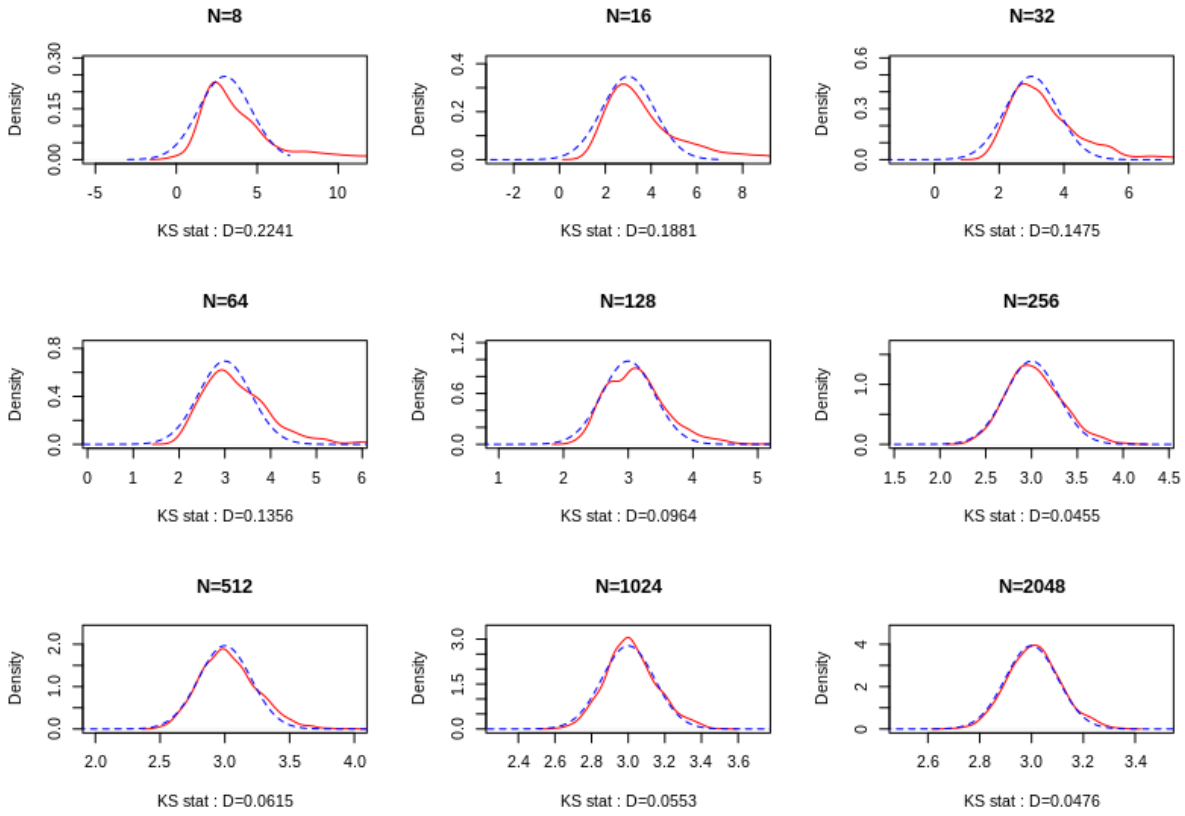

Figure 2.2 – Estimation of F_2: Distribution of $\widehat{\theta}_N$ for different values of N. For each $N \in \{8, 16, 32, 64, 128, 256, 512, 1024, 2048\}$, we simulate $K = 1000$ networks with $\lambda = 1$, $F_2 = 3$, $G_2 = 2$. For each simulated network, we estimate F_2 with the estimator $\widehat{\theta}_N$. The empirical distributions (solid red lines) are interpolated using the density() function from base R stats package. The dashed curves in blue correspond to the normal distribution densities with mean $F_2 = 3$ and variance V^δ/N. Under each plot, the value of the Kolmogorov-Smirnov test statistic D between the empirical distribution of $\widehat{\theta}_N$ and the normal distribution with mean $F_2 = 3$ and variance V^δ/N is given. $D = \sup_x |F_{emp}(x) - F(x)|$ where F_{emp} is the empirical c.d.f. of $\widehat{\theta}_N$ and $F(x)$ the c.d.f. of the normal distribution with mean $F_2 = 3$ and variance V^δ/N.

2.3.3. Network comparison

Some methods have been developed to compare networks. Network statistics, graph spectra, network motifs or graph alignment methods can be used to build a distance (or similarity scores) between two networks (Emmert-Streib et al., 2016; Tantardini et al., 2019). In the context of random networks, fewer comparison methods rely on generative random graph models and they are relatively recent (Asta and Shalizi, 2015; Maugis et al., 2020). A model-based approach offers two advantages. First, by suggesting a distribution on the networks, one might be able to design a distance with known distribution and therefore use statistical tests to compare networks. Second, the use of a generative model makes it considerably easier to interpret, one can use to the model parameters to design a suitable distance to compare the networks, giving insights into the underlying process generating them. Such ability to interpret is particularly interesting in applications such as ecology, where it is crucial to understand how and why the networks differ (Pellissier et al., 2018). In this section, we show how one can extend the usage of U-statistics to network comparison, providing a framework for model-based network comparison.

In the previous example, our analysis has been carried out on a single network. Now, consider two independent networks Y^A and Y^B and we wish to compare their row heterogeneity. Assume that they are respectively generated by the BEDD parameters $\Theta^A = (\lambda^A, f^A, g^A)$ and $\Theta^B = (\lambda^B, f^B, g^B)$. Then, each network is associated with their respective values F_2^A and F_2^B. The data consists in two observed networks $Y^A_{N_A}$ and $Y^B_{N_B}$, which are assumed to be extracted from the first m_A (respectively m_B) rows and n_A (respectively n_B) columns of the infinite matrices Y_A and Y_B. We would like to perform the following test: $\mathcal{H}_0 : F_2^A = F_2^B$ vs. $\mathcal{H}_1 : F_2^A \neq F_2^B$ using the two observed networks.

General method Let $N := N_A + N_B$. Suppose that $N_A/N \xrightarrow[N\to+\infty]{} \rho \in]0,1[$. Then one can simply notice that $\widehat{\delta}_N(Y^A, Y^B) := \widehat{\theta}_{N_A}(Y^A) - \widehat{\theta}_{N_B}(Y^B)$, where $\widehat{\theta}_N(Y)$ is the estimator of Proposition 2.3.5 taken on the matrix Y, is still asymptotically normal from equation (2.9)

$$\sqrt{\frac{N}{V(\Theta^A, \Theta^B)}}\left(\widehat{\delta}_N(Y^A, Y^B) - (F_2^A - F_2^B)\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1),$$

with $V(\Theta^A, \Theta^B) = V^\delta(\Theta^A)/\rho + V^\delta(\Theta^B)/(1-\rho)$ and for BEDD parameters Θ, $V^\delta(\Theta)$ is given by (2.10).

Using the estimators $\widehat{V}_N^\delta$ stemming from the delta-method (2.17), we build $\widehat{V}_N(Y^A, Y^B)$ a consistent estimator for $V(\Theta^A, \Theta^B)$

$$\widehat{V}_N(Y^A, Y^B) = \frac{1}{\rho}\widehat{V}_N^\delta(Y^A) + \frac{1}{1-\rho}\widehat{V}_N^\delta(Y^B).$$

Hence, Slutsky's theorem ensures that

$$\sqrt{\frac{N}{\widehat{V}_N(Y^A, Y^B)}}\left(\widehat{\delta}_N(Y^A, Y^B) - (F_2^A - F_2^B)\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

In this example, we consider the statistical test $\mathcal{H}_0 : F_2^A = F_2^B$ vs. $\mathcal{H}_1 : F_2^A \neq F_2^B$. So we use the test statistic

$$Z_N(Y^A, Y^B) = \sqrt{\frac{N}{\widehat{V}_N(Y^A, Y^B)}}\widehat{\delta}_N(Y^A, Y^B), \tag{2.20}$$

for which Slutsky's theorem applies

$$Z_N(Y^A, Y^B) - \sqrt{\frac{N}{\widehat{V}_N(Y^A, Y^B)}}(F_2^A - F_2^B) \xrightarrow[N\to+\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

Under $\mathcal{H}_0$, $Z_N(Y^A, Y^B) \xrightarrow[N\to+\infty]{\mathcal{D}} \mathcal{N}(0,1)$ which allows us to build asymptotic acceptance intervals for this test at level α with the $(1-\alpha/2)^{th}$ percentile $q_{1-\alpha/2}$ of the standard normal distribution:

$$I(\alpha) = [-q_{1-\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}].$$

Simulations Figure 2.3 shows simulation results for this test. Once again, we consider networks generated by the Poisson-BEDD model with power law functions f and g. To perform the test, we generate couples of observed networks $(Y_{N_A}^A, Y_{N_B}^B)$ with fixed and identical $\lambda^A = \lambda^B$ and $g^A = g^B$. f^A is also fixed, but we let f^B vary by setting the parameter α_{f^B} of the power law, which is used to set F_2^B. The empirical power for this test with varying F_2^B is evaluated for several values of N. It is compared with the asymptotic theoretical power $\psi_N(\Theta^A, \Theta^B)$ for this test. Let $\mu_N(\Theta^A, \Theta^B) := \sqrt{\frac{N}{V(\Theta^A, \Theta^B)}}(F_2^A - F_2^B)$. If a random variable $\tilde{Z}_N$ is such that $\tilde{Z}_N - \mu_N(\Theta^A, \Theta^B) \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1)$, then $\psi_N(\Theta^A, \Theta^B) = \mathbb{P}\left(\tilde{Z}_N \in I(\alpha)\right)$, so it can be computed with

$$\psi_N(\Theta^A, \Theta^B) = F_{(\Theta^A, \Theta^B)}\left(q_{1-\frac{\alpha}{2}}\right) - F_{(\Theta^A, \Theta^B)}\left(-q_{1-\frac{\alpha}{2}}\right) \tag{2.21}$$

where $F_{(\Theta^A, \Theta^B)}(t)$ is the cumulative distribution function of a Gaussian variable with mean $\mu_N(\Theta^A, \Theta^B)$ and variance 1. We notice that the empirical power becomes very close to the asymptotic theoretical power as N grows, which suggests that this test works well for networks with $N \gtrsim 100$.

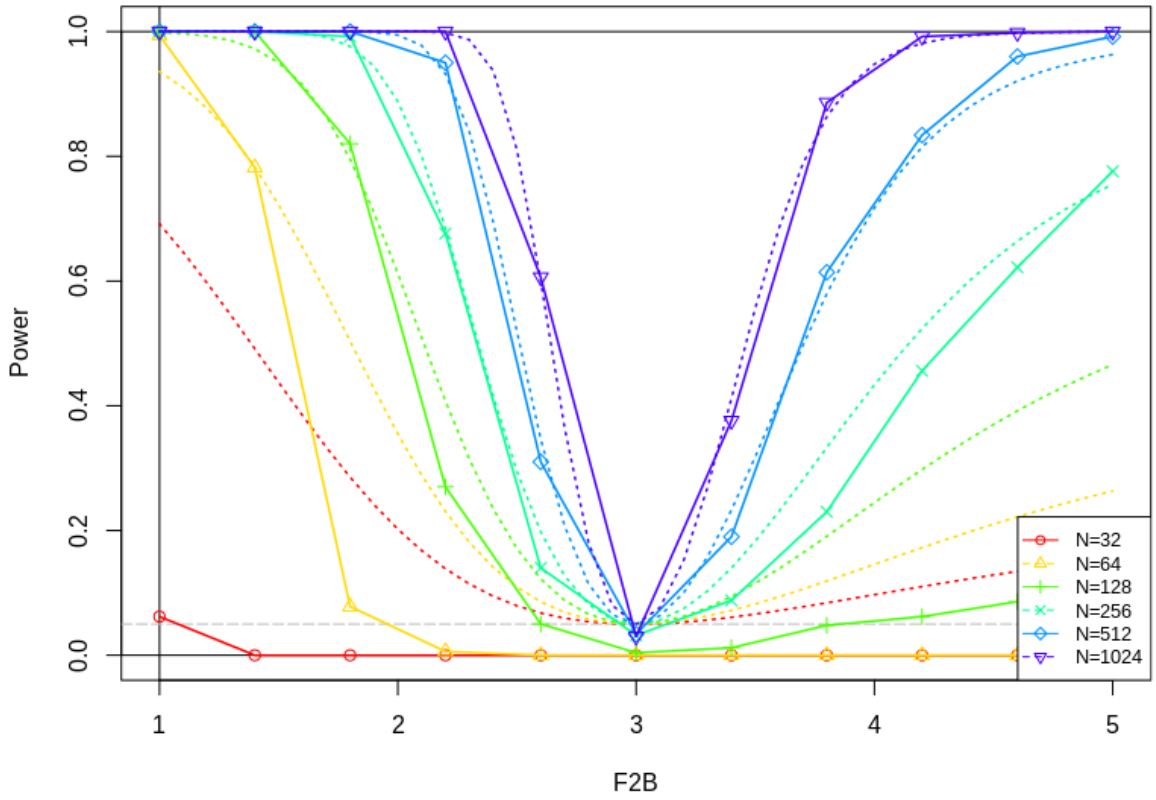

Figure 2.3 – Comparison of F_2 for two networks: Power of the test $\mathcal{H}_0 : F_2^A = F_2^B$ vs. $\mathcal{H}_1 : F_2^A \neq F_2^B$ using the statistic $Z_N(Y^A, Y^B)$ defined by (2.20). We set $\lambda^A = \lambda^B = 1$, $G_2^A = G_2^B = 2$, $c^A = c^B = 0.5$. The value of F_2^A is fixed at 3. Only N and F_2^B will vary. Several values of F_2^B are considered between 1 and 5. For each $N \in \{32, 64, 128, 256, 512, 1024\}$, for each F_2^B, we generate $K = 200$ couple of networks of same size $N^A = N^B = N/2$ with respective F_2 values F_2^A and F_2^B. On each couple of networks (Y^A, Y^B), we compute $Z_N(Y^A, Y^B)$ and we reject the hypothesis $\mathcal{H}_0$ if $Z_N(Y^A, Y^B) \notin I(\alpha)$. The empirical power (solid lines) is the frequency with which the hypothesis is admitted among the K simulations. The theoretical power (dashed lines) is the function $\psi_N(\Theta^A, \Theta^B)$, which only depends on F_2^B since the other parameters are constant, computed with equation (2.21).

2.3.4. Motif frequencies

A motif is a small-size subgraph. The frequencies of occurences of motifs (sometimes called network moments) are widely studied in network theory. Motifs frequencies have known asymptotic distribution under many generative models, so they can be used to analyze binary networks

(Stark, 2001; Picard et al., 2008; Reinert and Röllin, 2010; Bickel et al., 2011; Bhattacharyya and Bickel, 2015; Levin and Levina, 2019; Maugis et al., 2020; Naulet et al., 2021; Ouadah et al., 2022). Many probabilistic graph models also rely on motif frequencies, such as the Exponential Random Graph Model (Frank and Strauss, 1986) or the *dk*-random graphs (Orsini et al., 2015). For many real networks, one can interpret the frequencies of certain motifs, see examples for transcriptional networks (Shen-Orr et al., 2002), protein networks (Pržulj et al., 2004), social networks (Bearman et al., 2004), evolutionary trait networks (Przytycka, 2006), ecological food webs (Bascompte and Melián, 2005; Stouffer et al., 2007), ecological mutualistic networks (Baker et al., 2015; Simmons et al., 2019).

It naturally arises that frequencies of bipartite motifs of size 2×2 can be expressed as quadruplet U-statistics and can be integrated in our framework. If Y is a binary adjacency matrix, then one can count the motifs using a kernel and obtain statistical guarantees. For example, the motif represented in Figure 2.4 can be counted with the kernel

$$h_7(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{4}\Big(Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_1}(1-Y_{i_2j_2}) + Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_2}(1-Y_{i_2j_1}) + Y_{i_1j_1}Y_{i_2j_1}Y_{i_2j_2}(1-Y_{i_1j_2}) + Y_{i_1j_2}Y_{i_2j_1}Y_{i_2j_2}(1-Y_{i_1j_1})\Big).$$

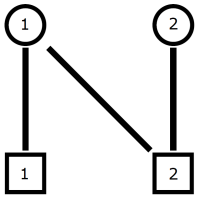

Figure 2.4 – Motif counted by $U_N^{h_7}$. The circles and the squares represent the two types of nodes of a bipartite network. Assuming that the circles correspond to the rows and the squares to the columns of the adjacency matrix, then the submatrix associated to this subgraph is $Y_{(1,2;1,2)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ (figure taken from Ouadah et al., 2022).

Theorem 2.2.7 shows that the associated U-statistic $U_N^{h_7}$ converges to the theoretical frequency T of this motif given the network model and it is asymptotically normal. Suppose $Y \sim$ Bernoulli-BEDD(Θ), where $\Theta = (\lambda, f, g)$ are BEDD parameters, then

$$\sqrt{\frac{N}{V^{h_7}}}\Big(U_N^{h_7} - T(\Theta)\Big) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{2.22}$$

where following derivations given in Lemma 2.H.6, $T(\Theta) = \lambda^3 F_2 G_2 (1 - \lambda F_2 G_2)$ and

$$\begin{aligned} V^{h_7} = &\frac{4\lambda^6}{c} G_2^2 \left[\lambda^2 F_4 F_2^2 G_2^2 - \lambda F_4 F_2 G_2 - \lambda F_3 F_2^2 G_2 + \frac{1}{2} F_3 F_2 + \frac{1}{4} F_4 + \frac{1}{4} F_2^3 \right] \\ &+ \frac{4\lambda^6}{1-c} F_2^2 \left[\lambda^2 G_4 G_2^2 F_2^2 - \lambda G_4 G_2 F_2 - \lambda G_3 G_2^2 F_2 + \frac{1}{2} G_3 G_2 + \frac{1}{4} G_4 + \frac{1}{4} G_2^3 \right] \\ &- \frac{4}{c(1-c)} \left(\lambda^3 F_2 G_2 (1 - \lambda F_2 G_2) \right)^2 . \end{aligned}$$

The quantities λ, $(F_k)_{k\geq 1}$ and $(G_k)_{k\geq 1}$ appearing in the expression of the asymptotic variance V^{h_7} can be consistently estimated using U-statistics of larger subgraphs. For any (p, q), define the kernel $h_{p,q}$ of submatrices $Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}$ of size $p \times q$ as follows:

$$h_{p,q}(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}) = \prod_{u=1}^{p} \prod_{v=1}^{q} Y_{i_u j_v}.$$

Then the U-statistic associated to $h_{p,q}$ is

$$U_N^{h_{p,q}} = \binom{m_N}{p}^{-1} \binom{n_N}{q}^{-1} \sum_{1\leq i_1 < \dots < i_p \leq m_N} \sum_{1 \leq j_1 < \dots < j_q \leq n_N} h_{p,q}(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}).$$

Lemma 2.H.7 states that $\mathbb{E}[h_{1,q}(Y_{(1;1,\dots,q)})] = \lambda^q F_q$ and $\mathbb{E}[h_{p,1}(Y_{(1,\dots,p;1)})] = \lambda^p G_p$. Since Y is a RCE matrix, Kallenberg's law of large number (Lem. 12 of Kallenberg, 1999) applies to these U-statistics and $U_N^{h_{1,q}} \xrightarrow[N\to\infty]{a.s.} \lambda^q F_q$ and $U_N^{h_{p,1}} \xrightarrow[N\to\infty]{a.s.} \lambda^p G_p$.

Using these consistent estimators for λ, and the $(F_k)_{k\geq 1}$ and $(G_k)_{k\geq 1}$, we build an example of consistent estimator for V^{h_7}

$$\begin{aligned} \widehat{V}_N = &\frac{4(U_N^{h_{2,1}})^2}{c} \left[\frac{U_N^{h_{1,4}} (U_N^{h_{1,2}})^2 (U_N^{h_{2,1}})^2}{(U_N^{h_{1,1}})^8} - \frac{U_N^{h_{1,4}} U_N^{h_{1,2}} U_N^{h_{2,1}}}{(U_N^{h_{1,1}})^5} - \frac{U_N^{h_{1,3}} (U_N^{h_{1,2}})^2 U_N^{h_{2,1}}}{(U_N^{h_{1,1}})^6} \right. \\ &\left. + \frac{1}{2} \frac{U_N^{h_{1,3}} U_N^{h_{1,2}}}{(U_N^{h_{1,1}})^3} + \frac{1}{4} \frac{(U_N^{h_{1,4}})^4}{(U_N^{h_{1,1}})^2} + \frac{1}{4} \frac{(U_N^{h_{1,2}})^3}{(U_N^{h_{1,1}})^4} \right] \\ &+ \frac{4(U_N^{h_{1,2}})^2}{1-c} \left[\frac{U_N^{h_{4,1}} (U_N^{h_{2,1}})^2 (U_N^{h_{1,2}})^2}{(U_N^{h_{1,1}})^8} - \frac{U_N^{h_{4,1}} U_N^{h_{2,1}} U_N^{h_{1,2}}}{(U_N^{h_{1,1}})^5} - \frac{U_N^{h_{3,1}} (U_N^{h_{2,1}})^2 U_N^{h_{1,2}}}{(U_N^{h_{1,1}})^6} \right. \\ &\left. + \frac{1}{2} \frac{U_N^{h_{3,1}} U_N^{h_{2,1}}}{(U_N^{h_{1,1}})^3} + \frac{1}{4} \frac{(U_N^{h_{4,1}})^4}{(U_N^{h_{1,1}})^2} + \frac{1}{4} \frac{(U_N^{h_{2,1}})^3}{(U_N^{h_{1,1}})^4} \right] \\ &- \frac{4}{c(1-c)} (U_N^{h_7})^2. \end{aligned}$$

This expression may seem complex at first, however it is computationally simple as one only needs to compute $U_N^{h_{p,1}}$ for $1 \leq p \leq 4$ and $U_N^{h_{1,q}}$ for $1 \leq q \leq 4$ which can be easily done (see Appendix 2.I).

From Slutsky's theorem, it follows that

$$\sqrt{\frac{N}{\widehat{V}_N}}\left(U_N^{h_7} - T(\Theta)\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{2.23}$$

This result can be used to build asymptotic confidence intervals for the motif frequency $T(\Theta)$, like in the previous examples of Sections 2.3.2 and 2.3.3.

Simulations We simulate networks with the Bernoulli-BEDD model, with power law functions for f and g, similar to which of previous examples. For α_f and α_g in $[0, +\infty[$, $f(u) = (\alpha_f + 1)u^{\alpha_f}$ and $g(v) = (\alpha_g + 1)v^{\alpha_g}$. α_f and α_g can be used to set F_2 and G_2. α_f and α_g also determine the maximum value for λ, as we should have $\lambda \leq \lambda_M = (f(1)g(1))^{-1} = (\alpha_f + 1)^{-1}(\alpha_g + 1)^{-1}$. The c constant remains at 0.5.

Figure 2.5 represents the frequency with which the 2 confidence intervals built from equations (2.22) and (2.23) contain the true value of the target motif frequency $T(\Theta)$. We see that as N grows larger than 250, the frequency becomes very close to 0.95, although the variance is still underestimated until $N \approx 2000$. In contrast to the example of Section 2.3.2, the frequencies for N smaller than 16 are very low (0.5 and lower). This is an expected result as the estimator $U_N^{h_7}$ counts the motifs contained in the networks. The maximum number of motifs in the network is $M_N = \binom{m_N}{2}\binom{n_N}{2}$. For a fixed N, $U_N^{h_7}$ can only take discrete values in $\left(kM_N^{-1}\right)_{0\leq k\leq M_N}$. The support of $U_N^{h_7}$ is more and more restricted as N becomes smaller, which makes the empirical distribution of $U_N^{h_7}$ more dissimilar from a Gaussian distribution.

This is also reflected in Figure 2.6 as the discrete support still appears very clearly in the yet smoothed distribution density of $U_N^{h_7}$ for $N = 8$ and $N = 16$. Nevertheless, we see that the empirical distribution converges quickly to a Gaussian distribution, even faster than in the F_2 estimation example of Section 2.3.2, as the Kolmogorov-Smirnov statistics are smaller if N is larger than 32.

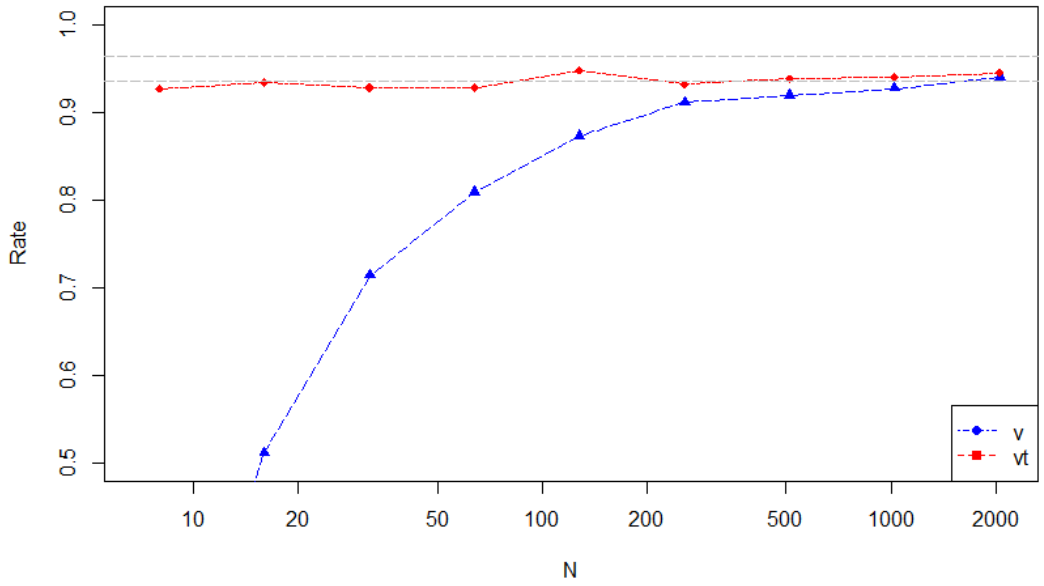

Figure 2.5 – Motif counts: Frequency of the confidence intervals that contain the theoretical value $T(\Theta)$ for different values of N (on a logarithmic scale). For each $N \in \{8, 16, 32, 64, 128, 256, 512, 1024, 2048\}$, we simulate $K = 1000$ networks with $F_2 = 2$, $G_2 = 2$, $\lambda = 0.9\lambda_M$. For each simulated network, we determine the motif frequency with the estimator $U_N^{h_7}$ and at level $1 - \alpha = 0.95$, we build the asymptotic confidence intervals from the weak convergence results: [vt] built from (2.22) (true value of V^{h_7}) and [v] built from (2.23) (estimated value of V^{h_7} by $\widehat{V}_N$). The horizontal dashed lines represent the confidence interval at level 0.95 of the frequency $Z = X/K$, if X follows the binomial distribution with parameters K and α.

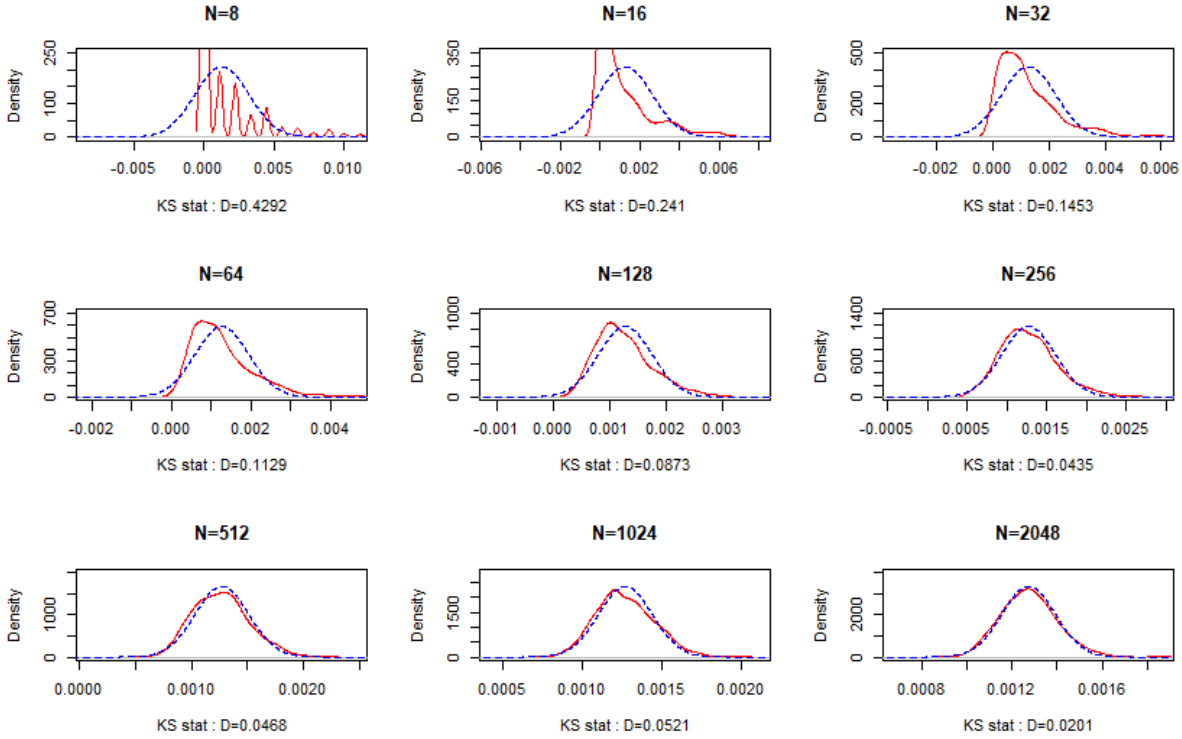

Figure 2.6 – Motif counts: Distribution of $U_N^{h_7}$ for different values of N. For each $N \in \{8, 16, 32, 64, 128, 256, 512, 1024, 2048\}$, we simulate $K = 1000$ networks with $F_2 = 2$, $G_2 = 2$, $\lambda = 0.9\lambda_M$. For each simulated network, we determine the motif frequency with the estimator $U_N^{h_7}$. The empirical distributions (solid red lines) are interpolated using the density() function from base R stats package. The dashed curves in blue correspond to the normal distribution densities with mean $T(\Theta)$ and variance V^{h_7}/N. Under each plot, the value of the Kolmogorov-Smirnov test statistic D between the empirical distribution of $U_N^{h_7}$ and the normal distribution with mean $T(\Theta)$ and variance V^{h_7}/N is given. $D = \sup_x |F_{emp}(x) - F(x)|$ where F_{emp} is the empirical c.d.f. of $U_N^{h_7}$ and $F(x)$ the c.d.f. of the normal distribution with mean $T(\Theta)$ and variance V^{h_7}/N.

2.4. Discussion

In this paper, we have proven a weak convergence result for quadruplet U-statistics over RCE matrices, using a backward martingale approach. We use the Aldous-Hoover representation of RCE matrices and a Hewitt-Savage type argument to extend this result and obtain a CLT in the dissociated case. Using this CLT, we provide a general framework to perform statistical inference on bipartite exchangeable networks through several examples.

Indeed, U-statistics can be used to build estimators. The advantage of taking quadruplets is to define functions over several interactions of the same row or column. This allows us to extract information on the row and column distribution. The CLT then guarantees an asymptotic normality result of the estimators, where the only unknown is their asymptotic variances, which have to be estimated then plugged in with Slutsky's theorem.

Computational cost One interesting feature of the kernels chosen in Section 2.3 is their computational simplicity. This simplicity comes naturally when considering quadruplet kernels consisting of small products. Indeed, if we denote $Y_N := (Y_{ij})_{1\leq i\leq m_N, 1\leq j\leq n_N}$, one can write $U_N^{h_1}$ and $U_N^{h_2}$ used in the F_2 estimation example (Sect. 2.3.2 and 2.3.3) as

$$\begin{aligned} U_N^{h_1} &= \frac{1}{n_N m_N (m_N-1)} \left[|Y_N^T Y_N|_1 - \mathrm{Tr}(Y_N^T Y_N) \right] \\ U_N^{h_2} &= \frac{1}{m_N(m_N-1) n_N (n_N-1)} \left[(|Y_N|_1)^2 - |Y_N^T Y_N|_1 + \mathrm{Tr}(Y_N^T Y_N) \right. \\ &\quad \left. - |Y_N Y_N^T|_1 + \mathrm{Tr}(Y_N Y_N^T) - |Y_N^{\odot 2}|_1 \right]. \end{aligned} \tag{2.24}$$

where Tr is the trace operator. We see that $U_N^{h_1}$ and $U_N^{h_2}$ can be computed using only basic operations on matrices, which are optimized in most computing software. This can also be said for all the other U-statistics used in this example, and by extension for the estimators $\widehat{\theta}_N$ and $\widehat{V}_N$. Expressions for the remaining U-statistics are given in Appendix 2.I.

In the motif count example (Sect. 2.3.4), the U-statistic $U_N^{h_7}$ can also be easily computed, despite the $h_{p,q}$ being kernels over submatrices larger than a quadruplet (at least one dimension greater than 2). The $U_N^{h_{p,q}}$ U-statistics normally involve more complex summations but fortunately, we show in Appendix 2.I that simpler expressions can be found for $p = 1$ or $q = 1$.

Other models: graphons We have seen that for a class of BEDD models (those falling under Asm. 2.3.3), the quadruplet U-statistics are particularly interesting because a single quadruplet contains all the information of the model. The Bernoulli-BEDD used in Section 2.3.4 is an example of model where this assumption does not hold. Still, one can build estimators, apply

Theorem 2.2.7 and perform statistical inference on this model, like in Section 2.3.4. In fact, the only conditions on the model are that it should be RCE and dissociated, i.e. it can be written as a bipartite W-graph model (see Sect. 2.2.3). For example, given the W-graph model $Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{P}(\lambda w(\xi_i, \eta_j))$ with $\iint w = 1$, one could have tested if it is of product form, i.e. if w can be written as $w(u,v) = f(u)g(v)$ (as in the BEDD models). An appropriate kernel for this test would be

$$\begin{aligned} h(Y_{(i_1,i_2;j_1,j_2)}) =& \frac{1}{4} Y_{i_1j_1}Y_{i_2j_2}(Y_{i_1j_1} + Y_{i_2j_2} - Y_{i_1j_2} - Y_{i_1j_2} - 2) \\ &+ \frac{1}{4} Y_{i_1j_2}Y_{i_2j_1}(Y_{i_1j_2} + Y_{i_2j_1} - Y_{i_1j_1} - Y_{i_2j_2} - 2) \end{aligned}$$

as $\mathbb{E}[h(Y_{(i_1,i_2;j_1,j_2)})] = \iint w(u,v)(w(u,v) - f(u)g(v))dudv$ with $f(u) = \int w(u,v)dv$ and $g(v) = \int w(u,v)du$ and should be equal to 0 if the hypothesis is true.

Extension to larger subgraphs It is legitimate to wonder if one can extend our framework to U-statistics over submatrices of size different from 2×2, for example $Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}$ of size $p \times q$, i.e.

$$U_N^h = \left[\binom{m_N}{p}\binom{n_N}{q}\right]^{-1} \sum_{1 \le i_1 < \dots < i_p \le m_N} \sum_{1 \le j_1 < \dots < j_q \le n_N} h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}).$$

Such generalization opens up many possibilities by building new estimators.

First, as seen in Section 2.3.4, in the Bernoulli-BEDD model, the quantities F_k and G_k cannot be retrieved by a quadruplet for $k \ge 3$, but F_k can be retrieved with subgraphs of size $1 \times k$ and G_k with subgraphs of size $k \times 1$. Second, our framework can also be used to count motifs of size larger that 2×2, since the maximum size of the motifs is determined by the size of the kernel. Finally, in the row heterogeneity example where we used formula (2.11) to derive an asymptotic confidence interval for F_2, we notice that one could have estimated the term $\lambda^4 F_4$ appearing in V with a kernel over submatrices of size 1×4 such as $h(Y_{(i_1;j_1,j_2,j_3,j_4)}) = Y_{i_1j_1}Y_{i_1j_2}Y_{i_1j_3}Y_{i_1j_4}$ and $\mathbb{E}[h(Y_{(i_1;j_1,j_2,j_3,j_4)})] = \lambda^4 F_4$.

Actually, our theorem can be extended to U-statistics over larger subgraphs under similar conditions. All the steps of our proof can be adapted to U-statistics of larger subgraphs. These U-statistics are indeed backward martingales and the equivalent of Propositions 2.D.1 and 2.E.1 require more calculus. As a consequence, the asymptotic variance also has a different expression. On the one hand, such an extension would allow more flexibility in the choice of the kernel, hence the ability to build more complex estimators that are asymptotically normal. On the other hand, in practice, the computation of such U-statistics may also be more complex and computationally demanding, whereas simple functions on quadruplets can easily be expressed with matrix operations.

Degeneracy Degenerate cases are of interest because they are rather common. The degeneracy depends on the kernels and the distribution of Y. As an example, assume that one is interested to test $\mathcal{H}_0 : f \equiv 1$ vs. $\mathcal{H}_1 : f \not\equiv 1$ for a Poisson-BEDD model. Under $\mathcal{H}_0$, $F_2 = 1$ whereas under $\mathcal{H}_1$, $F_2 > 1$. We plan to use the same estimator of F_2 than in Section 2.3.2. Equation (2.16) of the second method could be also obtained applying Theorem 2.2.7 to the kernel $h = h_1 - F_2 h_2$. From equation (2.10), we see that from that the asymptotic variance $V^t = V^\delta = 0$ under $\mathcal{H}_0$, since $F_2 = F_3 = F_4 = 1$. Thus, under $\mathcal{H}_0$, this is a degenerate case and Theorem 2.2.7 does not apply and the limiting distribution of a test statistic using this estimator is unidentified.

Theorems 2.2.5 and 2.2.7 avoid degeneracy by deliberately excluding the case where $V = 0$ almost surely. However, these theorems would remain valid in degenerate cases. Indeed, if $V = 0$ almost surely, then Theorems 2.2.5 and 2.2.7 would yield $\sqrt{N}(U_N^h - U_\infty^h) \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

This can be proven to be true. First, notice that from Corollary 2.H.2, $\mathbb{V}[U_N^h \mid \mathcal{F}_\infty] = V/N + o(1/N)$, therefore

$$\begin{aligned}\mathbb{V}[\sqrt{N}(U_N^h - U_\infty^h)] &= N\mathbb{E}[\mathbb{V}[U_N^h - U_\infty^h \mid \mathcal{F}_\infty]] + N\mathbb{V}[\mathbb{E}[U_N^h - U_\infty^h \mid \mathcal{F}_\infty]] \\ &= N\mathbb{E}[\mathbb{V}[U_N^h \mid \mathcal{F}_\infty]] \\ &= \mathbb{E}[V] + o(1)\end{aligned}$$

where we denote $\mathbb{V}[X]$ the variance of a random variable X and we used the fact that $\mathrm{Cov}(U_N^h, U_\infty^h) = \mathrm{Cov}(U_N^h, \mathbb{E}[U_N^h \mid \mathcal{F}_\infty]) = \mathbb{V}[\mathbb{E}[U_N^h \mid \mathcal{F}_\infty]] = \mathbb{V}[U_\infty^h]$. If $V = 0$ almost surely, then $\mathbb{V}[\sqrt{N}(U_N^h - U_\infty^h)] = o(1)$. By Chebyshev's inequality, we get $\sqrt{N}(U_N^h - U_\infty^h) \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Austern and Orbanz (2022) also comes to this conclusion if $\eta^2 = 0$ in their Theorem 17. However, they do not explicitly discuss the implications of this case.

In fact, if $V = 0$ a.s., the U-statistic is degenerate and the rate of convergence of $U_N^h - U_\infty^h$ is faster than $\sqrt{N}$. This behaviour is similar to regular U-statistic of i.i.d. variables as described by Lee (1990) or Arcones and Gine (1992). In the proof of Lemma 2.H.1, one could go further in the derivation of the covariance and developed more the content of the $o(1/N)$ term. This would have yielded a decomposition of the form:

$$\mathbb{V}[U_N^h \mid \mathcal{F}_\infty] = \frac{V^{(1)}}{N} + \frac{V^{(2)}}{N^2} + \frac{V^{(3)}}{N^3} + \frac{V^{(4)}}{N^4} + o\left(\frac{1}{N^4}\right),$$

where $V^{(1)} = V$ of Theorem 2.2.5 and $V^{(2)}$, $V^{(3)}$ and $V^{(4)}$ are non-negative $\mathcal{F}_\infty$-measurable random variables. The derivation of closed-form expressions for $V^{(2)}$, $V^{(3)}$ and $V^{(4)}$ is possible but out of scope.

If $V = V^{(1)} = 0$ a.s. but $\mathbb{P}(V^{(2)} > 0) > 0$, then we say that the U-statistic is degenerate of order 1 and the above formula indicates that the right normalization is $N(U_N^h - U_\infty^h)$ instead of

$\sqrt{N}(U_N^h - U_\infty^h)$. We can generalize this intuition as follows: for $2 \le d \le 4$, if $V^{(d')} = 0$ a.s. for all $1 \le d' \le d-1$ and $\mathbb{P}(V^{(d)} > 0) > 0$, then we say that the U-statistic is degenerate of order $d-1$ and $N^{\frac{d}{2}}(U_N^h - U_\infty^h)$ converges in distribution to some random variable. However, the asymptotic distribution for degenerate U-statistics is not trivial in general. Even for U-statistics of i.i.d. variables, the limit is very dependent of the kernel h and it involves combinations of products of independent Gaussian variables in a form that is not always tractable (Rubin and Vitale, 1980; Lee, 1990).

Further work might be carried out to investigate the degenerate cases. One lead is to derive some Hoeffding-type decomposition (see for example Chiang et al., 2021 for jointly and separately exchangeable arrays, Wang et al., 2015 for kernels of size 2) but for quadruplet kernels taken on RCE matrices. Hoeffding-type decompositions can help identify the limiting distribution of degenerate U-statistics, as shown by Lee (1990) and Arcones and Gine (1992) in the i.i.d. case.

Berry-Esseen Further studies might be carried out to investigate the rate of convergence of $\sqrt{N}(U_N^h - U_\infty^h)$ to its limiting distribution. For specific applications, one can for now rely on simulations to assess how quickly it converges. A possible direction to find theoretical guarantees is the derivation of a Berry-Esseen-type bound, similar to Austern and Orbanz (2022) in their limit theorems.

Choice of the optimal kernels It is simple to design new U-statistics using various kernels. So when it comes to estimate some particular parameter, one may have the choice between several kernels. Even though they have the same expectation, not only the asymptotic variances might differ, but their rates of convergence to their asymptotic distributions can also vary. In addition, one should remain careful that the derived U-statistics are easily computable. In conclusion, each kernel does not necessarily lead to the same statistical and computational guarantees. The art of designing the best estimation procedures or statistical tests using our approach relies on finding the optimal kernels, depending on the situation.

Appendix 2.A Properties of m_N and n_N

In this appendix, we provide the proofs for Proposition 2.2.2 and further properties of the sequences m_N and n_N defined as $m_N = 2 + \lfloor c(N+1) \rfloor$ and $n_N = 2 + \lfloor (1-c)(N+1) \rfloor$ for all $N \ge 1$, where c is an irrational number (Definition 2.2.1).

Proof of Proposition 2.2.2. The second result stems from the fact that

$$m_N + n_N = 4 + \lfloor c(N+1) \rfloor + \lfloor (1-c)(N+1) \rfloor = 4 + \lfloor c(N+1) \rfloor + \lfloor -c(N+1) \rfloor + N + 1$$

and $\lfloor c(N+1) \rfloor + \lfloor -c(N+1) \rfloor = -1$ because $c(N+1)$ is not an integer since c is irrational. Then, the first result simply follows as

$$\frac{m_N}{m_N + n_N} = \frac{\lfloor c(N+1) \rfloor + 2}{N+4} \underset{N}{\sim} \frac{c(N+1)+2}{N+4} \underset{N}{\sim} \frac{cN}{N},$$

where $\underset{N}{\sim}$ denotes the asymptotic equivalence when N grows to infinity, i.e. $a_N \underset{N}{\sim} b_N$ if and only if $a_N / b_N \underset{N \to \infty}{\longrightarrow} 1$. □

Proof of Corollary 2.2.3. As m_N and n_N are non-decreasing, the corollary is a direct consequence of $m_N + n_N = 4 + N$, because then $m_{N+1} + n_{N+1} = 4 + N + 1 = m_N + n_N + 1$. □

The following definition and proposition pertain to the partition of $\mathbb{N}^*$ which will be helpful in later proofs.

Definition 2.A.1. *We define $\mathcal{B}_c$ and $\mathcal{B}_{1-c}$ two complementary subsets of $\mathbb{N}^*$ as*

$$\mathcal{B}_c = \{N \in \mathbb{N}^* : m_N = m_{N-1} + 1\} \quad \text{and} \quad \mathcal{B}_{1-c} = \{N \in \mathbb{N}^* : n_N = n_{N-1} + 1\}.$$

Proposition 2.A.2. *Set $\kappa_c(m) := \left\lfloor \frac{m-2}{c} \right\rfloor$ and $\kappa_{1-c}(n) := \left\lfloor \frac{n-2}{1-c} \right\rfloor$. If $N \in \mathcal{B}_c$, then $N = \kappa_c(m_N)$. Similarly, if $N \in \mathcal{B}_{1-c}$, then $N = \kappa_{1-c}(n_N)$.*

Proof. Remember that c is an irrational number, so if $N \in \mathcal{B}_c$, then

$$cN + 2 < \lfloor cN \rfloor + 3 = m_{N-1} + 1 = m_N = \lfloor c(N+1) \rfloor + 2 < c(N+1) + 2,$$

which means that $\frac{m_N - 2}{c} - 1 < N < \frac{m_N - 2}{c}$, thus $N = \left\lfloor \frac{m_N - 2}{c} \right\rfloor$. □

Appendix 2.B Backward martingales

In this appendix, we recall the definition of decreasing filtrations, backward martingales and their convergence theorem. The proof of Theorem 2.B.3 can be found in Doob (1953), Section 7, Theorem 4.2.

Definition 2.B.1. *A decreasing filtration is a decreasing sequence of σ-fields $\mathcal{F} = (\mathcal{F}_n)_{n \geq 1}$, i.e. such that for all $n \geq 1$, $\mathcal{F}_{n+1} \subset \mathcal{F}_n$.*

Definition 2.B.2. *Let $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$ be a decreasing filtration and $M = (M_n)_{n\geq 1}$ a sequence of integrable random variables adapted to $\mathcal{F}$. $(M_n, \mathcal{F}_n)_{n\geq 1}$ is a backward martingale if and only if for all $n \geq 1$, $\mathbb{E}[M_n \mid \mathcal{F}_{n+1}] = M_{n+1}$.*

Theorem 2.B.3. *Let $(M_n, \mathcal{F}_n)_{n\geq 1}$ be a backward martingale. Then, $(M_n)_{n\geq 1}$ is uniformly integrable, and, denoting $M_\infty = \mathbb{E}[M_1 \mid \mathcal{F}_\infty]$ where $\mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \mathcal{F}_n$, we have*

$$M_n \xrightarrow[n\to\infty]{a.s., L_1} M_\infty.$$

Furthermore, if $(M_n)_{n\geq 1}$ is square-integrable, then $M_n \xrightarrow[n\to\infty]{L_2} M_\infty$.

Appendix 2.C Square-integrable backward martingale

In this appendix, we prove Proposition 2.C.1, which states that U_N^h is a square-integrable backward martingale.

Proposition 2.C.1. *Let Y be a RCE matrix. Let h be a quadruplet kernel such that $\mathbb{E}[h(Y_{(1,2;1,2)})^2] < \infty$. Let $\mathcal{F}_N = \sigma\big((U_{k,l}^h, k \geq m_N, l \geq n_N)\big)$ and $\mathcal{F}_\infty = \bigcap_{N=1}^{\infty} \mathcal{F}_N$. Set $U_\infty^h := \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty]$. Then $(U_N^h, \mathcal{F}_N)_{N\geq 1}$ is a square-integrable backward martingale and $U_N^h \xrightarrow[N\to\infty]{a.s., L_2} U_\infty^h = \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_\infty]$.*

The proof relies on the following lemma.

Lemma 2.C.2. *For all $1 \leq i_1 < i_2 \leq m_N$ and $1 \leq j_1 < j_2 \leq n_N$, $\mathbb{E}[h(Y_{(i_1,i_2;j_1,j_2)}) \mid \mathcal{F}_N] = \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_N]$.*

Proof. In the proof of this lemma, we specify the matrices over which the U-statistics are taken, i.e. we denote $U_{k,l}^h(Y)$ instead of $U_{k,l}^h$ the U-statistic of kernel h and of size $k \times l$ taken on Y.

By construction, for all $k \geq m_N, l \geq n_N$, for all matrix permutations $\Phi \in \mathbb{S}_{m_N} \times \mathbb{S}_{n_N}$ (only acting on the first m_N rows and n_N columns), we have $U_{k,l}^h(\Phi Y) = U_{k,l}^h(Y)$. Moreover, since Y is RCE, we also have $\Phi Y \overset{\mathcal{D}}{=} Y$. Therefore,

$$\Phi Y \mid (U_{k,l}^h(Y), k \geq m_N, l \geq n_N) \overset{\mathcal{D}}{=} Y \mid (U_{k,l}^h(Y), k \geq m_N, l \geq n_N).$$

That means that conditionally on $\mathcal{F}_N$, the first m_N rows and n_N columns of Y are exchangeable and the result to prove follows from this. □

Proof of Proposition 2.C.1. First, we remark that as $\mathbb{E}[h(Y_{(1,2;1,2)})^2] < \infty$, then for all N, $\mathbb{E}[(U_N^h)^2] < \infty$. Thus, the $(U_N^h)_{N\geq 1}$ are square-integrable. Second, $\mathcal{F} = (\mathcal{F}_N)_{N\geq 1}$ is a decreasing filtration and for all N, U_N^h is $\mathcal{F}_N$-measurable.

Now using lemma 2.C.2, we have for all $K \le N$,

$$\begin{aligned}\mathbb{E}[U_K^h \mid \mathcal{F}_N] &= \binom{m_K}{2}^{-2}\binom{n_K}{2}^{-2} \sum_{\substack{1\le i_1<i_2\le m_K\\ 1\le j_1<j_2\le n_K}} \mathbb{E}[h(Y_{(i_1,i_2;j_1,j_2)}) \mid \mathcal{F}_N] \\ &= \binom{m_K}{2}^{-2}\binom{n_K}{2}^{-2} \sum_{\substack{1\le i_1<i_2\le m_K\\ 1\le j_1<j_2\le n_K}} \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_N] \\ &= \mathbb{E}[h(Y_{(1,2;1,2)}) \mid \mathcal{F}_N],\end{aligned}$$

In particular, $\mathbb{E}[U_{N-1}^h \mid \mathcal{F}_N] = \mathbb{E}[U_N^h \mid \mathcal{F}_N] = U_N^h$, which concludes the proof that $(U_N^h, \mathcal{F}_N)_{N\ge1}$ is a square-integrable backward martingale. Finally, Theorem 2.B.3 ensures that $U_N^h \xrightarrow[N\to\infty]{a.s.,L_2} U_\infty^h$. □

Appendix 2.D Asymptotic variance

We prove Proposition 2.D.1 which gives the convergence and an expression for the asymptotic variance. The proof involves some tedious calculations. Before that, we introduce some notations to make the proof of Proposition 2.D.1 more readable.

Notation. In this appendix and in Appendix 2.E, we denote

$$\begin{aligned} X_{[i_1,i_2;j_1,j_2]} &:= h(Y_{\{i_1,i_2;j_1,j_2\}}), & Z_{NK} &:= \sqrt{N}(U_K - U_{K+1}), \\ S_{NK} &:= \mathbb{E}[Z_{NK}^2 \mid \mathcal{F}_{K+1}], & V_N &:= \sum_{K=N}^{\infty} S_{NK}. \end{aligned} \tag{2.25}$$

The exchangeability of Y implies that $\mathbb{E}[X_{\{i_1,i_2;j_1,j_2\}}X_{\{i_1',i_2';j_1',j_2'\}} \mid \mathcal{F}_K]$ only depends on the numbers of rows and columns shared by both $\{i_1,i_2;j_1,j_2\}$ and $\{i_1',i_2';j_1',j_2'\}$. For $0 \le p \le 2$ and $0 \le q \le 2$, we set

$$c_K^{(p,q)} := \mathbb{E}[X_{\{i_1,i_2;j_1,j_2\}}X_{\{i_1',i_2';j_1',j_2'\}} \mid \mathcal{F}_K],$$

and

$$c_\infty^{(p,q)} := \mathbb{E}[X_{\{i_1,i_2;j_1,j_2\}}X_{\{i_1',i_2';j_1',j_2'\}} \mid \mathcal{F}_\infty],$$

where they share p rows and q columns.

Proposition 2.D.1. *Let $(V_N)_{N\ge1}$ be as defined in (2.25). Then, under the hypotheses of Theorem 2.2.5, we have*

$$V_N \xrightarrow[N\to\infty]{\mathbb{P}} V = 4c^{-1}(c_\infty^{(1,0)} - (U_\infty^h)^2) + 4(1-c)^{-1}(c_\infty^{(0,1)} - (U_\infty^h)^2).$$

The proof of Proposition 2.D.1 will be based on the following five lemmas.

Lemma 2.D.2. *If $K \in \mathcal{B}_c$ (see Definition 2.A.1), then*

$$Z_{N,K-1} = \sqrt{N}\frac{2}{m_K - 2}(U_K^h - \delta_K),$$

where

$$\delta_K = (m_K - 1)^{-1}\binom{n_K}{2}^{-1} \sum_{\substack{1 \le i_1 \le m_K - 1 \\ 1 \le j_1 < j_2 \le n_K}} X_{\{i_1, m_K; j_1, j_2\}}.$$

Proof. Observe that

$$\sum_{\substack{1 \le i_1 < i_2 \le m_K \\ 1 \le j_1 < j_2 \le n_K}} X_{\{i_1,i_2;j_1,j_2\}} = \sum_{\substack{1 \le i_1 < i_2 \le m_K - 1 \\ 1 \le j_1 < j_2 \le n_K}} X_{\{i_1,i_2;j_1,j_2\}} + \sum_{\substack{1 \le i_1 \le m_K - 1 \\ 1 \le j_1 < j_2 \le n_K}} X_{\{i_1,m_K;j_1,j_2\}}. \tag{2.26}$$

But if $K \in \mathcal{B}_c$, then $m_{K-1} = m_K - 1$ and $n_{K-1} = n_K$. Therefore, equation (2.26) is equivalent to

$$\binom{m_K}{2}\binom{n_K}{2}U_K^h = \binom{m_K - 1}{2}\binom{n_K}{2}U_{K-1}^h + (m_K - 1)\binom{n_K}{2}\delta_K,$$

so

$$U_{K-1}^h = \frac{1}{m_K - 2}\left(m_K U_K^h - 2\delta_K\right).$$

This concludes the proof since $Z_{N,K-1} = \sqrt{N}(U_{K-1}^h - U_K^h)$. □

We now calculate S_{NK} in the following lemmas.

Lemma 2.D.3. *For all $0 \le p \le 2$ and $0 \le q \le 2$, $c_N^{(p,q)} \xrightarrow[N\to\infty]{a.s., L_1} c_\infty^{(p,q)}$.*

Proof. This follows from the fact that $(c_N^{(p,q)}, \mathcal{F}_N)_{N \ge 1}$ is a backward martingale. □

Lemma 2.D.4. *If $K \in \mathcal{B}_c$ (see Definition 2.A.1), then*

$$S_{N,K-1} = 4N\left(\frac{(n_K - 2)(n_K - 3)}{(m_K - 1)(m_K - 2)n_K(n_K - 1)}c_K^{(1,0)} - \frac{1}{(m_K - 2)^2}(U_K^h)^2 + \psi(K)\right),$$

where ψ does not depend on N and $\psi(K) = o(m_K^{-2})$.

Proof. Because of Lemma 2.D.2 and the $\mathcal{F}_K$-measurability of U_K^h,

$$S_{N,K-1} = \frac{4N}{(m_K - 2)^2}(\mathbb{E}[\delta_K^2 \mid \mathcal{F}_K] + (U_K^h)^2 - 2U_K^h\mathbb{E}[\delta_K \mid \mathcal{F}_K]).$$

First, Lemma 2.C.2 implies that

$$\mathbb{E}[\delta_K \mid \mathcal{F}_K] = U_K^h.$$

Then, we can calculate

$$\mathbb{E}[\delta_K^2 \mid \mathcal{F}_K] = (m_K-1)^{-2}\binom{n_K}{2}^{-2} \sum_{\substack{1\le i_1\le m_K-1\\ 1\le j_1<j_2\le n_K}} \sum_{\substack{1\le i_1'\le m_K-1\\ 1\le j_1'<j_2'\le n_K}} \mathbb{E}[X_{\{i_1,m_K;j_1,j_2\}}X_{\{i_1',m_K;j_1',j_2'\}} \mid \mathcal{F}_K].$$

Each term of the sum only depends on the number of rows and columns the quadruplets in $X_{\{i_1,m_K;j_1,j_2\}}$ and $X_{\{i_1',m_K;j_1',j_2'\}}$ have in common. For example, if they share p rows and q columns, it is equal to $c_K^{(p,q)}$. So by breaking down the different cases for p and q, we may count the number of possibilities. For example, if $(p,q)=(1,2)$, then the number of possibilities is $(m_K-1)(m_K-2)\binom{n_K}{2}$. This gives

$$\begin{aligned}\mathbb{E}[\delta_K^2 \mid \mathcal{F}_K] = (m_K-1)^{-1}\binom{n_K}{2}^{-1}\Big\{&\frac{1}{2}(m_K-2)(n_K-2)(n_K-3)c_K^{(1,0)} + 2(m_K-2)(n_K-2)c_K^{(1,1)}\\ &+ (m_K-2)c_K^{(1,2)} + \frac{1}{2}(n_K-2)(n_K-3)c_K^{(2,0)} + 2(n_K-2)c_K^{(2,1)} + c_K^{(2,2)}\Big\}.\end{aligned}$$

Finally, setting

$$\begin{aligned}\psi(K) := (m_K-1)^{-3}\binom{n_K}{2}^{-1}\Big\{&2(m_K-2)(n_K-2)c_K^{(1,1)} + (m_K-2)c_K^{(1,2)}\\ &+ \frac{1}{2}(n_K-2)(n_K-3)c_K^{(2,0)} + 2(n_K-2)c_K^{(2,1)} + c_K^{(2,2)}\Big\},\end{aligned}$$

we obtain the desired result, with $\psi(K)=o(m_K^{-2})$ since $\frac{m_K}{c} \underset{K}{\sim} \frac{n_K}{1-c} \underset{K}{\sim} K$. □

Remark. In the case where $K \in \mathcal{B}_{1-c}$, the equivalent formulas to those of Lemmas 2.D.2 and 2.D.4 are derived from similar proofs. If $K \in \mathcal{B}_{1-c}$, then

$$Z_{N,K-1} = \sqrt{N}\frac{2}{n_K-2}(U_K^h - \gamma_K),$$

where

$$\gamma_K = (n_K-1)^{-1}\binom{m_K}{2}^{-1} \sum_{\substack{1\le i_1<i_2\le m_K\\ 1\le j_1\le n_K-1}} X_{\{i_1,i_2;j_1,n_K\}},$$

and

$$S_{N,K-1} = 4N\left(\frac{(m_K-2)(m_K-3)}{(n_K-1)(n_K-2)m_K(m_K-1)}c_K^{(0,1)} - \frac{1}{(n_K-2)^2}(U_K^h)^2 + \varphi(K)\right),$$

where φ does not depend on N and $\varphi(K)=o(n_K^{-2})$.

Lemma 2.D.5. *Let $(R_n)_{n\ge1}$ be a sequence of random variables and $(\lambda_n)_{n\ge1}$ a sequence of real positive numbers. Set $C_n := n\sum_{k=n}^{\infty}\lambda_k R_k$. If*

- *$n\sum_{k=n}^{\infty}\lambda_k \xrightarrow[n\to\infty]{} 1$, and*
- *there exists a random variable R_∞ such that $R_n \xrightarrow[n\to\infty]{a.s.} R_\infty$,*

then $C_n \xrightarrow[n\to\infty]{a.s.} R_\infty$. Furthermore, if $R_n \xrightarrow[n\to\infty]{L_1} R_\infty$, then $C_n \xrightarrow[n\to\infty]{L_1} R_\infty$.

Proof. Notice that

$$\begin{aligned}|C_n - R_\infty| &= \Big|n\sum_{k=n}^{\infty}\lambda_k R_k - R_\infty\Big| \\ &\le \Big|n\sum_{k=n}^{\infty}\lambda_k R_k - n\sum_{k=n}^{\infty}\lambda_k R_\infty\Big| + \Big|n\sum_{k=n}^{\infty}\lambda_k R_\infty - R_\infty\Big| \\ &\le \left(n\sum_{k=n}^{\infty}\lambda_k\right)\times\sup_{k\ge n}|R_k - R_\infty| + \Big|n\sum_{k=n}^{\infty}\lambda_k - 1\Big|\times|R_\infty|.\end{aligned}$$

If $n\sum_{k=n}^{\infty}\lambda_k \xrightarrow[n\to\infty]{} 1$ and $R_n \xrightarrow[n\to\infty]{a.s.} R_\infty$, then for all ω fixed except a set of neglectable size, $C_n(\omega) \xrightarrow[n\to\infty]{} R_\infty(\omega)$, which gives the a.s. convergence. Now, consider also that

$$\begin{aligned}\mathbb{E}\Big[|C_n - R_\infty|\Big] &\le n\sum_{k=n}^{\infty}\lambda_k\mathbb{E}\Big[|R_k - R_\infty|\Big] + \Big|n\sum_{k=n}^{\infty}\lambda_k - 1\Big|\mathbb{E}\Big[|R_\infty|\Big] \\ &\le \left(n\sum_{k=n}^{\infty}\lambda_k\right)\times\sup_{k\ge n}\mathbb{E}\Big[|R_k - R_\infty|\Big] + \Big|n\sum_{k=n}^{\infty}\lambda_k - 1\Big|\mathbb{E}\Big[|R_\infty|\Big].\end{aligned}$$

So if $R_n \xrightarrow[n\to\infty]{L_1} R_\infty$, then $\mathbb{E}\big[|R_n - R_\infty|\big] \xrightarrow[n\to\infty]{L_1} 0$ and $\sup_{k\ge n}\mathbb{E}\big[|R_k - R_\infty|\big] \xrightarrow[n\to\infty]{L_1} 0$. Since $n\sum_{k=n}^{\infty}\lambda_k \xrightarrow[n\to\infty]{} 1$, the first term converges to 0, and the second term too because $\mathbb{E}\big[|R_\infty|\big] < \infty$. Finally, $\mathbb{E}\big[|C_n - R_\infty|\big] \xrightarrow[n\to\infty]{} 0$. □

Lemma 2.D.6. *Let $(Q_n)_{n\ge1}$ be a sequence of random variables. Set $C_n := n\sum_{k=n}^{\infty}Q_k$. If there exists a random variable C_∞ such that $n^2Q_n \xrightarrow[n\to\infty]{a.s.} C_\infty$, then $C_n \xrightarrow[n\to\infty]{a.s.} C_\infty$. Furthermore, if $n^2Q_n \xrightarrow[n\to\infty]{L_1} C_\infty$, then $C_n \xrightarrow[n\to\infty]{L_1} C_\infty$.*

Proof. This is a direct application of Lemma 2.D.5, where $R_n := n^2Q_n$ and $\lambda_n := n^{-2}$, as $n\sum_{k=n}^{\infty}k^{-2} \xrightarrow[n\to\infty]{} 1$. □

Proof of Proposition 2.D.1. Recall that from Corollary 2.2.3, $\mathcal{B}_c$ and $\mathcal{B}_{1-c}$ form a partition of the set of the positive integers $\mathbb{N}^*$, so that we can write

$$V_N = V_N^{(c)} + V_N^{(1-c)},$$

where $V_N^{(c)} = \sum_{\substack{K=N+1\\K\in\mathcal{B}_c}}^{\infty}S_{N,K-1}$ and $V_N^{(1-c)} = \sum_{\substack{K=N+1\\K\in\mathcal{B}_{1-c}}}^{\infty}S_{N,K-1}$. Here, we only detail the computation of $V_N^{(c)}$, as one would proceed analogously with $V_N^{(1-c)}$.

In $V_N^{(c)}$, the sum is over the $K\in\mathcal{B}_c$. So, from Lemma 2.D.4,

$$S_{N,K-1} = 4N\left(\frac{(n_K-2)(n_K-3)}{(m_K-1)(m_K-2)n_K(n_K-1)}c_K^{(1,0)} - \frac{1}{(m_K-2)^2}(U_K^h)^2 + \psi(K)\right).$$

Now we use Proposition 2.A.2 to replace K with $\kappa_c(m_K) = \left\lfloor \frac{m_K-2}{c} \right\rfloor$ and

$$S_{N,\kappa_c(m_K)-1} = 4N\Bigg(\frac{(\kappa_c(m_K) - m_K + 2)(\kappa_c(m_K) - m_K + 1)}{(m_K-1)(m_K-2)(\kappa_c(m_K) - m_K + 4)(\kappa_c(m_K) - m_K + 3)} c^{(1,0)}_{\kappa_c(m_K)} - \frac{1}{(m_K-2)^2}(U^h_{\kappa_c(m_K)})^2 + \psi(\kappa_c(m_K)) \Bigg).$$

Therefore, because for all $K \in \mathcal{B}_c$ we have $m_K = m_{K-1} + 1$, we can then transform the sum over K into a sum over m and

$$V_N^{(c)} = \sum_{\substack{K=N+1 \\ K\in\mathcal{B}_c}}^{\infty} S_{N,K-1} = \sum_{m=m_{N+1}}^{\infty} S_{N,\kappa_c(m)-1} = N \sum_{m=m_{N+1}}^{\infty} R_m,$$

where $R_m := S_{N,\kappa_c(m)-1}/N$, i.e.

$$R_m = \frac{4(\kappa_c(m) - m + 2)(\kappa_c(m) - m + 1)}{(m-1)(m-2)(\kappa_c(m) - m + 4)(\kappa_c(m) - m + 3)} c^{(1,0)}_{\kappa_c(m)} - \frac{4}{(m-2)^2}(U^h_{\kappa_c(m)})^2 + 4\psi(\kappa_c(m)).$$

But we notice that since $\psi(\kappa_c(m)) = o(m^{-2})$, then Lemma 2.D.3 and Proposition 2.C.1 give for all N,

$$m^2 R_m \xrightarrow[m\to\infty]{a.s.,L_1} 4(c^{(1,0)}_\infty - (U^h_\infty)^2).$$

And since $\frac{m_{N+1}}{N} \xrightarrow[N\to\infty]{} c$ from Proposition 2.2.2, we find with Lemma 2.D.6 that

$$V_N^{(c)} = \frac{N}{m_{N+1}} \times m_{N+1} \sum_{m=m_{N+1}}^{\infty} R_m \xrightarrow[N\to\infty]{a.s.,L_1} \frac{4}{c}(c^{(1,0)}_\infty - (U^h_\infty)^2).$$

We can proceed likewise with $V_N^{(1-c)}$, where all the terms have $K \in \mathcal{B}_{1-c}$, to get

$$V_N^{(1-c)} \xrightarrow[N\to\infty]{a.s.,L_1} \frac{4}{1-c}(c^{(0,1)}_\infty - (U^h_\infty)^2),$$

which finally gives

$$V_N = V_N^{(c)} + V_N^{(1-c)} \xrightarrow[N\to\infty]{a.s.,L_1} V := \frac{4}{c}(c^{(1,0)}_\infty - (U^h_\infty)^2) + \frac{4}{1-c}(c^{(0,1)}_\infty - (U^h_\infty)^2).$$

□

Appendix 2.E Conditional Lindeberg condition

We verify the conditional Lindeberg condition as stated by Proposition 2.E.1. We use the notations defined in Appendix 2.D.

Proposition 2.E.1. *Let $\epsilon > 0$. Then the conditional Lindeberg condition is satisfied:*

$$\sum_{K=N}^{\infty} \mathbb{E}\big[Z^2_{NK} \mathbb{1}_{\{|Z_{NK}|>\epsilon\}} \big| \mathcal{F}_{K+1}\big] \xrightarrow[N\to\infty]{\mathbb{P}} 0$$

The proof relies on the four following lemmas.

Lemma 2.E.2. *Let $(Q_n)_{n\geq 1}$ be a sequence of random variables. Set $C_n := n\sum_{k=n}^{\infty} Q_k$. If $n^2\mathbb{E}\big[|Q_n|\big] \xrightarrow[n\to\infty]{} 0$, then $C_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.*

Proof. Lemma 2.D.6 and the triangular inequality give $\mathbb{E}\big[|C_n|\big] \leq n\sum_{k=n}^{\infty}\mathbb{E}\big[|Q_k|\big] \xrightarrow[n\to\infty]{} 0$. Let some $\epsilon > 0$, then Markov's inequality ensures that

$$\mathbb{P}(|C_n| > \epsilon) \leq \frac{\mathbb{E}\big[|C_n|\big]}{\epsilon} \xrightarrow[n\to\infty]{} 0.$$

□

Lemma 2.E.3. *For sequences of random variables T_n and sets B_n, if $T_n \xrightarrow[n\to\infty]{L_2} T$ and $\mathbb{1}(B_n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$, then $\mathbb{E}[T_n^2\mathbb{1}(B_n)] \xrightarrow[n\to\infty]{} 0$.*

Proof. Note that for all n, $a > 0$,

$$\begin{aligned}\mathbb{E}\big[T_n^2\mathbb{1}(B_n)\big] &= \mathbb{E}\big[T_n^2\mathbb{1}(B_n)\mathbb{1}(T_n^2 > a)\big] + \mathbb{E}\big[T_n^2\mathbb{1}(B_n)\mathbb{1}(T_n^2 \leq a)\big]\\ &\leq \mathbb{E}\big[T_n^2\mathbb{1}(T_n^2 > a)\big] + \mathbb{E}\big[a\mathbb{1}(B_n)\big]\\ &\leq \mathbb{E}\big[T_n^2\mathbb{1}(T_n^2 > a)\big] + a\mathbb{P}(B_n)\end{aligned}$$

Let $\epsilon > 0$. $T_n \xrightarrow[n\to\infty]{L_2} T$, so $(T_n^2)_{n\geq 1}$ is uniformly integrable and there exists $a > 0$ such that $\mathbb{E}[T_n^2\mathbb{1}(T_n^2 > a)] \leq \sup_k \mathbb{E}[T_k^2\mathbb{1}(T_k^2 > a)] \leq \frac{\epsilon}{2}$. Moreover, $\mathbb{1}(B_n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$, which translates to $\mathbb{P}(B_n) \xrightarrow[n\to\infty]{} 0$ and there exists an integer n_0 such that for all $n > n_0$, $\mathbb{P}(B_n) \leq \frac{\epsilon}{2a}$. Choosing such a real number a, we can always find an integer n_0 such that for $n > n_0$, we have $\mathbb{E}[T_n^2\mathbb{1}(B_n)] \leq \epsilon$. □

Lemma 2.E.4. *For sequences of random variables M_n and sets B_n, if $(M_n)_{n\geq 1}$ is a backward martingale with respect to some filtration and $\mathbb{1}(B_n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$, then $\mathbb{E}[M_n\mathbb{1}(B_n)] \xrightarrow[n\to\infty]{} 0$.*

Proof. We notice that from Theorem 2.B.3, $(M_n)_{n\geq 1}$ is uniformly integrable, then the proof is similar to that of Lemma 2.E.3. □

Lemma 2.E.5. *Set $A_K := m_K^{-1}\binom{n_K}{2}^{-1}\sum_{\substack{2\leq i_2\leq m_K+1\\ 1\leq j_1<j_2\leq n_K}} X_{\{1,i_2;j_1,j_2\}}$. If $K \in \mathcal{B}_c$ (see Definition 2.A.1), then $A_K \overset{\mathcal{D}}{=} \delta_K$, where δ_K is defined in Lemma 2.D.2.*

Proof. Remember that if $K \in \mathcal{B}_c$, then by symmetry of h, $\delta_K = (m_K - 1)^{-1}\binom{n_K}{2}^{-1}\sum_{\substack{1\leq i_2\leq m_K-1\\ 1\leq j_1<j_2\leq n_K}} X_{\{m_K,i_2;j_1,j_2\}}$. The exchangeability of Y says that all permutations on the

rows and the columns of Y leave its distribution unchanged, hence for all $(\sigma_1, \sigma_2) \in \mathbb{S}_{m_K} \times \mathbb{S}_{n_K}$, we have

$$\delta_K \overset{\mathcal{L}}{=} (m_K - 1)^{-1} \binom{n_K}{2}^{-1} \sum_{\substack{1 \le i_2 \le m_K - 1 \\ 1 \le j_1 < j_2 \le n_K}} X_{\{\sigma_1(m_K), \sigma_1(i_2); \sigma_2(j_1), \sigma_2(j_2)\}}.$$

Consider σ_2 to be the identity and $\sigma_1 \in \mathbb{S}_{m_K}$ the permutation defined by:

- $\sigma_1(i) = i + 1$ if $i < m_K$,
- $\sigma_1(m_K) = 1$,
- $\sigma_1(i) = i$ if $i > m_K$.

Then $A_K = (m_K - 1)^{-1} \binom{n_K}{2}^{-1} \sum_{\substack{1 \le i_2 \le m_K - 1 \\ 1 \le j_1 < j_2 \le n_K}} X_{\{\sigma_1(m_K), \sigma_1(i_2); \sigma_2(j_1), \sigma_2(j_2)\}}$, hence $A_K \overset{\mathcal{L}}{=} \delta_K$. □

Proof of Proposition 2.E.1. Similarly to the proof of the Proposition 2.D.1, we can verify the conditional Lindeberg condition by decomposing the sum along with $K+1 \in \mathcal{B}_c$ and $K+1 \in \mathcal{B}_{1-c}$ (Cor. 2.2.3), so here we only consider $\sum_{\substack{K=N+1 \\ K \in \mathcal{B}_c}}^{\infty} \mathbb{E}\big[Z^2_{N,K-1} \mathbb{1}_{\{|Z_{N,K-1}| > \epsilon\}} \mid \mathcal{F}_K\big]$.

Like previously, using Proposition 2.A.2, we can transform the sum over K into a sum over m:

$$\sum_{\substack{K=N+1 \\ K \in \mathcal{B}_c}}^{\infty} \mathbb{E}\big[Z^2_{N,K-1} \mathbb{1}_{\{|Z_{N,K-1}| > \epsilon\}} \mid \mathcal{F}_K\big] = \sum_{m=m_{N+1}}^{\infty} \mathbb{E}\big[Z^2_{N,\kappa_c(m)-1} \mathbb{1}_{\{|Z_{N,\kappa_c(m)-1}| > \epsilon\}} \mid \mathcal{F}_{\kappa_c(m)}\big],$$

where $\kappa_c(m) = \lfloor \frac{m-2}{c} \rfloor$.

We remark that using Lemma 2.D.2, for $m \ge m_{N+1} = m_N + 1 > c(N+1) + 2$,

$$\begin{aligned}
\mathbb{1}_{\{|Z_{N,\kappa_c(m)-1}| > \epsilon\}} &\le \mathbb{1}_{\left\{\frac{2\sqrt{N}}{m-2} |U^h_{\kappa_c(m)} - \delta_{\kappa_c(m)}| > \epsilon\right\}} \\
&\le \mathbb{1}_{\left\{|U^h_{\kappa_c(m)} - \delta_{\kappa_c(m)}| > \frac{m-2}{2\sqrt{\frac{m-2}{c}}} \epsilon\right\}} \\
&\le \mathbb{1}_{\left\{|U^h_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4} \epsilon\right\}} + \mathbb{1}_{\left\{|\delta_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4} \epsilon\right\}}.
\end{aligned}$$

So, using Lemma 2.D.2 again and the identity $(U^h_{\kappa_c(m)} - \delta_{\kappa_c(m)})^2 \le 2\Big((U^h_{\kappa_c(m)})^2 + \delta^2_{\kappa_c(m)}\Big)$, we get for $m \ge m_{N+1}$,

$$\begin{aligned}
&\mathbb{E}\big[Z^2_{N,\kappa_c(m)-1} \mathbb{1}_{\{|Z_{N,\kappa_c(m)-1}| > \epsilon\}} \mid \mathcal{F}_{\kappa_c(m)}\big] \\
&\le \frac{8N}{(m-2)^2} \mathbb{E}\left[\Big((U^h_{\kappa_c(m)})^2 + \delta^2_{\kappa_c(m)}\Big) \left(\mathbb{1}_{\left\{|U^h_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4} \epsilon\right\}} + \mathbb{1}_{\left\{|\delta_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4} \epsilon\right\}} \right) \middle| \mathcal{F}_{\kappa_c(m)} \right].
\end{aligned}$$

This inequality and Lemma 2.E.2 imply that a sufficient condition to have the conditional Lindeberg condition is

$$\mathbb{E}\left[\Big((U^h_{\kappa_c(m)})^2 + \delta^2_{\kappa_c(m)}\Big) \left(\mathbb{1}_{\left\{|U^h_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4} \epsilon\right\}} + \mathbb{1}_{\left\{|\delta_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4} \epsilon\right\}} \right) \right] \xrightarrow[m \to \infty]{} 0. \tag{2.27}$$

Next, we prove that this condition is satisfied.

First, note that from Markov's inequality,

$$\mathbb{P}\left(|U^h_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4}\epsilon\right) \leq \frac{4\mathbb{E}[|U^h_{\kappa_c(m)}|]}{\epsilon\sqrt{c(m-2)}} \xrightarrow[m\to\infty]{} 0$$

and

$$\mathbb{P}\left(|\delta_{\kappa_c(m)}| > \frac{\sqrt{c(m-2)}}{4}\epsilon\right) \leq \frac{4\mathbb{E}[|\delta_{\kappa_c(m)}|]}{\epsilon\sqrt{c(m-2)}} \xrightarrow[m\to\infty]{} 0.$$

Now, remember that from Proposition 2.C.1, $U^h_K \xrightarrow[K\to\infty]{L_2} U^h_\infty$, therefore $U^h_{\kappa_c(m)} \xrightarrow[m\to\infty]{L_2} U^h_\infty$ and Lemma 2.E.3 can be applied, which gives

$$\mathbb{E}\left[(U^h_{\kappa_c(m)})^2\left(\mathbb{1}_{\left\{|U^h_{\kappa_c(m)}|>\frac{\sqrt{c(m-2)}}{4}\epsilon\right\}} + \mathbb{1}_{\left\{|\delta_{\kappa_c(m)}|>\frac{\sqrt{c(m-2)}}{4}\epsilon\right\}}\right)\right] \xrightarrow[m\to\infty]{} 0. \tag{2.28}$$

Likewise, we calculated $\mathbb{E}[\delta_K^2 \mid \mathcal{F}_K]$ in the proof of Lemma 2.D.4. The application of Lemma 2.D.3 shows that $\mathbb{E}[\delta^2_{\kappa_c(m)} \mid \mathcal{F}_{\kappa_c(m)}]$ is a backward martingale. It follows from Lemma 2.E.4 that

$$\mathbb{E}\left[\delta^2_{\kappa_c(m)}\mathbb{1}_{\left\{|U^h_{\kappa_c(m)}|>\frac{\sqrt{c(m-2)}}{4}\epsilon\right\}}\right] = \mathbb{E}\left[\mathbb{E}[\delta^2_{\kappa_c(m)} \mid \mathcal{F}_{\kappa_c(m)}]\mathbb{1}_{\left\{|U^h_{\kappa_c(m)}|>\frac{\sqrt{c(m-2)}}{4}\epsilon\right\}}\right] \xrightarrow[m\to\infty]{} 0. \tag{2.29}$$

Finally, applying Lemma 2.E.5, we obtain

$$\mathbb{E}\left[\delta^2_{\kappa_c(m)}\mathbb{1}_{\left\{|\delta_{\kappa_c(m)}|>\frac{\sqrt{c(m-2)}}{4}\epsilon\right\}}\right] = \mathbb{E}\left[A^2_{\kappa_c(m)}\mathbb{1}_{\left\{|A_{\kappa_c(m)}|>\frac{\sqrt{c(m-2)}}{4}\epsilon\right\}}\right], \tag{2.30}$$

where $A_K = m_K^{-1}\binom{n_K}{2}^{-1}\sum_{\substack{2\leq i_2\leq m_K+1\\ 1\leq j_1<j_2\leq n_K}} X_{\{1,i_2;j_1,j_2\}}$. Using similar arguments as in the proof of Proposition 2.C.1, it can be shown that A_K is a square integrable backward martingale with respect to the decreasing filtration $\mathcal{F}^A_K = \sigma(A_K, A_{K+1}, \ldots)$. Therefore, Theorem 2.B.3 ensures that there exists A_∞ such that $A_K \xrightarrow[K\to\infty]{L_2} A_\infty$. This proves that $A_{\kappa_c(m)} \xrightarrow[m\to\infty]{L_2} A_\infty$, so applying Lemma 2.E.3 again, we obtain

$$\mathbb{E}\left[A^2_{\kappa_c(m)}\mathbb{1}_{\left\{|A_{\kappa_c(m)}|>\frac{\sqrt{c(m-2)}}{4}\epsilon\right\}}\right] \xrightarrow[m\to\infty]{} 0. \tag{2.31}$$

Combining (2.28), (2.29), (2.30) and (2.31), we deduce that the sufficient condition (2.27) is satisfied, thus concluding the proof.

□

Appendix 2.F Hewitt-Savage theorem

Proof of Theorem 2.2.13. This proof adapts the steps taken by Feller (1971) and detailed by Durrett (2019) to our case. Let $A \in \mathcal{E}_\infty$.

First, let $\mathcal{A}_N = \sigma\big((\xi_i)_{1\le i\le m_N}, (\eta_j)_{1\le j\le n_N}, (\zeta_{ij})_{1\le i\le m_N, 1\le j\le n_N}\big)$, the σ-field generated by the random variables associated with the first m_N rows and n_N columns. Notice that $A \in \mathcal{A} := \bigcap_{n=1}^{\infty} \mathcal{A}_N$. Since $\mathcal{A}$ is the limit of $\mathcal{A}_N$, then for all $\epsilon > 0$, there exists a N and an associated set $A_N \in \mathcal{A}_N$ such that $\mathbb{P}(A - A \cap A_N) < \epsilon$ and $\mathbb{P}(A_N - A \cap A_N) < \epsilon$, so that $\mathbb{P}(A \Delta A_N) < 2\epsilon$, where Δ is the symmetric difference operator, i.e. $B\Delta C = (B - C) \cup (C - B)$. Therefore, we can pick a sequence of sets A_N such that $\mathbb{P}(A\Delta A_N) \longrightarrow 0$.

Next, we consider the row-column permutation $\Phi^{(N)} = (\sigma_1^{(N)}, \sigma_2^{(N)}) \in \mathbb{S}_{m_N} \times \mathbb{S}_{n_N}$ defined by

$$\sigma_1^{(N)}(i) = \begin{cases} i + m_N & \text{if } 1 \le i \le m_N, \\ i - m_N & \text{if } m_N + 1 \le i \le 2m_N, \\ i & \text{if } 2m_N + 1 \le i. \end{cases}$$

$$\sigma_2^{(N)}(j) = \begin{cases} j + n_N & \text{if } 1 \le j \le n_N, \\ j - n_N & \text{if } n_N + 1 \le j \le 2n_N, \\ j & \text{if } 2n_N + 1 \le j. \end{cases}$$

Since $A \in \mathcal{E}_\infty$, by the definition of $\mathcal{E}_\infty$, it follows that

$$\left\{\omega : \Phi^{(N)}\omega \in A\right\} = \{\omega : \omega \in A\} = A.$$

Using this, if we denote $A'_N := \left\{\omega : \Phi^{(N)}\omega \in A_N\right\}$, then we can write that

$$\left\{\omega : \Phi^{(N)}\omega \in A_N \Delta A\right\} = \left\{\omega : \omega \in A'_N \Delta A\right\} = A'_N \Delta A.$$

Furthermore, the $(\xi_i)_{1\le i<\infty}$, $(\eta_j)_{1\le j<\infty}$ and $(\zeta_{ij})_{1\le i<\infty, 1\le j<\infty}$ are i.i.d., so

$$\mathbb{P}(A_N \Delta A) = \mathbb{P}(\omega : \omega \in A_N \Delta A) = \mathbb{P}(\omega : \Phi^{(N)}\omega \in A_N \Delta A).$$

and we conclude that $\mathbb{P}(A'_N \Delta A) = \mathbb{P}(A_N \Delta A) \longrightarrow 0$.

From this, we derive that $\mathbb{P}(A_N) \longrightarrow \mathbb{P}(A)$ and $\mathbb{P}(A'_N) \longrightarrow \mathbb{P}(A)$. We also remark that $\mathbb{P}(A_N \Delta A'_N) \le \mathbb{P}(A_N \Delta A) + \mathbb{P}(A'_N \Delta A) \longrightarrow 0$, so $\mathbb{P}(A_N \cap A'_N) \longrightarrow \mathbb{P}(A)$.

But A_N and A'_N are independent, so we have $\mathbb{P}(A_N \cap A'_N) = \mathbb{P}(A_N)\mathbb{P}(A'_N) \longrightarrow \mathbb{P}(A)^2$, therefore $\mathbb{P}(A) = \mathbb{P}(A)^2$, which means that $\mathbb{P}(A) = 0$ or 1. □

Appendix 2.G Identifiability of the BEDD model

2.G.1 Proof of Theorem 2.3.2

First, we define the generalized inverse of a cumulative distribution function and we prove some useful properties. We need Lemmas 2.G.3 and 2.G.4 to prove Theorem 2.3.2.

Definition 2.G.1. *For any increasing, bounded, càdlàg function $t : \mathbb{R} \to \mathbb{R}$, we define its generalized inverse by $t^{-1} : \mathbb{R} \to \mathbb{R}$ as follows:*

$$t^{-1}(x) = \inf\{u \in \mathbb{R} : t(u) > x\}.$$

Lemma 2.G.2 (de La Fortelle, 2020, Proposition 2.2.)**.** *Let t be an increasing, bounded, càdlàg function. Then $(t^{-1})^{-1} = t$.*

Lemma 2.G.3. *Let U be a random variable such that $U \sim \mathcal{U}[0,1]$. Let $D = f(U)$, where f is an increasing, bounded, càdlàg function. Let $\tilde{f}$ be the cumulative distribution function of D. Then $f^{-1} = \tilde{f}$ and $\tilde{f}^{-1} = f$.*

Proof. Since f is right-continuous, for all $x \in \mathbb{R}$, $\mathbb{P}(f(U) \leq x) = \inf\{u \in [0,1] : f(u) > x\}$ so that means $\tilde{f} = f^{-1}$ according to Definition 2.G.1, so the first equation is proven. Also Lemma 2.G.2 ensures that $f = (f^{-1})^{-1} = \tilde{f}^{-1}$, which concludes the proof of this lemma. □

Lemma 2.G.4. *Let D be a random variable such for all $k \in \mathbb{N}$, $\mathbb{E}[D^k] \leq \alpha^k$, for some $\alpha > 0$. Then, the distribution of D is uniquely characterized by its moments.*

Proof. We show that the exponential generating series of the moments of D $M(r) = \sum_{k=1}^{\infty} \mathbb{E}[D^k]\frac{r^k}{k!}$ has a positive radius of convergence.

Using the fact that $k!/k^k \xrightarrow[k\to\infty]{} 1$ and that $\mathbb{E}[D^k] \leq \alpha^k$ for all $k \in \mathbb{N}$, we see that

$$\begin{aligned}\limsup_{k\to\infty} \left(\frac{\mathbb{E}[D^k]}{k!}\right)^{\frac{1}{k}} &= \limsup_{k\to\infty} \frac{\mathbb{E}[D^k]^{\frac{1}{k}}}{k} \\ &\leq \alpha \limsup_{k\to\infty} \frac{1}{k} \\ &< \infty.\end{aligned}$$

So by the Cauchy-Hadamard's theorem, the series $\sum_k \frac{\mathbb{E}[D^k]r^k}{k!}$ converges for any $r > 0$, which is a sufficient condition so that D is determined by its moments $(\mathbb{E}[D^k])_{k\geq 1}$ (see Sect. 9.2, Thm. 2 of Billingsley (1995)). □

Using Lemmas 2.G.3 and 2.G.4, we can finally prove Theorem 2.3.2.

Proof of Theorem 2.3.2. Let $\Theta = (\lambda, f, g)$ be BEDD parameters. Here, we prove that f is uniquely characterised by $(F_k)_{k\geq 1}$. In order to do that, we introduce a random variable D which both has moments $(F_k)_{k\geq 1}$ and f as the generalized inverse of its cumulative distribution function (Definition 2.G.1). We show that D is uniquely characterised by f and then, by its moments.

1. Let U be a random variable such that $U \sim \mathcal{U}[0,1]$. Let $D = f(U)$. For all $k \geq 1$, $\mathbb{E}[D^k] = F_k$.
2. Since f is bounded, we notice that for all $k \in \mathbb{N}$, $F_k = \int f^k \leq \sup_{[0,1]} f^k = \|f\|_\infty^k$, therefore $\mathbb{E}[D^k] \leq \|f\|_\infty^k$.

   So Lemma 2.G.4 ensures that the distribution of D is uniquely characterised by its moments $(F_k)_{k\geq 1}$.
3. Now, for some other increasing, bounded, càdlàg function f^*. Let $D^* = f^*(U)$ and $\tilde{f}^*$ its cumulative distribution function. If $D \sim D^*$, then $\tilde{f} = \tilde{f}^*$. Therefore, using the generalized inverse (Definition 2.G.1), we have $\tilde{f}^{-1} = (\tilde{f}^*)^{-1}$. Finally, Lemma 2.G.3 implies that $f = f^*$. Thus, the distribution of D is uniquely characterised by f.

We can conclude by stating that the moments $(F_k)_{k\geq 1}$ of D are then uniquely characterised by f.

By symmetry, the same follows for g and $(G_k)_{k\geq 1}$. □

2.G.2 Proof of Theorem 2.3.4

Theorem 2.3.4 is proven by induction using two lemmas.

Lemma 2.G.5. *Let $\Theta = (\lambda, f, g)$ be BEDD parameters and $Y \sim \mathcal{L}$-BEDD(Θ). For all $i \in \mathbb{N}$ and for all $(j_1, j_2) \in \mathbb{N}^2$ such that $j_1 \neq j_2$,*

$$\mathbb{E}[\Psi_k(Y_{ij_1}) \times Y_{ij_2}] = \lambda^{k+1} F_{k+1} G_k.$$

Proof. Let $\Theta = (\lambda, f, g)$ be BEDD parameters and $Y \sim \mathcal{L}$-BEDD(Θ), for any $i \in \mathbb{N}$ and $(j_1, j_2) \in$

$\mathbb{N}^2$ such that $j_1 \neq j_2$,

$$\begin{aligned}
\mathbb{E}[\Psi_k(Y_{ij_1}) \times Y_{ij_2}] &= \mathbb{E}\big[\mathbb{E}[\Psi_k(Y_{ij_1}) \times Y_{ij_2} \mid \xi_i, \eta_{j_1}, \eta_{j_2}]\big] \\
&= \mathbb{E}\big[\mathbb{E}[\Psi_k(Y_{ij_1}) \mid \xi_i, \eta_{j_1}] \times \mathbb{E}[Y_{ij_2} \mid \xi_i, \eta_{j_2}]\big] \\
&= \mathbb{E}\big[\lambda^k f(\xi_i)^k g(\eta_{j_1})^k \times \lambda f(\xi_i) g(\eta_{j_2})\big] \\
&= \lambda^{k+1} \mathbb{E}\big[f(\xi_i)^{k+1}\big] \mathbb{E}\big[g(\eta_{j_1})^k\big] \mathbb{E}\big[g(\eta_{j_2})\big] \\
&= \lambda^{k+1} F_{k+1} G_k.
\end{aligned}$$

□

Lemma 2.G.6. *Let $\Theta = (\lambda, f, g)$ be BEDD parameters and $Y \sim BEDD(\Theta)$. For all $(i_1, i_2) \in \mathbb{N}^2$ such that $i_1 \neq i_2$ and for all $j \in \mathbb{N}$,*

$$\mathbb{E}[\Psi_k(Y_{i_1 j_1}) \times Y_{i_2 j_1}] = \lambda^{k+1} F_k G_{k+1}.$$

Proof. Let $\Theta = (\lambda, f, g)$ be BEDD parameters and $Y \sim \mathcal{L}\text{-BEDD}(\Theta)$, for any $(i_1, i_2) \in \mathbb{N}^2$ such that $i_1 \neq i_2$ and for any $j \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}[\Psi_k(Y_{i_1 j}) \times Y_{i_2 j}] &= \mathbb{E}\big[\mathbb{E}[\Psi_k(Y_{i_1 j}) \times Y_{i_2 j} \mid \xi_{i_1}, \xi_{i_2}, \eta_j]\big] \\
&= \mathbb{E}\big[\mathbb{E}[\Psi_k(Y_{i_1 j}) \mid \xi_{i_1}, \eta_j] \times \mathbb{E}[Y_{i_2 j} \mid \xi_{i_2}, \eta_j]\big] \\
&= \mathbb{E}\big[\lambda^k f(\xi_{i_1})^k g(\eta_j)^k \times \lambda f(\xi_{i_2}) g(\eta_j)\big] \\
&= \lambda^{k+1} \mathbb{E}\big[f(\xi_{i_1})^k\big] \mathbb{E}\big[f(\xi_{i_2})\big] \mathbb{E}\big[g(\eta_j)^{k+1}\big] \\
&= \lambda^{k+1} F_k G_{k+1}.
\end{aligned}$$

□

Proof of Theorem 2.3.4. Let $\Theta = (\lambda, f, g)$ be BEDD parameters and $Y \sim \mathcal{L}\text{-BEDD}(\Theta)$. Let $Y_{(i_1, i_2; j_1, j_2)}$ be a quadruplet of Y. Since Assumption 2.3.3 holds, we set the $(\Psi_k)_{k \geq 1}$ such that $\mathbb{E}[\Psi_k(X)] = \mu^k$ for all $k \in \mathbb{N}$. We know that $\mathbb{E}[Y_{i_1 j_1}] = \lambda$.

- First, recall that $F_1 = G_1 = 1$.
- Then, having recovered λ and all the F_ℓ, G_ℓ for $1 \leq \ell \leq k$, we can recover F_{k+1} and G_{k+1} as from Lemmas 2.G.5 and 2.G.6,

$$\begin{aligned}
F_{k+1} &= \lambda^{-(k+1)} G_k^{-1} \mathbb{E}[\Psi_k(Y_{i_1 j_1}) \times Y_{i_1 j_2}] \\
G_{k+1} &= \lambda^{-(k+1)} F_k^{-1} \mathbb{E}[\Psi_k(Y_{i_1 j_1}) \times Y_{i_2 j_1}]
\end{aligned}$$

Then, the theorem is proven by induction. □

Appendix 2.H Derivation of variances

In this section, we derive a general formula for the covariance of two U-statistics and then we derive asymptotic variances for specific kernels used in Section 2.3. We denote for any $k > 0$, $F_k := \int f^k$ and $G_k := \int g^k$.

Lemma 2.H.1. *Let Y be a RCE matrix. Let h_1 and h_2 be two quadruplet kernels. Let $U_N^{h_1} := U_{m_N,n_N}^{h_1}$ and $U_N^{h_2} := U_{m_N,n_N}^{h_2}$, defined by Formula (2.5) and Definition 2.2.1. For any σ-field $\mathcal{F}$,*

$$\begin{aligned}\mathrm{Cov}(U_N^{h_1}, U_N^{h_2} \mid \mathcal{F}) = \frac{4}{cN}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(1,3;3,4)}) \mid \mathcal{F}) \\ + \frac{4}{(1-c)N}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(3,4;1,3)}) \mid \mathcal{F}) + o\left(\frac{1}{N}\right).\end{aligned}$$

Proof. Similar to the proof of Lemma 2.D.4, using the exchangeability of the quadruplets, we deduce that

$$\begin{aligned}&\mathrm{Cov}(U_N^{h_1}, U_N^{h_2} \mid \mathcal{F}) \\ &= \binom{m_N}{2}^{-2}\binom{n_N}{2}^{-2}\mathrm{Cov}\left(\sum_{\substack{1\le i_1<i_2\le m_N\\ 1\le j_1<j_2\le n_N}} h_1(Y_{(i_1,i_2;j_1,j_2)}), \sum_{\substack{1\le i_1<i_2\le m_N\\ 1\le j_1<j_2\le n_N}} h_2(Y_{(i_1,i_2;j_1,j_2)})\middle|\mathcal{F}\right), \\ &= \binom{m_N}{2}^{-2}\binom{n_N}{2}^{-2}\sum_{\substack{1\le i_1<i_2\le m_N\\ 1\le j_1<j_2\le n_N}}\sum_{\substack{1\le i_1'<i_2'\le m_N\\ 1\le j_1'<j_2'\le n_N}}\mathrm{Cov}(h_1(Y_{(i_1,i_2;j_1,j_2)}), h_2(Y_{(i_1',i_2';j_1',j_2')}) \mid \mathcal{F}), \\ &= \binom{m_N}{2}^{-1}\binom{n_N}{2}^{-1}\sum_{\substack{1\le i_1'<i_2'\le m_N\\ 1\le j_1'<j_2'\le n_N}}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(i_1',i_2';j_1',j_2')}) \mid \mathcal{F}), \\ &= \binom{m_N}{2}^{-1}\binom{n_N}{2}^{-1}\Big((m_N-2)(n_N-2)(n_N-3)\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(1,3;3,4)}) \mid \mathcal{F}) \\ &+ (m_N-2)(m_N-3)(n_N-2)\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(3,4;1,3)}) \mid \mathcal{F}) \\ &+ o(m_N n_N^2) + o(m_N^2 n_N)\Big), \\ &= \frac{4}{c^2(1-c)^2N^4}\Big(c(1-c)^2N^3\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(1,3;3,4)}) \mid \mathcal{F}) \\ &+ c^2(1-c)N^3\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(3,4;1,3)}) \mid \mathcal{F}) + o(N^3)\Big), \\ &= \frac{4}{cN}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(1,3;3,4)}) \mid \mathcal{F}) \\ &+ \frac{4}{(1-c)N}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(3,4;1,3)}) \mid \mathcal{F}) + o\left(\frac{1}{N}\right).\end{aligned}$$

□

Corollary 2.H.2. *Let Y be a RCE matrix. Let h_1 and h_2 be two quadruplet kernels. Let $U_N^{h_1} := U_{m_N,n_N}^{h_1}$ and $U_N^{h_2} := U_{m_N,n_N}^{h_2}$, defined by Formula (2.5) and Definition 2.2.1. For any σ-field $\mathcal{F}$,*

$$\begin{aligned}\mathbb{V}[U_N^h \mid \mathcal{F}] = \frac{4}{cN}\mathrm{Cov}(h(Y_{(1,2;1,2)}), h(Y_{(1,3;3,4)}) \mid \mathcal{F}) \\ + \frac{4}{(1-c)N}\mathrm{Cov}(h(Y_{(1,2;1,2)}), h(Y_{(3,4;1,3)}) \mid \mathcal{F}) + o\left(\frac{1}{N}\right).\end{aligned}$$

Lemma 2.H.3. *Let Y be a RCE matrix generated by the Poisson-BEDD model, with density λ and degree functions f and g. Let h be a quadruplet kernel defined by*

$$h(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_1j_2} + Y_{i_2j_1}Y_{i_2j_2}),$$

then a closed-form expression for V of Theorem 2.2.7 is

$$V = \frac{\lambda^4}{c}(F_4 - F_2^2) + \frac{4\lambda^4}{1-c}F_2^2(G_2 - 1).$$

Proof. Using the fact that the $(Y_{ij})_{(i,j)\in\mathbb{N}^2}$ are independent conditionally on $\xi = (\xi_i)_{i\in\mathbb{N}}$ and $\eta = (\eta_j)_{j\in\mathbb{N}}$, we find that

$$\begin{aligned}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(1,3;3,4)})\big) &= \frac{1}{4}\mathrm{Cov}\big(Y_{11}Y_{12} + Y_{21}Y_{22}, Y_{13}Y_{14} + Y_{33}Y_{34}\big), \\ &= \frac{1}{4}\mathrm{Cov}\big(Y_{11}Y_{12}, Y_{13}Y_{14}\big), \\ &= \frac{1}{4}\left(\mathbb{E}[Y_{11}Y_{12}Y_{13}Y_{14}] - \mathbb{E}[Y_{11}Y_{12}]\mathbb{E}[Y_{13}Y_{14}]\right), \\ &= \frac{1}{4}\left(\mathbb{E}[Y_{11}Y_{12}Y_{13}Y_{14}] - \mathbb{E}[Y_{11}Y_{12}]^2\right), \\ &= \frac{1}{4}\left(\mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{13}Y_{14} \mid \xi,\eta]] - \mathbb{E}[\mathbb{E}[Y_{11}Y_{12} \mid \xi,\eta]]^2\right), \\ &= \frac{1}{4}\Big(\mathbb{E}[\lambda^4 f(\xi_1)^4 g(\eta_1)g(\eta_2)g(\eta_3)g(\eta_4)] \\ &\quad -\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1)g(\eta_2)]^2\Big), \\ &= \frac{\lambda^4}{4}(F_4 - F_2^2),\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(3,4;1,3)})\big) &= \frac{1}{4}\mathrm{Cov}\big(Y_{11}Y_{12} + Y_{21}Y_{22}, Y_{31}Y_{33} + Y_{41}Y_{43}\big),\\
&= \frac{1}{4} \times 4\mathrm{Cov}\big(Y_{11}Y_{12}, Y_{31}Y_{33}\big),\\
&= \mathbb{E}[Y_{11}Y_{12}Y_{31}Y_{33}] - \mathbb{E}[Y_{11}Y_{12}]\mathbb{E}[Y_{31}Y_{33}],\\
&= \mathbb{E}[Y_{11}Y_{12}Y_{31}Y_{33}] - \mathbb{E}[Y_{11}Y_{12}]^2,\\
&= \mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{31}Y_{33} \mid \xi, \eta]] - \mathbb{E}[\mathbb{E}[Y_{11}Y_{12} \mid \xi, \eta]]^2,\\
&= \mathbb{E}[\lambda^4 f(\xi_1)^2 f(\xi_3)^2 g(\eta_1)^2 g(\eta_2) g(\eta_3)]\\
&\quad - \mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1) g(\eta_2)]^2,\\
&= \lambda^4 (F_2^2 G_2 - F_2^2),\\
&= \lambda^4 F_2^2 (G_2 - 1).
\end{aligned}$$

The proof is concluded injecting these results in the expression of

$$V = \frac{4}{c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(1,3;3,4)})\big) + \frac{4}{1-c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(3,4;1,3)})\big)$$

given by Theorem 2.2.7. □

Lemma 2.H.4. *Let Y be a RCE matrix generated by the Poisson-BEDD model, with density λ and degree functions f and g. Let h be a quadruplet kernel defined by*

$$h(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{2}(Y_{i_1 j_1} Y_{i_2 j_2} + Y_{i_1 j_2} Y_{i_2 j_1}),$$

then a closed-form expression for V of Theorem 2.2.7 is

$$V = \frac{4\lambda^4}{c}(F_2 - 1) + \frac{4\lambda^4}{1-c}(G_2 - 1).$$

Proof. Using the fact that the $(Y_{ij})_{(i,j)\in\mathbb{N}^2}$ are independent conditionally on $\xi = (\xi_i)_{i\in\mathbb{N}}$ and $\eta = (\eta_j)_{j\in\mathbb{N}}$, we find that

$$\begin{aligned}
\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(1,3;3,4)})\big) &= \frac{1}{4}\mathrm{Cov}\big(Y_{11}Y_{22} + Y_{12}Y_{21}, Y_{13}Y_{34} + Y_{14}Y_{33}\big),\\
&= \frac{1}{4} \times 4\mathrm{Cov}\big(Y_{11}Y_{22}, Y_{13}Y_{34}\big),\\
&= \mathbb{E}[Y_{11}Y_{22}Y_{13}Y_{34}] - \mathbb{E}[Y_{11}Y_{22}]\mathbb{E}[Y_{13}Y_{34}],\\
&= \mathbb{E}[Y_{11}Y_{22}Y_{13}Y_{34}] - \mathbb{E}[Y_{11}Y_{22}]^2,\\
&= \mathbb{E}[\mathbb{E}[Y_{11}Y_{22}Y_{13}Y_{34} \mid \xi, \eta]] - \mathbb{E}[\mathbb{E}[Y_{11}Y_{22} \mid \xi, \eta]]^2,\\
&= \mathbb{E}[\lambda^4 f(\xi_1)^2 f(\xi_2) f(\xi_3) g(\eta_1) g(\eta_2) g(\eta_3) g(\eta_4)]\\
&\quad - \mathbb{E}[\lambda^2 f(\xi_1) f(\xi_2) g(\eta_1) g(\eta_2)]^2,\\
&= \lambda^4 (F_2 - 1),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(3,4;1,3)})\big) &= \frac{1}{4}\mathrm{Cov}\big(Y_{11}Y_{22} + Y_{12}Y_{21}, Y_{31}Y_{43} + Y_{33}Y_{41}\big),\\
&= \frac{1}{4} \times 4\mathrm{Cov}\big(Y_{11}Y_{22}, Y_{31}Y_{43}\big),\\
&= \mathbb{E}[Y_{11}Y_{22}Y_{31}Y_{43}] - \mathbb{E}[Y_{11}Y_{22}]\mathbb{E}[Y_{31}Y_{43}],\\
&= \mathbb{E}[Y_{11}Y_{22}Y_{31}Y_{43}] - \mathbb{E}[Y_{11}Y_{22}]^2,\\
&= \mathbb{E}[\mathbb{E}[Y_{11}Y_{22}Y_{31}Y_{43} \mid \xi, \eta]] - \mathbb{E}[\mathbb{E}[Y_{11}Y_{22} \mid \xi, \eta]]^2,\\
&= \mathbb{E}[\lambda^4 f(\xi_1)f(\xi_2)f(\xi_3)f(\xi_4)g(\eta_1)^2 g(\eta_2)g(\eta_3)]\\
&\quad - \mathbb{E}[\lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2)]^2,\\
&= \lambda^4(G_2 - 1).
\end{aligned}$$

The proof is concluded injecting these results in the expression of

$$V = \frac{4}{c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(1,3;3,4)})\big) + \frac{4}{1-c}\mathrm{Cov}\big(h(Y_{(1,2;1,2)}), h(Y_{(3,4;1,3)})\big)$$

given by Theorem 2.2.7. □

Lemma 2.H.5. *Let Y be a RCE matrix generated by the Poisson-BEDD model, with density λ and degree functions f and g. Let h_1 and h_2 be two quadruplet kernels defined by*

$$h_1(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_1j_2} + Y_{i_2j_1}Y_{i_2j_2}),$$

and

$$h_2(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_2j_2} + Y_{i_1j_2}Y_{i_2j_1}).$$

Let $U_N^{h_1} := U_{m_N,n_N}^{h_1}$ and $U_N^{h_2} := U_{m_N,n_N}^{h_2}$, defined by Formula (2.5) and Definition 2.2.1. Set $C^{h_1,h_2} := \lim_{N\to+\infty} N\mathrm{Cov}(U_N^{h_1}, U_N^{h_2})$, then

$$C^{h_1,h_2} = \frac{2\lambda^4}{c}(F_3 - F_2) + \frac{4\lambda^4}{1-c}F_2(G_2 - 1).$$

Proof. First, using Lemma 2.H.1, we deduce that

$$\begin{aligned}
\mathrm{Cov}(U_N^{h_1}, U_N^{h_2}) &= \frac{4}{cN}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(1,3;3,4)}))\\
&\quad + \frac{4}{(1-c)N}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(3,4;1,3)})) + o\left(\frac{1}{N}\right),
\end{aligned}$$

so that

$$C^{h_1,h_2} = \frac{4}{c}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(1,3;3,4)})) + \frac{4}{1-c}\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(3,4;1,3)})).$$

To conclude the proof, we proceed analogously to the proofs of Lemma 2.H.3 and 2.H.4 to derive the expressions of

$$\begin{aligned}
\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(1,3;3,4)})) &= \frac{1}{4}\mathrm{Cov}(Y_{11}Y_{12} + Y_{21}Y_{22}, Y_{13}Y_{34} + Y_{33}Y_{14}),\\
&= \frac{1}{4}\mathrm{Cov}(Y_{11}Y_{12}, Y_{13}Y_{34} + Y_{33}Y_{14}),\\
&= \frac{1}{4} \times 2\mathrm{Cov}(Y_{11}Y_{12}, Y_{13}Y_{34}),\\
&= \frac{1}{2}\left(\mathbb{E}[Y_{11}Y_{12}Y_{13}Y_{34}] - \mathbb{E}[Y_{11}Y_{12}]\mathbb{E}[Y_{13}Y_{34}]\right),\\
&= \frac{1}{2}\left(\mathbb{E}[Y_{11}Y_{12}Y_{13}Y_{34}] - \mathbb{E}[Y_{11}Y_{12}]\mathbb{E}[Y_{11}Y_{22}]\right),\\
&= \frac{1}{2}\left(\mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{13}Y_{34} \mid \xi, \eta]]\right.\\
&\quad \left. -\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} \mid \xi, \eta]]\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} \mid \xi, \eta]]\right),\\
&= \frac{1}{2}\left(\mathbb{E}[\lambda^4 f(\xi_1)^3 f(\xi_3) g(\eta_1) g(\eta_2) g(\eta_3) g(\eta_4)]\right.\\
&\quad \left. -\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1) g(\eta_2)]\mathbb{E}[\lambda^2 f(\xi_1) f(\xi_2) g(\eta_1) g(\eta_2)]\right),\\
&= \frac{1}{2}\lambda^4(F_3 - F_2),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}(h_1(Y_{(1,2;1,2)}), h_2(Y_{(3,4;1,3)})) &= \frac{1}{4}\mathrm{Cov}(Y_{11}Y_{12} + Y_{21}Y_{22}, Y_{31}Y_{43} + Y_{41}Y_{33}),\\
&= \frac{1}{4} \times 4\mathrm{Cov}(Y_{11}Y_{12}, Y_{31}Y_{43}),\\
&= \mathbb{E}[Y_{11}Y_{12}Y_{31}Y_{43}] - \mathbb{E}[Y_{11}Y_{12}]\mathbb{E}[Y_{31}Y_{43}],\\
&= \mathbb{E}[Y_{11}Y_{12}Y_{31}Y_{43}] - \mathbb{E}[Y_{11}Y_{12}]\mathbb{E}[Y_{11}Y_{22}],\\
&= \mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{31}Y_{43} \mid \xi, \eta]]\\
&\quad - \mathbb{E}[\mathbb{E}[Y_{11}Y_{12} \mid \xi, \eta]]\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} \mid \xi, \eta]],\\
&= \mathbb{E}[\lambda^4 f(\xi_1)^2 f(\xi_3) f(\xi_4) g(\eta_1)^2 g(\eta_2) g(\eta_3)]\\
&\quad - \mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1) g(\eta_2)]\mathbb{E}[\lambda^2 f(\xi_1) f(\xi_2) g(\eta_1) g(\eta_2)],\\
&= \lambda^4 F_2(G_2 - 1).
\end{aligned}$$

□

Lemma 2.H.6. *Let Y be a RCE matrix generated by the Bernoulli-BEDD model, with density λ and degree functions f and g. Let h be the quadruplet kernel defined by*

$$\begin{aligned}
h(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{4}\Big(&Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_1}(1 - Y_{i_2j_2}) + Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_2}(1 - Y_{i_2j_1})\\
&+ Y_{i_1j_1}Y_{i_2j_1}Y_{i_2j_2}(1 - Y_{i_1j_2}) + Y_{i_1j_2}Y_{i_2j_1}Y_{i_2j_2}(1 - Y_{i_1j_1})\Big),
\end{aligned}$$

then $\mathbb{E}[h(Y_{(1,2;1,2)})] = \lambda^3 F_2 G_2 (1 - \lambda F_2 G_2)$ *and a closed-form expression for* V *of Theorem 2.2.7 is*

$$\begin{aligned} V = {} & \frac{4\lambda^6}{c} G_2^2 \left[\lambda^2 F_4 F_2^2 G_2^2 - \lambda F_4 F_2 G_2 - \lambda F_3 F_2^2 G_2 + \frac{1}{2} F_3 F_2 + \frac{1}{4} F_4 + \frac{1}{4} F_2^3 \right] \\ & + \frac{4\lambda^6}{1-c} F_2^2 \left[\lambda^2 G_4 G_2^2 F_2^2 - \lambda G_4 G_2 F_2 - \lambda G_3 G_2^2 F_2 + \frac{1}{2} G_3 G_2 + \frac{1}{4} G_4 + \frac{1}{4} G_2^3 \right] \\ & - \frac{4\lambda^6}{c(1-c)} F_2^2 G_2^2 (1 - \lambda F_2 G_2)^2. \end{aligned}$$

Proof. First, the expectation of h is

$$\begin{aligned} \mathbb{E}[h(Y_{(1,2;1,2)})] &= \mathbb{E}[Y_{11} Y_{12} Y_{21} (1 - Y_{22})] \\ &= \mathbb{E}[Y_{11} Y_{12} Y_{21}] - \mathbb{E}[Y_{11} Y_{12} Y_{21} Y_{22}] \\ &= \mathbb{E}[\mathbb{E}[Y_{11} Y_{12} Y_{21} \mid \xi, \eta]] - \mathbb{E}[\mathbb{E}[Y_{11} Y_{12} Y_{21} Y_{22} \mid \xi, \eta]] \\ &= \mathbb{E}[\lambda^3 f(\xi_1)^2 f(\xi_2) g(\xi_1)^2 g(\xi_2)] - \mathbb{E}[\lambda^4 f(\xi_1)^2 f(\xi_2)^2 g(\xi_1)^2 g(\xi_2)^2] \\ &= \lambda^3 F_2 G_2 - \lambda^4 F_2^2 G_2^2. \end{aligned}$$

Now we derive the expression of V. From the expression given by Theorem 2.2.7, we deduce the following form

$$V = \frac{4}{c} \mathbb{E}[h(Y_{(1,2;1,2)}) h(Y_{(1,3;3,4)})] + \frac{4}{1-c} \mathbb{E}[h(Y_{(1,2;1,2)}) h(Y_{(3,4;1,3)})] - \frac{4}{c(1-c)} \mathbb{E}[h(Y_{(1,2;1,2)})]^2.$$

Since the calculation of $\mathbb{E}[h(Y_{(1,2;1,2)}) h(Y_{(1,3;3,4)})]$ and $\mathbb{E}[h(Y_{(1,2;1,2)}) h(Y_{(3,4;1,3)})]$ is completely symmetric in this case, we only need to prove that

$$\mathbb{E}[h(Y_{(1,2;1,2)}) h(Y_{(1,3;3,4)})] = \lambda^6 G_2^2 \left[\lambda^2 F_4 F_2^2 G_2^2 - \lambda F_4 F_2 G_2 - \lambda F_3 F_2^2 G_2 + \frac{1}{2} F_3 F_2 + \frac{1}{4} F_4 + \frac{1}{4} F_2^3 \right]. \tag{2.32}$$

The direct derivation of this quantity is more tedious than technical. Using symmetries and

exchangeability, one can decompose it.

$$\begin{aligned}
\mathbb{E}[h(Y_{(1,2;1,2)})h(Y_{(1,3;3,4)})] &= \frac{1}{16}\mathbb{E}\Bigg[\Big(Y_{11}Y_{12}Y_{21}(1-Y_{22}) + Y_{11}Y_{12}Y_{22}(1-Y_{21}) \\
&\quad + Y_{11}Y_{21}Y_{22}(1-Y_{12}) + Y_{12}Y_{21}Y_{22}(1-Y_{11})\Big) \\
&\quad \times \Big(Y_{13}Y_{14}Y_{33}(1-Y_{34}) + Y_{13}Y_{14}Y_{34}(1-Y_{33}) \\
&\quad + Y_{13}Y_{33}Y_{34}(1-Y_{14}) + Y_{14}Y_{33}Y_{34}(1-Y_{13})\Big)\Bigg] \\
&= \frac{1}{4}\mathbb{E}\Bigg[\Big(Y_{11}Y_{12}Y_{21}(1-Y_{22}) + Y_{11}Y_{21}Y_{22}(1-Y_{12})\Big) \\
&\quad \times \Big(Y_{13}Y_{14}Y_{33}(1-Y_{34}) + Y_{13}Y_{33}Y_{34}(1-Y_{14}))\Big)\Bigg] \\
&= \frac{1}{4}\Big(4\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{14}Y_{33}Y_{34}] \\
&\quad - 4\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{14}Y_{33}] \\
&\quad - 4\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{33}Y_{34}] \\
&\quad + 2\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{13}Y_{33}Y_{34}] \\
&\quad + \mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{13}Y_{14}Y_{33}] \\
&\quad + \mathbb{E}[Y_{11}Y_{21}Y_{22}Y_{13}Y_{33}Y_{34}]\Big).
\end{aligned}$$

Now we derive each simpler expectation.

$$\begin{aligned}
\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{14}Y_{33}Y_{34}] &= \mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{14}Y_{33}Y_{34} \mid \xi, \eta]] \\
&= \mathbb{E}[\lambda^8 f(\xi_1)^4 f(\xi_2)^2 f(\xi_3)^2 g(\eta_1)^2 g(\eta_2)^2 g(\eta_3)^2 g(\eta_4)^2] \\
&= \lambda^8 F_4 F_2^2 G_2^4.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{14}Y_{33}] &= \mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{14}Y_{33} \mid \xi, \eta]] \\
&= \mathbb{E}[\lambda^8 f(\xi_1)^4 f(\xi_2)^2 f(\xi_3) g(\eta_1)^2 g(\eta_2)^2 g(\eta_3)^2 g(\eta_4)] \\
&= \lambda^7 F_4 F_2 G_2^3.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{33}Y_{34}] &= \mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{22}Y_{13}Y_{33}Y_{34} \mid \xi, \eta]] \\
&= \mathbb{E}[\lambda^7 f(\xi_1)^3 f(\xi_2)^2 f(\xi_3)^2 g(\eta_1)^2 g(\eta_2)^2 g(\eta_3)^2 g(\eta_4)] \\
&= \lambda^7 F_3 F_2 G_2^3.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{13}Y_{33}Y_{34}] &= \mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{13}Y_{33}Y_{34} \mid \xi, \eta]] \\
&= \mathbb{E}[\lambda^6 f(\xi_1)^3 f(\xi_2) f(\xi_3)^2 g(\eta_1)^2 g(\eta_2) g(\eta_3)^2 g(\eta_4)] \\
&= \lambda^6 F_3 F_2 G_2^2.
\end{aligned}$$

$$\begin{aligned}\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{13}Y_{14}Y_{33}] &= \mathbb{E}[\mathbb{E}[Y_{11}Y_{12}Y_{21}Y_{13}Y_{14}Y_{33} \mid \xi, \eta]] \\ &= \mathbb{E}[\lambda^6 f(\xi_1)^4 f(\xi_2) f(\xi_3) g(\eta_1)^2 g(\eta_2) g(\eta_3)^2 g(\eta_4)] \\ &= \lambda^6 F_4 G_2^2.\end{aligned}$$

$$\begin{aligned}\mathbb{E}[Y_{11}Y_{21}Y_{22}Y_{13}Y_{33}Y_{34}] &= \mathbb{E}[\mathbb{E}[Y_{11}Y_{21}Y_{22}Y_{13}Y_{33}Y_{34} \mid \xi, \eta]] \\ &= \mathbb{E}[\lambda^6 f(\xi_1)^2 f(\xi_2)^2 f(\xi_3)^2 g(\eta_1)^2 g(\eta_2) g(\eta_3)^2 g(\eta_4)] \\ &= \lambda^6 F_2^3 G_2^2.\end{aligned}$$

Injecting these expressions into (2.32), we find the correct expression for $\mathbb{E}[h(Y_{(1,2;1,2)})h(Y_{(1,3;3,4)})]$, which concludes the proof. □

Lemma 2.H.7. *Let Y be a RCE matrix generated by the Bernoulli-BEDD model, with density λ and degree functions f and g. For $p \geq 1$ and $q \geq 1$, let $h_{p,q}$ be the quadruplet kernel defined by*

$$h_{p,q}(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}) = \prod_{u=1}^{p}\prod_{v=1}^{q} Y_{i_u j_v}.$$

Then $\mathbb{E}[h_{p,q}(Y_{(1,\dots,p;1,\dots,q)})] = \lambda^{pq} F_q^p G_p^q$.

Proof. Direct derivation gives

$$\begin{aligned}\mathbb{E}[h_{p,q}(Y_{(1,\dots,p;1,\dots,q)})] &= \mathbb{E}[\prod_{i=1}^{p}\prod_{j=1}^{q} Y_{ij}] \\ &= \mathbb{E}[\mathbb{E}[\prod_{i=1}^{p}\prod_{j=1}^{q} Y_{ij} \mid \xi, \eta]] \\ &= \mathbb{E}[\prod_{i=1}^{p}\prod_{j=1}^{q} \mathbb{E}[Y_{ij} \mid \xi, \eta]] \\ &= \mathbb{E}[\prod_{i=1}^{p}\prod_{j=1}^{q} \lambda f(\xi_i) g(\eta_j)] \\ &= \lambda^{pq} \prod_{i=1}^{p} \mathbb{E}[f(\xi_i)^q] \prod_{j=1}^{q} \mathbb{E}[g(\eta_j)^p] \\ &= \lambda^{pq} F_q^p G_p^q.\end{aligned}$$

□

Appendix 2.I Some U-statistics written with matrix operations

We denote for all $k \in \mathbb{N}$, $Y^{\odot k}$ the matrix (or vector) Y elevated to the element-wise power k, i.e. $Y_{ij}^{\odot k} = Y_{ij}^k$ for all i and j.

Following formula (2.24), we write all the quadruplet U-statistics considered in the examples described in Sections 2.3.2 and 2.3.3 as simple operations on matrices. $U_N^{h_1}$ and $U_N^{h_2}$ are already given in these sections and

$$\begin{aligned}
U_N^{h_3} &= \frac{1}{m_N(m_N-1)n_N}\left[|Y_N Y_N^T|_1 - \mathrm{Tr}(Y_N Y_N^T)\right],\\
U_N^{h_4} &= \frac{1}{m_N n_N(n_N-1)}\left[|\tilde{Y}_N^T \tilde{Y}_N|_1 - \mathrm{Tr}(\tilde{Y}_N^T \tilde{Y}_N)\right],\\
U_N^{h_5} &= \frac{1}{m_N n_N}|Y_N|_1,\\
U_N^{h_6} &= \frac{1}{m_N n_N(n_N-1)}\left[|\tilde{Y}_N^T Y_N|_1 - \mathrm{Tr}(\tilde{Y}_N^T Y_N)\right].
\end{aligned}$$

where Tr is the trace operator and $\tilde{Y}$ is defined by $\tilde{Y}_{ij} = Y_{ij}^2 - Y_{ij}$.

The motif-counting U-statistic of Section 2.3.4 can be written as

$$\begin{aligned}
U_N^{h_7} = \frac{1}{m_N(m_N-1)n_N(n_N-1)}&\left[|Y_N^T Y_N Y_N^T|_1 - |(Y_N^{\odot 2})^T Y_N|_1 - |Y_N^{\odot 2} Y_N^T|_1 + \mathrm{Tr}(Y_N^{\odot 2} Y_N^T)\right.\\
&\left. -\mathrm{Tr}(Y_N^T Y_N Y_N^T Y_N) + |(Y_N^{\odot 2})^T Y_N^{\odot 2}|_1 + |Y_N^{\odot 2}(Y_N^{\odot 2})^T|_1 - \mathrm{Tr}((Y_N^{\odot 2})^T Y_N^{\odot 2})\right].
\end{aligned}$$

The kernels $h_{p,q}$ are not quadruplet kernels, but they can also be simply computed if $p = 1$ or $q = 1$. We define respectively $r(Y_N)$ and $c(Y_N)$ the vector of row sums (degrees) and the vector of column sums (degrees) of the matrix Y_N. For all $1 \le i \le m_N$, $r(Y_N)_i = \sum_{j=1}^{n_N} Y_{ij}$ and for all $1 \le j \le n_N$, $c(Y_N)_j = \sum_{i=1}^{m_N} Y_{ij}$.

$$\begin{aligned}
U_N^{h_{1,p}} &= \left[m_N \binom{n_N}{p}\right]^{-1} \sum_{i=1}^{m_N} \binom{r(Y_N)_i}{p},\\
U_N^{h_{q,1}} &= \left[\binom{m_N}{q} n_N\right]^{-1} \sum_{j=1}^{n_N} \binom{c(Y_N)_j}{q}.
\end{aligned}$$

We also notice that

$$\begin{aligned}
U_N^{h_{1,1}} &= U_N^{h_5},\\
U_N^{h_{1,2}} &= U_N^{h_1},\\
U_N^{h_{2,1}} &= U_N^{h_3}.
\end{aligned}$$

Chapter 3

Hoeffding-type decomposition for U-statistics on bipartite networks

This chapter corresponds to the following article, submitted in a statistical journal:

Le Minh, T., Donnet, S., Massol, F., and Robin, S. (2023). Hoeffding-type decomposition for U-statistics on bipartite networks. *arXiv preprint arXiv:2308.14518*. https://doi.org/10.48550/arXiv.2308.14518

Abstract. We consider a broad class of random bipartite networks, the distribution of which is invariant under permutation within each type of nodes. We are interested in U-statistics defined on the adjacency matrix of such a network, for which we define a new type of Hoeffding decomposition. This decomposition enables us to characterize non-degenerate U-statistics – which are then asymptotically normal – and provides us with a natural and easy-to-implement estimator of their asymptotic variance.

We illustrate the use of this general approach on some typical random graph models and use it to estimate or test some quantities characterizing the topology of the associated network. We also assess the accuracy and the power of the proposed estimates or tests, via a simulation study.

Contents

3.1. Introduction

Networks are popular objects to represent a set of interacting entities. The last decades have witnessed an explosion in the number of networks datasets. The fields of application range from sociology to ecology, from economics to computer science. Understanding the organization of the network is a first step towards a better insight of the system it represents. Several strategies exist to study or describe the topology of a network. Many of them are based on the calculation of one or several numeric quantities (statistics) such as density, clustering coefficients or counts of given motifs to name but a few. These statistics generally rely on several nodes.

The calculation of these numerical quantities on a given network naturally leads to comparing them to a reference value, or to the value obtained on another network. The concept of hypothesis tests naturally meets this expectation. The challenging step of statistical hypothesis testing is identifying the statistic distribution under the null assumption. In particular, one class of statistic considered are the U-statistics which, in the context of network analysis, have complex dependencies.

Networks and dissociated RCE matrices In networks, entities are represented by nodes which are linked by edges when they interact. In bipartite networks, the nodes are divided into two types and the interactions only happen between nodes of the two different types. Some examples of bipartite networks connect users and items in recommender systems (Zhou et al., 2007), papers and scientists in authorship networks (Newman, 2001), or plants and pollinators in ecological interaction networks (Dormann et al., 2009). The networks are naturally encoded in matrices. In the adjacency matrix Y of a bipartite network (sometimes also called incidence matrix), the two types of nodes are represented by rows and columns, so that Y_{ij} encodes the interaction between entity i of the first type and entity j of the second type. In binary networks, $Y_{ij} = 1$ if i and j interact, else $Y_{ij} = 0$. Some networks are weighted, meaning that Y_{ij} represents the intensity of the interaction.

We consider the asymptotic framework where Y is an infinite adjacency matrix and the adjacency matrix of an observed network of size $m \times n$ is the submatrix extracted from the leading m rows and n columns of Y. Probabilistic models define a joint distribution on the values of the matrix entries. In random graph models, it is common to assume that the nodes of the networks are exchangeable, i.e. that the distribution of the network does not change if its nodes are permuted. For instance, the stochastic block model (Snijders and Nowicki, 1997), the random dot product graph model (Young and Scheinerman, 2007) or the latent space model (Hoff et al., 2002) are all node-exchangeable. On the corresponding adjacency matrix of a bipartite network, this assumption implies the row-column exchangeability. Y is said to be row-column exchangeable (RCE) if for any couple $\Phi = (\sigma_1, \sigma_2)$ of finite permutations of $\mathbb{N}$,

$$\Phi Y \overset{\mathcal{D}}{=} Y,$$

where $\Phi Y := (Y_{\sigma_1(i)\sigma_2(j)})_{i\geq 1, j\geq 1}$. Many exchangeable random graph models also have a dissociatedness property, i.e. their adjacency matrices are also dissociated (Silverman, 1976; Lauritzen et al., 2018). A RCE matrix is said to be dissociated if for all m and n, $(Y_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ is independent from $(Y_{ij})_{i>m, j>n}$. In the present work, we only consider RCE dissociated matrices.

U-statistics and Hoeffding decomposition U-statistics are a generalization of the empirical mean to functions of more than one variable. Many estimators fall under the category of U-statistics. Given a sequence of random variables $Y = (Y_i)_{i\geq 1}$ numbered with a unique index a U-statistic $U_n^h(Y)$ of order n and kernel function h is defined as the following average

$$U_n^h(Y) = \binom{n}{k}^{-1} \sum_{1\leq i_1 < i_2 < \ldots < i_k \leq n} h(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}),$$

where $h : \mathbb{R}^k \to \mathbb{R}$ is a symmetric function referred to as the kernel. Denote $[\![n]\!] := \{1, \ldots, n\}$ and for a set A, $\mathcal{P}_k(A)$ the set of all subsets of cardinal k of A. Let $\mathbf{i} = \{i_1, \ldots, i_k\} \in \mathcal{P}_k([\![n]\!])$, then by

symmetry of h, $h(Y_{i_1}, ..., Y_{i_k})$ does not depend on the order of the elements of $\mathbf{i}$. Therefore, we will denote $h(Y_{\mathbf{i}}) := h(Y_{i_1}, ..., Y_{i_k})$. Finally, the U-statistic $U_n^h(Y)$ can be formulated as follows:

$$U_n^h(Y) = \binom{n}{k}^{-1} \sum_{\mathbf{i} \in \mathcal{P}_k(\llbracket n \rrbracket)} h(Y_{\mathbf{i}}).$$

When Y is an exchangeable sequence, $h(Y_{\mathbf{i}})$ has the same distribution for all $\mathbf{i} \in \mathcal{P}_k(\llbracket n \rrbracket)$, therefore $U_n^h(Y)$ is an unbiased estimate of $h(Y_{\llbracket k \rrbracket})$. The case where Y is an i.i.d. sequence is well-studied: the U-statistics are known to be asymptotically normal (Hoeffding, 1948) and can be used for inference tasks such as estimation and hypothesis testing.

In the i.i.d. case, a useful technique to study the asymptotic behavior of U-statistics is the Hoeffding decomposition, formalized for the first time in Hoeffding (1961). For $1 \leq c \leq k$, define the function $\psi^c h : \mathbb{R}^c \to \mathbb{R}$ as

$$\psi^c h : (y_1, ..., y_c) \longmapsto \mathbb{E}[h(Y_1, ..., Y_k) \mid Y_1 = y_1, ..., Y_c = y_c].$$

Again, by symmetry of h, for some set $\mathbf{i} \in \mathcal{P}_c(\llbracket n \rrbracket)$, we can denote $\psi^c h(Y_{\mathbf{i}}) := \psi^c h(Y_{i_1}, ..., Y_{i_c})$ since the order of the elements of $\mathbf{i}$ does not matter. Set $p^0 h = \mathbb{E}[h(Y_{\llbracket k \rrbracket})]$ and define recursively

$$p^c h(Y_{\mathbf{i}}) = \psi^c h(Y_{\mathbf{i}}) - \sum_{c'=0}^{c-1} \sum_{\mathbf{i}' \in \mathcal{P}_{c'}(\mathbf{i})} p^{c'} h(Y_{\mathbf{i}'}).$$

for all subsets $\mathbf{i} \in \mathcal{P}_c(\llbracket n \rrbracket)$, for all $1 \leq c \leq k$. Then, for $\mathbf{i} \in \mathcal{P}_k(\llbracket n \rrbracket)$, $h(Y_{\mathbf{i}})$ can be written

$$h(Y_{\mathbf{i}}) = \sum_{0 \leq c \leq k} \sum_{\mathbf{i}' \in \mathcal{P}_c(\mathbf{i})} p^c h(Y_{\mathbf{i}'}).$$

The U-statistic U_n^h can be written as

$$U_n^h(Y) = \sum_{c=0}^{k} \binom{k}{c} P_n^c h(Y),$$

where for $1 \leq c \leq k$, $P_n^c h(Y) = \binom{n}{c}^{-1} \sum_{\mathbf{i} \in \mathcal{P}_c(\llbracket n \rrbracket)} p^c h(Y_{\mathbf{i}})$.

This decomposition is interesting as all the quantities $p^c h(Y_{\mathbf{i}})$ are orthogonal. By extension the U-statistics $P_n^c h$ are also orthogonal. The leading terms of this decomposition have been used by Hoeffding (1948) to prove the asymptotic normality of U-statistics. The decomposition also yields a decomposition of the variance of U-statistics.

By extension, a network U-statistic averages a function h defined over sub-matrices of size $p \times q$. Let Y be an infinite adjacency matrix from which we observe the first m rows and n columns. Let $h : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ be a function defined on $p \times q$ matrices, $1 \leq p \leq m$, $1 \leq q \leq n$, verifying the following symmetry property: for all $(\sigma_1, \sigma_2) \in \mathbb{S}_p \times \mathbb{S}_q$,

$$h(Y_{(i_{\sigma_1(1)}, ..., i_{\sigma_1(p)}; j_{\sigma_2(1)}, ..., j_{\sigma_2(q)})}) = h(Y_{(i_1, i_2, ..., i_p; j_1, j_2, ..., j_q)}), \tag{3.1}$$

where $Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}$ is the $p \times q$ sub-matrix consisting of the rows and columns of Y indexed by $i_1, \dots, i_p$ and $j_1, \dots, j_q$ respectively. Therefore, since the order of the elements of $\mathbf{i} = \{i_1, \dots, i_p\}$ and $\mathbf{j} = \{j_1, \dots, j_q\}$ does not matter, we can denote $h(Y_{\mathbf{i},\mathbf{j}}) := h(Y_{(i_1,i_2,\dots,i_p;j_1,j_2,\dots,j_q)})$. Then the associated U-statistic is

$$U^h_{m,n}(Y) = \binom{m}{p}^{-1}\binom{n}{q}^{-1} \sum_{\substack{\mathbf{i}\in\mathcal{P}_p([\![m]\!]) \\ \mathbf{j}\in\mathcal{P}_q([\![n]\!])}} h(Y_{\mathbf{i},\mathbf{j}}). \tag{3.2}$$

Note that the assumption on the symmetry of h can be made without loss of generality. Indeed, if $h^0 : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ is not symmetric, then $h : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ defined by

$$h(Y_{(i_1,i_2,\dots,i_p;j_1,j_2,\dots,j_q)}) = (p!q!)^{-1} \sum_{(\sigma_1,\sigma_2)\in\mathbb{S}_p\times\mathbb{S}_q} h^0(Y_{(i_{\sigma_1(1)},\dots,i_{\sigma_1(p)};j_{\sigma_2(1)},\dots,j_{\sigma_2(q)})}) \tag{3.3}$$

verifies Equation (3.1) and leads to the same U-statistic ($U^{h^0}_{m,n}(Y) = U^h_{m,n}(Y)$).

In the case where Y is a RCE matrix, Le Minh (2023) used a martingale approach to obtain a weak convergence result for $U^h_{m,n}$ when m and n grow to infinity at the same rate. Applying this result requires specific development to get the asymptotic variance. In this paper, we propose an Hoeffding decomposition-based approach. This strategy has the advantage to provide a method to estimate the asymptotic variance of the U-statistics. Indeed, obtaining an estimation of the (asymptotic) variance of U-statistics is a required condition to perform practical inference such as hypothesis testing. However, it remains a complex task and has been tackled with various methods in the literature.

Variance estimation of U-statistics (related work) The standard error of U-statistics is most often computed using resampling techniques such as the jackknife (Arvesen, 1969) and bootstrap (Efron, 1979; Bickel and Freedman, 1981) estimators of variance. Sen (1960, 1977), Callaert and Veraverbeke (1981) and Schucany and Bankson (1989) suggested various estimators of the asymptotic variance. However, all these estimators are biased for both the variance and the asymptotic variances. Callaert and Veraverbeke (1981); Schucany and Bankson (1989) also discussed unbiased estimators, but they are computationally more demanding than all the previous estimators and they find them to have a positive probability of being negative, which is undesirable.

Contribution We show how U-statistics of size $p \times q$ can be used for exchangeable network inference. In particular, we propose a Hoeffding-type decomposition to identify the asymptotic distribution of these U-statistics and build a consistent estimator of their variance. For that, we assume that the matrix dimensions m_N and n_N grow at the same rate with $m_N/N \to \rho$ and

$n_N/N \to 1-\rho$, where $\rho \in]0,1[$. For simplification, we denote $U_N^h(Y) := U_{m_N,n_N}^h(Y)$. First, we show that the distribution of distribution $\sqrt{N}(U_N^h(Y) - \mathbb{E}[h])$ converges to a Gaussian. Next, we build a consistent estimator for the variance of $U_N^h(Y)$ and we study its properties.

Outline In Section 3.2, we first extend the Hoeffding decomposition of a sub-graph statistic. The decomposition we propose is based on the Aldous-Hoover-Kallenberg (AHK) representation. Section 3.3 exploits this Hoeffding decomposition to demonstrate the asymptotic normality of the U-statistics we consider. In Section 3.4, we use the Hoeffding decomposition to build an estimator of the asymptotic variance of the U-statistics. Section 3.5 presents some models for RCE matrices and kernels that can be interesting for network analysis and comparisons. Section 3.6 and 3.7 are dedicated to simulation studies and an illustration of legislature dataset.

3.2. Hoeffding decomposition of a submatrix U-statistic

Aldous-Hoover-Kallenberg (AHK) representation Corollary 7.23 of Kallenberg (2005) states that for any dissociated RCE matrix Y, there exists $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ and $(\zeta_{ij})_{i,j\geq 1}$ arrays of i.i.d. random variables with uniform distribution over $[0,1]$ and a real measurable function ϕ such that for all $1 \leq i,j < \infty$, $Y_{ij} \overset{a.s.}{=} \phi(\xi_i, \eta_j, \zeta_{ij})$. With such a representation, the kernel function taken on a $p \times q$ submatrix indexed by the rows $\mathbf{i} \in \mathcal{P}_p(\mathbb{N})$ and columns $\mathbf{j} \in \mathcal{P}_q(\mathbb{N})$ can be written $h(Y_{\mathbf{i},\mathbf{j}}) \overset{a.s.}{=} h_\phi((\xi_i)_{i\in\mathbf{i}}; (\eta_j)_{j\in\mathbf{j}}; (\zeta_{ij})_{\substack{i\in\mathbf{i}\\ j\in\mathbf{j}}})$, where h_ϕ is some function depending on h and ϕ.

Note that the AHK decomposition is not unique. In the rest of the paper, we assume that for each dissociated RCE matrix Y, we have picked an AHK representation, i.e. a suitable function ϕ, and suitable i.i.d. random variables $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ and $(\zeta_{ij})_{\substack{i\geq 1\\ j\geq 1}}$. In the rest of the paper, we will write, abusively but without ambiguity,

$$Y_{ij} = \phi(\xi_i, \eta_j, \zeta_{ij}).$$

and

$$h(Y_{\mathbf{i},\mathbf{j}}) = h_\phi((\xi_i)_{i\in\mathbf{i}}; (\eta_j)_{j\in\mathbf{j}}; (\zeta_{ij})_{\substack{i\in\mathbf{i}\\ j\in\mathbf{j}}}).$$

For $\mathbf{i}' \in \mathcal{P}(\mathbb{N})$ and $\mathbf{j}' \in \mathcal{P}(\mathbb{N})$, we define the σ-algebra

$$\mathcal{A}_{\mathbf{i}',\mathbf{j}'} := \sigma((\xi_i)_{i\in\mathbf{i}'}, (\eta_j)_{j\in\mathbf{j}'}, (\zeta_{ij})_{\substack{i\in\mathbf{i}'\\ j\in\mathbf{j}'}}).$$

Therefore, it follows from our notations that

$$\mathbb{E}[h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\mathbf{i}',\mathbf{j}'}] = \mathbb{E}[h_\phi((\xi_i)_{i\in\mathbf{i}}; (\eta_j)_{j\in\mathbf{j}}; (\zeta_{ij})_{\substack{i\in\mathbf{i}\\ j\in\mathbf{j}}}) \mid (\xi_i)_{i\in\mathbf{i}'}; (\eta_j)_{j\in\mathbf{j}'}; (\zeta_{ij})_{\substack{i\in\mathbf{i}'\\ j\in\mathbf{j}'}}].$$

Now, notice that for fixed $\mathbf{i}'$ and $\mathbf{j}'$, the quantity $\mathbb{E}[h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\mathbf{i}',\mathbf{j}'}]$ only depends on the elements shared by $\mathbf{i}$ and $\mathbf{i}'$ and the elements shared by $\mathbf{j}$ and $\mathbf{j}'$, and not on the other elements of $\mathbf{i}$, $\mathbf{i}'$, $\mathbf{j}$ and $\mathbf{j}'$. Suppose $r = \mathrm{Card}(\mathbf{i} \cap \mathbf{i}')$ and $c = \mathrm{Card}(\mathbf{j} \cap \mathbf{j}')$. Without loss of generality, we can assume that $\mathbf{i}' \in \mathcal{P}_r(\mathbf{i})$ and $\mathbf{j}' \in \mathcal{P}_c(\mathbf{j})$ so $\mathbb{E}[h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\mathbf{i}',\mathbf{j}'}]$ only depends on the r elements of $\mathbf{i}'$ and the c elements of $\mathbf{j}'$. Therefore, we can define the quantities $\psi^{r,c}h(Y_{\mathbf{i}',\mathbf{j}'})$ such that

$$\psi^{r,c}h(Y_{\mathbf{i}',\mathbf{j}'}) := \mathbb{E}[h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\mathbf{i}',\mathbf{j}'}],$$

where the choice of $\mathbf{i}$ and $\mathbf{j}$ does not matter as long as $\mathbf{i}' \subset \mathbf{i}$ and $\mathbf{j}' \subset \mathbf{j}$. Note that $\psi^{r,c}h(Y_{\mathbf{i}',\mathbf{j}'})$ is simply a notation and not a function of $Y_{\mathbf{i}',\mathbf{j}'}$. If $\mathbf{i}' = \varnothing$ or $\mathbf{j}' = \varnothing$, we will still use this notation, for example

$$\psi^{r,c}h(Y_{\mathbf{i}',\varnothing}) = \mathbb{E}[h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\mathbf{i}',\varnothing}],$$

despite $Y_{\mathbf{i}',\varnothing}$ being undefined.

Hoeffding-type decomposition of the kernel In the following, for elements of $\mathbb{N}^2$, $(x,y) \leq (x',y')$ means that both $x \leq x'$ and $y \leq y'$; $(x,y) < (x',y')$ means that, in addition, $(x,y) \neq (x',y')$.

For all $\mathbf{i} \in \mathcal{P}_r(\mathbb{N})$ and $\mathbf{j} \in \mathcal{P}_c(\mathbb{N})$, we define by recursion the following quantity

$$p^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) = \psi^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) - \sum_{(0,0)\leq(r',c')<(r,c)} \sum_{\substack{\mathbf{i}'\in\mathcal{P}_{r'}(\mathbf{i}) \\ \mathbf{j}'\in\mathcal{P}_{c'}(\mathbf{j})}} p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'}). \tag{3.4}$$

Since $\psi^{p,q}h(Y_{\mathbf{i},\mathbf{j}}) = h(Y_{\mathbf{i},\mathbf{j}})$ for $\mathbf{i} \in \mathcal{P}_p(\mathbb{N})$ and $\mathbf{j} \in \mathcal{P}_q(\mathbb{N})$, (3.4) yields the decomposition of the kernel function h

$$h(Y_{\mathbf{i},\mathbf{j}}) = \sum_{(0,0)\leq(r,c)\leq(p,q)} \sum_{\substack{\mathbf{i}'\in\mathcal{P}_{r}(\mathbf{i}) \\ \mathbf{j}'\in\mathcal{P}_{c}(\mathbf{j})}} p^{r,c}h(Y_{\mathbf{i}',\mathbf{j}'}). \tag{3.5}$$

Remark 1. From this formula, we see that $h(Y_{\mathbf{i},\mathbf{j}})$ is a linear combination of the $\left(p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'})\right)_{\substack{0\leq r'\leq p, 0\leq c'\leq q \\ \mathbf{i}'\in\mathcal{P}_{r'}(\mathbf{i}),\mathbf{j}'\in\mathcal{P}_{c'}(\mathbf{j})}}$, therefore, this is a linear combination of the $\left(\psi^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'})\right)_{\substack{0\leq r'\leq r, 0\leq c'\leq c \\ \mathbf{i}'\in\mathcal{P}_{r'}(\mathbf{i}),\mathbf{j}'\in\mathcal{P}_{c'}(\mathbf{j})}}$.

Now, we show that for $\mathbf{i}' \subset \mathbf{i}$ and $\mathbf{j}' \subset \mathbf{j}$, $p^{r,c}h(Y_{\mathbf{i}',\mathbf{j}'})$ is actually the projection of $h(Y_{\mathbf{i},\mathbf{j}})$ on the subspace generated by L_2 functions of all the AHK variables of $\mathcal{A}_{\mathbf{i}',\mathbf{j}'}$, orthogonally to the sub-spaces generated by L_2 functions of all the variables of $\mathcal{A}_{\mathbf{i}'',\mathbf{j}''}$, for $\mathbf{i}'' \subset \mathbf{i}'$ and $\mathbf{j}'' \subset \mathbf{j}'$. This system of projection is analogous to the Hoeffding decomposition for the kernel functions of usual U-statistics on i.i.d. data. We prove that the following orthogonality properties hold.

Proposition 3.2.1. *Let h_1 and h_2 two kernel functions of respective size $p_1 \times q_1$ and $p_2 \times q_2$.*

1. *Let $(0,0) \leq (r_1,c_1) \leq (p_1,q_1)$ and $(0,0) \leq (r_2,c_2) \leq (p_2,q_2)$ such that $(r_1,c_1) \neq (r_2,c_2)$. Let $(\mathbf{i}_1,\mathbf{j}_1) \in \mathcal{P}_{r_1}(\llbracket m \rrbracket) \times \mathcal{P}_{c_1}(\llbracket n \rrbracket)$ and $(\mathbf{i}_2,\mathbf{j}_2) \in \mathcal{P}_{r_2}(\llbracket m \rrbracket) \times \mathcal{P}_{c_2}(\llbracket n \rrbracket)$, then*

$$\mathrm{Cov}(p^{r_1,c_1}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1}), p^{r_2,c_2}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2})) = 0.$$

2. *Let* (r,c) *such that* $(0,0) \leq (r,c) \leq (p_1,q_1)$ *and* $(0,0) \leq (r,c) \leq (p_2,q_2)$. *Let* $(\mathbf{i}_1,\mathbf{j}_1)$ *and* $(\mathbf{i}_2,\mathbf{j}_2)$ *two elements of* $\mathcal{P}_r([\![m]\!]) \times \mathcal{P}_c([\![n]\!])$. *If* $(\mathbf{i}_1,\mathbf{j}_1) \neq (\mathbf{i}_2,\mathbf{j}_2)$, *then*

$$\mathrm{Cov}(p^{r,c}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1}), p^{r,c}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2})) = 0.$$

This proposition relies on the fact that the projections are "conditionnally centered". This property is given by the following lemma, the proof of which is provided in Appendix 3.B.

Lemma 3.2.2. *Let* h *be a kernel function of size* $p \times q$. *Let* $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_r(\mathbb{N}) \times \mathcal{P}_c(\mathbb{N})$, *where* $(0,0) < (r,c) \leq (p,q)$. *For all* $\underline{\mathbf{i}} \subset \mathbf{i}$ *and* $\underline{\mathbf{j}} \subset \mathbf{j}$, *we have*

$$\mathbb{E}[p^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] = 0.$$

Proof of Proposition 3.2.1. The two properties are proven similarly and derive from the fact that $(\mathbf{i}_1,\mathbf{j}_1) \neq (\mathbf{i}_2,\mathbf{j}_2)$. This is true for both properties.

Consider any (possibly equal) (r_1,c_1) and (r_2,c_2) and associated $(\mathbf{i}_1,\mathbf{j}_1) \neq (\mathbf{i}_2,\mathbf{j}_2)$. Then $\mathbf{i}_1 \neq \mathbf{i}_2$ or $\mathbf{j}_1 \neq \mathbf{j}_2$. Without loss of generality, assume that $\mathbf{i}_1 \neq \mathbf{i}_2$ so there is an element $i_2 \in \mathbf{i}_2$ which is not included in $\mathbf{i}_1$. Then

$$\begin{aligned}
\mathbb{E}[p^{r_1,c_1}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1})p^{r_2,c_2}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2})] &= \mathbb{E}[\mathbb{E}[p^{r_1,c_1}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1})p^{r_2,c_2}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2}) \mid \mathcal{A}_{\{i_2\},\varnothing}]] \\
&= \mathbb{E}[p^{r_1,c_1}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1})\mathbb{E}[p^{r_2,c_2}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2}) \mid \mathcal{A}_{\{i_2\},\varnothing}]] \\
&= 0,
\end{aligned}$$

since $\mathbb{E}[p^{r_2,c_2}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2}) \mid \mathcal{A}_{\{i_2\},\varnothing}] = 0$ from Lemma 3.2.2.

We have proven both properties, since the case $(r_1,c_1) \neq (r_2,c_2)$ corresponds to the first property and the case $(r_1,c_1) = (r_2,c_2)$ corresponds to the second property. □

Decomposition of U-statistics Using the Hoeffding-type decomposition of kernel functions (3.5) the U-statistic (3.2) can be reformulated as:

$$\begin{aligned}
U^h_{m,n}(Y) &= \binom{m}{p}^{-1}\binom{n}{q}^{-1} \sum_{\substack{1\leq i_1<\ldots<i_p\leq m \\ 1\leq j_1<\ldots<j_q\leq n}} \sum_{(0,0)\leq(r,c)\leq(p,q)} \sum_{\substack{\mathbf{i}\in\mathcal{P}_r(\{i_1,\ldots,i_p\}) \\ \mathbf{j}\in\mathcal{P}_c(\{j_1,\ldots,j_q\})}} p^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) \\
&= \binom{m}{p}^{-1}\binom{n}{q}^{-1} \sum_{(0,0)\leq(r,c)\leq(p,q)} \binom{m-r}{p-r}\binom{n-c}{q-c} \sum_{\substack{\mathbf{i}\in\mathcal{P}_r([\![m]\!]) \\ \mathbf{j}\in\mathcal{P}_c([\![n]\!])}} p^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) \\
&= \sum_{(0,0)\leq(r,c)\leq(p,q)} \binom{p}{r}\binom{q}{c}\binom{m}{r}^{-1}\binom{n}{c}^{-1} \sum_{\substack{\mathbf{i}\in\mathcal{P}_r([\![m]\!]) \\ \mathbf{j}\in\mathcal{P}_c([\![n]\!])}} p^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) \\
&= \sum_{(0,0)\leq(r,c)\leq(p,q)} \binom{p}{r}\binom{q}{c} P^{r,c}_{m,n}h(Y)
\end{aligned}$$

where for all $0 \le r \le p$ and $0 \le c \le q$,

$$P^{r,c}_{m,n} h(Y) = \binom{m}{r}^{-1} \binom{n}{c}^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_r(\llbracket m \rrbracket) \\ \mathbf{j} \in \mathcal{P}_c(\llbracket n \rrbracket)}} p^{r,c} h(Y_{\mathbf{i},\mathbf{j}})$$

is the U-statistic of kernel function $p^{r,c}h$ taken on the first $m \times n$ rows and columns of the matrix Y. A consequence of the orthogonality of the projections of h is the orthogonality of these U-statistics, as stated in the following corollary which is proven in Appendix 3.B.

Corollary 3.2.3. *Let h_1 and h_2 two kernel functions of respective sizes $p_1 \times q_1$ and $p_2 \times q_2$. Let $(0,0) \le (r_1, c_1) < (p_1, q_1)$ and $(0,0) \le (r_2, c_2) < (p_2, q_2)$.*

1. *If $(r_1, c_1) \neq (r_2, c_2)$, then*

$$\mathrm{Cov}(P^{r_1,c_1}_{m,n} h_1(Y), P^{r_2,c_2}_{m,n} h_2(Y)) = 0.$$

2. *If $(r_1, c_1) = (r_2, c_2) = (r, c)$, then*

$$\mathrm{Cov}(P^{r,c}_{m,n} h_1(Y), P^{r,c}_{m,n} h_2(Y)) = \binom{m}{r}^{-1} \binom{n}{c}^{-1} \mathrm{Cov}(p^{r,c} h_1(Y_{\llbracket r \rrbracket, \llbracket c \rrbracket}), p^{r,c} h_2(Y_{\llbracket r \rrbracket, \llbracket c \rrbracket})).$$

The orthogonality between the $P^{r_1,c_1}_{m,n} h_1(Y)$ and $P^{r_2,c_2}_{m,n} h_2(Y)$ allows to decompose the covariance of two U-statistics into a few covariance terms.

Corollary 3.2.4.

$$\begin{aligned}
&\mathrm{Cov}(U^{h_1}_{m,n}(Y), U^{h_2}_{m,n}(Y)) \\
&\quad = \sum_{(0,0) < (r,c) \le (p,q)} \binom{p}{r}^2 \binom{q}{c}^2 \mathrm{Cov}(P^{r,c}_{m,n} h_1(Y), P^{r,c}_{m,n} h_2(Y)) \\
&\quad = \sum_{(0,0) < (r,c) \le (p,q)} \binom{p}{r}^2 \binom{q}{c}^2 \binom{m}{r}^{-1} \binom{n}{c}^{-1} \mathrm{Cov}(p^{r,c} h_1(Y_{\llbracket r \rrbracket, \llbracket c \rrbracket}), p^{r,c} h_2(Y_{\llbracket r \rrbracket, \llbracket c \rrbracket})).
\end{aligned}$$

Corollary 3.2.4 is a direct consequence of Corollary 3.2.3. This result will be helpful when considering the asymptotic properties of the U-statistics. Indeed, one can see that each covariance term is associated with a binomial coefficient depending on different orders of m and n.

3.3. Asymptotic normality of U-statistics

For the following sections, we will use simplified notations, summarizing the couple (m_N, n_N) into N. We recall that $U^h_N(Y) = U^h_{m_N,n_N}(Y)$. We also denote $P^{r,c}_N h(Y) := P^{r,c}_{m_N,n_N} h(Y)$. When

this is unambiguous, we will omit to mention Y, so we will simply write U_N^h, $P_N^{r,c}h$, $p_{(\mathbf{i},\mathbf{j})}^{r,c}h$ and $\psi_{(\mathbf{i},\mathbf{j})}^{r,c}h$ instead of $U_N^h(Y)$, $P_N^{r,c}h(Y)$, $p^{r,c}h(Y_{\mathbf{i},\mathbf{j}})$ and $\psi^{r,c}h(Y_{\mathbf{i},\mathbf{j}})$.

The following theorem is a Central Limit Theorem for U_N^h. Denote $v_h^{r,c} := \mathbb{V}[\psi_{(\llbracket r\rrbracket,\llbracket c\rrbracket)}^{r,c}h]$.

Theorem 3.3.1. *Let Y be a dissociated RCE matrix. Let h be a $p\times q$ kernel function such that $\mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2] < \infty$. Let $(m_N,n_N)_{N\geq 1}$ be a sequence of dimensions for the U-statistics, such that $\frac{m_N}{N} \xrightarrow[N\to\infty]{} \rho$ and $\frac{n_N}{N} \xrightarrow[N\to\infty]{} 1-\rho$, where $\rho \in]0,1[$. Let $(U_N^h)_{N\geq 1}$ be the sequence of U-statistics associated with h defined by $U_N^h := U_{m_N,n_N}^h$. Set $U_\infty^h = \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})]$ and*

$$V^h = \frac{p^2}{\rho}v_h^{1,0} + \frac{q^2}{1-\rho}v_h^{0,1}.$$

If $V^h > 0$, then

$$\sqrt{N}(U_N^h - U_\infty^h) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,V^h).$$

This theorem comes from the decomposition of $\sqrt{N}(U_N^h - U_\infty^h)$ into three different terms, the limits of which are given by the following lemmas (proofs in Appendix 3.C).

Lemma 3.3.2. *If $v_h^{1,0} > 0$, then we have*

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m} p_{(\{i\},\varnothing)}^{1,0}h \xrightarrow[m\to\infty]{\mathcal{D}} \mathcal{N}(0,v_h^{1,0})$$

and if $v_h^{0,1} > 0$, then we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} p_{(\{j\},\varnothing)}^{0,1}h \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,v_h^{0,1})$$

Lemma 3.3.3. *Let $A_N := \sqrt{N}\sum_{\substack{(0,0)<(r,c)\leq(p,q)\\(r,c)\neq(1,0)\neq(0,1)}} \binom{p}{r}\binom{q}{c}P_N^{r,c}h$. Then $A_N \xrightarrow[N\to\infty]{\mathbb{P}} 0$.*

Proof of Theorem 3.3.1. We have

$$\begin{aligned} U_N^h &= \sum_{(0,0)\leq(r,c)\leq(p,q)} \binom{p}{r}\binom{q}{c}P_N^{r,c}h \\ &= P_N^{0,0}h + pP_N^{1,0}h + qP_N^{0,1}h + \sum_{\substack{(0,0)<(r,c)\leq(p,q)\\(r,c)\neq(1,0)\neq(0,1)}} \binom{p}{r}\binom{q}{c}P_N^{r,c}h. \end{aligned}$$

First, we see that $P_N^{0,0}h = U_\infty^h$. Next, A_N being defined in Lemma 3.3.3, we have

$$\sqrt{N}(U_N^h - U_\infty^h) = \frac{\sqrt{N}p}{m_N}\sum_{i=1}^{m_N} p_{(\{i\},\varnothing)}^{1,0}h + \frac{\sqrt{N}q}{n_N}\sum_{j=1}^{n_N} p_{(\varnothing,\{j\})}^{0,1}h + A_N.$$

From Lemma 3.3.3, $A_N \xrightarrow[N\to\infty]{\mathbb{P}} 0$. So by Slutsky's theorem, $\sqrt{N}(U_N^h - U_\infty^h)$ has the same limiting distribution as the two main terms of this decomposition. From Lemma 3.3.2, this is the sum of two centered Gaussians of respective variance $\frac{p^2}{\rho}v_h^{1,0}$ and $\frac{q^2}{1-\rho}v_h^{0,1}$. Furthermore, $\sum_{i=1}^m p^{1,0}_{(\{i\},\varnothing)}h$ and $\sum_{j=1}^n p^{0,1}_{(\{j\},\varnothing)}h$ are independent, so the two Gaussians are independent, which concludes the proof. □

Remark 2. The expression of V^h could have been predicted with Corollary 3.2.4. Indeed, this corollary implies that $\mathbb{V}[U_N^h] = \sum_{(0,0)<(r,c)\le(p,q)} \binom{p}{r}^2\binom{q}{c}^2\binom{m_N}{r}^{-1}\binom{n_N}{c}^{-1}\mathbb{V}[p^{r,c}_{([\![r]\!],[\![c]\!])}h]$, so $\lim_{N\to\infty} N\mathbb{V}[U_N^h] = \lim_{N\to\infty}\left(\frac{p^2N}{m_N}\mathbb{V}[p^{1,0}_{(\{1\},\varnothing)}h] + \frac{q^2N}{n_N}\mathbb{V}[p^{0,1}_{(\varnothing,\{1\})}h]\right) = \frac{p^2}{c}\mathbb{V}[\psi^{1,0}_{(\{1\},\varnothing)}h] + \frac{q^2}{1-c}\mathbb{V}[\psi^{0,1}_{(\varnothing,\{1\})}h]$.

Now, we show that the limiting distribution of a vector of U-statistics is a multivariate normal distribution under the condition that all the kernel functions are linearly independent. However, if the kernels functions are of different sizes, the notion of linear independence is unclear. We need to define the concept of kernel extension to enunciate the corollary.

Definition 3.3.4. *Let h be a kernel function of size $p\times q$. Let $p'\ge p$ and $q'\ge q$. We define the extension of h to the size $p'\times q'$ by $\tilde{h}$ such that for all $\mathbf{i}'\in\mathcal{P}_{p'}(\mathbb{N})$ and $\mathbf{j}'\in\mathcal{P}_{q'}(\mathbb{N})$,*

$$\tilde{h}(Y_{\mathbf{i}',\mathbf{j}'}) = \left[\binom{p'}{p}\binom{q'}{q}\right]^{-1}\sum_{\substack{\mathbf{i}\subset\mathcal{P}_p(\mathbf{i}')\\ \mathbf{j}\subset\mathcal{P}_q(\mathbf{j}')}} h(Y_{\mathbf{i},\mathbf{j}}).$$

The extension of a kernel actually shares some properties with its kernel, as shown in Lemma 3.C.1. With this definition, we can define the linear independence of kernel functions needed for the following corollary as the linear independence of their kernel extensions. Denote $c^{r,c}_{h_k,h_\ell} := \mathrm{Cov}\left(\psi^{r,c}_{([\![r]\!],[\![c]\!])}h_k, \psi^{r,c}_{([\![r]\!],[\![c]\!])}h_\ell\right)$. The proof of this corollary can be found in Appendix 3.C.

Corollary 3.3.5. *Let Y be a dissociated RCE matrix. Let $(h_1,h_2,...,h_D)$ be a vector of kernel functions of respective sizes $p_1\times q_1, p_2\times q_2, ..., p_D\times q_D$ such that*

1. *Theorem 3.3.1 applies for each kernel function, i.e. $\mathbb{E}[h_k(Y_{(1,2;1,2)})^2]<\infty$ and $U_\infty^{h_k}$ and V^{h_k} are as defined in Theorem 3.3.1 for each kernel h_k, $1\le k\le D$,*
2. *for $t\in\mathbb{R}^D$, $\sum_{k=1}^D t_k\tilde{h}_k \equiv 0$ if and only if $t=(0,...,0)$, where for $1\le k\le D$, $\tilde{h}_k$ is the extension of h_k to size $\max_k(p_k)\times\max_k(q_k)$.*

Then

$$\sqrt{N}\left(\begin{pmatrix}U_N^{h_1}\\ U_N^{h_2}\\ ...\\ U_N^{h_D}\end{pmatrix} - \begin{pmatrix}U_\infty^{h_1}\\ U_\infty^{h_2}\\ ...\\ U_\infty^{h_D}\end{pmatrix}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,\Sigma^{h_1,...,h_D}),$$

with

$$\Sigma^{h_1,\dots,h_D} = \left(C^{h_k,h_\ell}\right)_{1\le k,\ell\le D},$$

where $C^{h_k,h_\ell} = \frac{p^2}{\rho} c^{1,0}_{h_k,h_\ell} + \frac{q^2}{1-\rho} c^{0,1}_{h_k,h_\ell}$ *for all* $1 \le k,\ell \le D$ *(and* $C^{h_k,h_k} = V^{h_k}$*).*

Although, through the first condition, the theorem requires the kernel functions of $(h_1, h_2, \dots, h_D)$ to be linearly independent, the corresponding U-statistics are not independent random variables, even asymptotically, because $\Sigma^{h_1,\dots,h_D}$ is not a diagonal matrix. One consequence of this corollary is that Theorem 3.3.1 can be extended to differentiable functions of U-statistics.

Corollary 3.3.6. *Let* $h_1, \dots, h_D$ *be* D *kernel functions such that Corollary 3.3.5 applies and* $\Sigma^{h_1,\dots,h_D} = \left(C^{h_k,h_\ell}\right)_{1\le k,\ell\le D}$. *Denote* $\theta = (U^{h_1}_\infty, \dots, U^{h_D}_\infty)$. *Let* $g : \mathbb{R}^d \to \mathbb{R}$ *be a differentiable function at* θ. *Denote* ∇g *the gradient of* g *and* $V^\delta := \nabla g(\theta)^T \Sigma^{h_1,\dots,h_D} \nabla g(\theta)$. *If* $V^\delta > 0$, *then*

$$\sqrt{N}(g(U^{h_1}_N, \dots, U^{h_D}_N) - g(\theta)) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V^\delta).$$

Proof. The first-order Taylor expansion of g at θ is written

$$g(U^{h_1}_N, \dots, U^{h_D}_N) - g(\theta) = \nabla g(\theta)^T \left((U^{h_1}_N, \dots, U^{h_D}_N) - \theta\right) + o_P\left(\|(U^{h_1}_N, \dots, U^{h_D}_N) - \theta\|\right).$$

From Corollary 3.3.5, $\sqrt{N}\left((U^{h_1}_N, \dots, U^{h_D}_N) - \theta\right)$ converges to a multivariate normal distribution with asymptotic covariance matrix $\Sigma^{h_1,\dots,h_D}$, so the delta method (see Theorem 3.1 of Van der Vaart, 2000) can be applied to prove this proposition. □

3.4. Estimation of the asymptotic variance of a non-degenerated U-statistic

Theorem 3.3.1 shows the asymptotic normality of RCE submatrix U-statistics. In order to perform statistical inference using these U-statistics, one needs to estimate their variances. We see that

$$\mathbb{V}[U^h_N] = \left[\binom{m}{p}\binom{n}{q}\right]^{-2} \sum_{\substack{\mathbf{i}\in\mathcal{P}_p([\![m]\!])\\ \mathbf{j}\in\mathcal{P}_q([\![n]\!])}} \sum_{\substack{\mathbf{i}'\in\mathcal{P}_p([\![m]\!])\\ \mathbf{j}'\in\mathcal{P}_q([\![n]\!])}} \mathrm{Cov}\left(h(Y_{\mathbf{i},\mathbf{j}}), h(Y_{\mathbf{i}',\mathbf{j}'})\right).$$

By exchangeability, the covariance between the kernels $h(Y_{\mathbf{i},\mathbf{j}})$ and $h(Y_{\mathbf{i}',\mathbf{j}'})$ only depends on the number of row and column indices they share. Denote $\gamma^{r,c}_h := \mathrm{Cov}\left(h(Y_{\mathbf{i},\mathbf{j}}), h(Y_{\mathbf{i}',\mathbf{j}'})\right)$ where $Y_{\mathbf{i},\mathbf{j}}$ and $Y_{\mathbf{i}',\mathbf{j}'}$ share r row indices and c column indices: $\mathrm{Card}(\mathbf{i}\cap\mathbf{i}') = r$ and $\mathrm{Card}(\mathbf{j}\cap\mathbf{j}') = c$. We get

$$\mathbb{V}[U^h_N] = \left[\binom{m}{p}\binom{n}{q}\right]^{-1} \sum_{(0,0)\le(r,c)\le(p,q)} \binom{p}{r}\binom{q}{c}\binom{m-p}{p-r}\binom{n-q}{q-c}\gamma^{r,c}_h.$$

Each $\gamma_h^{r,c}$, $1 \le r \le p, 1 \le c \le q$, can be estimated using empirical covariance estimators, between kernel terms that share r rows and c columns and in particular. This leads to an unbiased estimator of $\mathbb{V}[U_N^h]$ similar to the one discussed by Schucany and Bankson (1989) for U-statistics of one-dimensional i.i.d. arrays. However, the estimation of these covariances is computationally intensive and the estimators can take negative values, which can lead to a negative variance estimation.

One approach is to estimate the asymptotic variance V^h. The asymptotic variance formula given by Theorem 3.3.1 is $V^h = \frac{p^2}{\rho} v_h^{1,0} + \frac{q^2}{1-\rho} v_h^{0,1}$. Remark that $v_h^{1,0} = \gamma_h^{1,0}$ and $v_h^{0,1} = \gamma_h^{0,1}$. It is often tedious to analytically calculate V^h, especially as it depends on h and the distribution of Y, see for example Section 3 of Le Minh (2023).

We present here a kernel-free and model-free estimator of V^h, taking advantage of the Hoeffding decomposition. Indeed, we will first define estimators for the conditional expectations $(\psi^{1,0}_{(\{i\},\varnothing)}h)_{1\le i\le m}$ and $(\psi^{0,1}_{(\varnothing,\{j\})}h)_{1\le j\le n}$. Then, using the fact that $v_h^{1,0} = \mathbb{V}[\psi^{1,0}_{(\{1\},\varnothing)}h]$ and $v_h^{0,1} = \mathbb{V}[\psi^{0,1}_{(\varnothing,\{1\})}h]$, we can derive a positive estimator for V^h.

V^h is the asymptotic variance of a single U-statistic. Afterwards, we explain how to generalize this estimator to estimate the asymptotic variance of a function of U-statistics to apply Corollary 3.3.6.

3.4.1. Some useful notations and results

First, we introduce further notations and a helpful lemma for this section.

For some $N > 0$, the size of the overall RCE matrix is $m_N \times n_N$. i being a row index means that $1 \le i \le m_N$ and j being a column index means that $1 \le j \le n_N$.

We denote

$$X_{\mathbf{i},\mathbf{j}} := h(Y_{\mathbf{i},\mathbf{j}}).$$

For N such that $m_N \ge p$ and $n_N \ge q$, we further denote

$$\mathcal{S}_N^{p,q} := \{(\mathbf{i},\mathbf{j}) : \mathbf{i} \in \mathcal{P}_p(\llbracket m_N \rrbracket), \mathbf{j} \in \mathcal{P}_q(\llbracket n_N \rrbracket)\},$$

so that the set of the kernels taken on all the $p \times q$ submatrices can be written as $(X_{\mathbf{i},\mathbf{j}})_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_N^{p,q}}$.

Let $\underline{\mathbf{i}}$ be a set of row indices of size $\underline{p}$ such that $0 \le \underline{p} \le p$ and $\underline{\mathbf{j}}$ a set of column indices of size $\underline{q}$ such that $0 \le \underline{q} \le q$. The subset of $\mathcal{S}_N^{p,q}$ where $\underline{\mathbf{i}}$ is included in the row indices and $\underline{\mathbf{j}}$ is included in the column indices is denoted

$$\mathcal{S}^{p,q}_{N,(\underline{\mathbf{i}},\underline{\mathbf{j}})} := \left\{(\mathbf{i},\mathbf{j}) \in \mathcal{S}_N^{p,q} : \underline{\mathbf{i}} \subset \mathbf{i}, \underline{\mathbf{j}} \subset \mathbf{j}\right\}.$$

For example, the set of the kernels taken on all the $p \times q$ submatrices containing the columns 1 and 2 can be written $(X_{\mathbf{i},\mathbf{j}})_{(\mathbf{i},\mathbf{j}) \in \mathcal{S}^{p,q}_{N,(\varnothing,\{1,2\})}}$.

It is obvious that

$$\mathrm{Card}(\mathcal{S}^{p,q}_N) = \binom{m_N}{p}\binom{n_N}{q} \qquad \text{and} \qquad \mathrm{Card}(\mathcal{S}^{p,q}_{N,(\underline{\mathbf{i}},\underline{\mathbf{j}})}) = \binom{m_N - \underline{p}}{p - \underline{p}}\binom{n_N - \underline{q}}{q - \underline{q}}.$$

Let $I_K = (\underline{\mathbf{i}}_1, ..., \underline{\mathbf{i}}_K)$ be a K-uplet of sets of row indices and $J_K = (\underline{\mathbf{j}}_1, ..., \underline{\mathbf{j}}_K)$ a K-uplet of subsets of column indices. Denote

$$\mathcal{T}^{p,q}_{N,(I_K,J_K)} := \mathcal{S}^{p,q}_{N,(\underline{\mathbf{i}}_1,\underline{\mathbf{j}}_1)} \times \mathcal{S}^{p,q}_{N,(\underline{\mathbf{i}}_2,\underline{\mathbf{j}}_2)} \times ... \times \mathcal{S}^{p,q}_{N,(\underline{\mathbf{i}}_K,\underline{\mathbf{j}}_K)}.$$

In the rest of the paper, we will often see averages of the type

$$T^{p,q}_N(I_K, J_K) := \frac{1}{\prod_{k=1}^K \mathrm{Card}(\mathcal{T}^{p,q}_{N,(I_K,J_K)})} \sum_{\mathcal{T}^{p,q}_{N,(I_K,J_K)}} \mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1} X_{\mathbf{i}_2,\mathbf{j}_2} ... X_{\mathbf{i}_K,\mathbf{j}_K}]. \tag{3.6}$$

As a remark, $T^{p,q}_N((\varnothing),(\varnothing)) = U^h_N$.

By exchangeability, the quantities $\mathbb{E}[X_{(\mathbf{i}_1,\mathbf{j}_1)} X_{(\mathbf{i}_2,\mathbf{j}_2)} ... X_{(\mathbf{i}_K,\mathbf{j}_K)}]$ do not depend on the row indices that do not belong to any pairwise intersection of the $(\mathbf{i}_k)_{1 \le k \le K}$. The same holds for column indices that do not belong to any pairwise intersection of the $(\mathbf{j}_k)_{1 \le k \le K}$. Therefore, assuming $m_N \ge \mathrm{Card}(\cup_{k=1}^K \underline{\mathbf{i}}_k)$ and $n_N \ge \mathrm{Card}(\cup_{k=1}^K \underline{\mathbf{j}}_k)$, we can define

$$\alpha(I_K, J_K) := \mathbb{E}[X_{\bar{\underline{\mathbf{i}}}_1,\bar{\underline{\mathbf{j}}}_1} X_{\bar{\underline{\mathbf{i}}}_2,\bar{\underline{\mathbf{j}}}_2} ... X_{\bar{\underline{\mathbf{i}}}_K,\bar{\underline{\mathbf{j}}}_K}] \tag{3.7}$$

where for $1 \le k \le K$, the p-uplet $\bar{\underline{\mathbf{i}}}_k$ only consists of elements of $\underline{\mathbf{i}}_k$ and elements that are not in any of the other $\underline{\mathbf{i}}_{k'}$, i.e. the $\bar{\underline{\mathbf{i}}}_k$ are of the form $\bar{\underline{\mathbf{i}}}_k = \underline{\mathbf{i}}_k \cup \tilde{\underline{\mathbf{i}}}_k$ where $\cap_{k=1}^K \tilde{\underline{\mathbf{i}}}_k = \varnothing$ and $(\cup_{k=1}^K \tilde{\underline{\mathbf{i}}}_k) \cap (\cup_{k=1}^K \underline{\mathbf{i}}_k) = \varnothing$.

The following lemma will be helpful in later proofs as it provides the asymptotic behaviour of $T^{p,q}_N(I_K, J_K)$. It shows that these averages can be reduced to one dominant expectation term given by $\alpha(I_K, J_K)$ and a remainder vanishing as N grows.

Lemma 3.4.1. *Let $I_K = (\underline{\mathbf{i}}_1, ..., \underline{\mathbf{i}}_K)$ and $J_K = (\underline{\mathbf{j}}_1, ..., \underline{\mathbf{j}}_K)$ be K-uplets of respectively row and column indices. Let $\alpha(I_K, J_K)$ defined by (3.7). Let :*

- $\underline{p}_k := \mathrm{Card}(\underline{\mathbf{i}}_k)$ *and* $\underline{q}_k := \mathrm{Card}(\underline{\mathbf{j}}_k)$, *for* $1 \le k \le K$,
- $\underline{P} := \sum_{k=1}^K \underline{p}_k$ *and* $\underline{Q} := \sum_{k=1}^K \underline{q}_k$,
- $\bar{p} := \mathrm{Card}(\cup_{k=1}^K \underline{\mathbf{i}}_k)$ *and* $\bar{q} := \mathrm{Card}(\cup_{k=1}^K \underline{\mathbf{j}}_k)$.

We have

$$T^{p,q}_N(I_K, J_K) = \alpha(I_K, J_K) + O\left(m_N^{-1} + n_N^{-1}\right).$$

3.4.2. Estimation of the conditional expectations

In this paragraph, for all $i \in [\![m_N]\!]$ and $j \in [\![n_N]\!]$, we define estimators for $\psi^{1,0}_{(\{i\},\varnothing)}h = \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)}) \mid \xi_i]$ and $\psi^{0,1}_{(\varnothing,\{j\})}h = \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)}) \mid \eta_j]$, where ξ_i and η_j have been defined in Section 3.2.

Let $\widehat{\mu}_N^{h,(i)}$ be the average of the kernel function applied on the $p \times q$ submatrices containing the row i. Symmetrically, let $\widehat{\nu}_N^{h,(j)}$ be the average of the kernel function applied on the $p \times q$ submatrices containing the column j. This means

$$\widehat{\mu}_N^{h,(i)} := \binom{m_N - 1}{p-1}^{-1} \binom{n_N}{q}^{-1} \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{S}^{p,q}_{N,(\{i\},\varnothing)}} h(Y_{\mathbf{i},\mathbf{j}}), \tag{3.8}$$

and

$$\widehat{\nu}_N^{h,(j)} := \binom{m_N}{p}^{-1} \binom{n_N - 1}{q-1}^{-1} \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{S}^{p,q}_{N,(\varnothing,\{j\})}} h(Y_{\mathbf{i},\mathbf{j}}). \tag{3.9}$$

Now, we establish some properties for these estimators.

Proposition 3.4.2. *If Y is a RCE matrix, then $\widehat{\mu}_N^{h,(i)}$ and $\widehat{\nu}_N^{h,(i)}$ are both conditionally unbiased $\psi^{1,0}_{(\{i\},\varnothing)}h$ and $\psi^{0,1}_{(\varnothing,\{j\})}h$, i.e. we have for all $N \in \mathbb{N}$:*

- $\mathbb{E}[\widehat{\mu}_N^{h,(i)} \mid \xi_i] = \psi^{1,0}_{(\{i\},\varnothing)}h$,
- $\mathbb{E}[\widehat{\nu}_N^{h,(j)} \mid \eta_j] = \psi^{0,1}_{(\varnothing,\{j\})}h$.

Proof. The first result can be found directly by the definition of $\psi^{1,0}_{(\{i\},\varnothing)}h$, since

$$\begin{aligned}
\mathbb{E}[\widehat{\mu}_N^{h,(i)} \mid \xi_i] &= \binom{m_N - 1}{p-1}^{-1} \binom{n_N}{q}^{-1} \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{S}^{p,q}_{N,(\{i\},\varnothing)}} \mathbb{E}[h(Y_{\mathbf{i},\mathbf{j}}) \mid \xi_i] \\
&= \binom{m_N - 1}{p-1}^{-1} \binom{n_N}{q}^{-1} \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{S}^{p,q}_{N,(\{i\},\varnothing)}} \psi^{1,0}_{(\{i\},\varnothing)}h \\
&= \psi^{1,0}_{(\{i\},\varnothing)}h.
\end{aligned}$$

The second result can be obtained analogously. □

Proposition 3.4.3. *If Y is a RCE matrix, then :*

- $\widehat{\mu}_N^{h,(i)} \xrightarrow[N\to\infty]{a.s.,L_1} \psi^{1,0}_{(\{i\},\varnothing)}h$,
- $\widehat{\nu}_N^{h,(j)} \xrightarrow[N\to\infty]{a.s.,L_1} \psi^{0,1}_{(\varnothing,\{j\})}h$.

As a consequence, $\widehat{\mu}_N^{h,(i)}$ and $\widehat{\nu}_N^{h,(j)}$ are consistent estimators for $\psi^{1,0}_{(\{i\},\varnothing)}h$ and $\psi^{0,1}_{(\varnothing,\{j\})}h$.

Proof. Let $N \in \mathbb{N}$ and $\mathcal{F}_N(Y) = \sigma((\widehat{\mu}_K^{h,(i)}(Y))_{K \geq N})$. Let $\Phi_N \in \mathbb{S}^{(i)}_{m_N} \times \mathbb{S}_{n_N}$ where $\mathbb{S}^{(i)}_{m_N}$ is the group of permutations σ_i of $[\![m_N]\!]$ such that $\sigma_i(i) = i$. If $\Phi_N = (\sigma_i, \tau)$, denote $\Phi_N Y = (Y_{\sigma_i(k)\tau(j)})_{\substack{1 \leq k \leq m_N \\ 1 \leq j \leq n_N}}$.

First, we observe that $\widehat{\mu}_N^{h,(i)}(Y) = \widehat{\mu}_N^{h,(i)}(\Phi_N Y)$, so $\mathcal{F}_N(\Phi_N Y) = \mathcal{F}_N(Y)$. Therefore, by the exchangeability of Y, we have

$$Y \mid \mathcal{F}_N(Y) \overset{\mathcal{D}}{=} \Phi_N Y \mid \mathcal{F}_N(Y).$$

This assertion is true for all $N \in \mathbb{N}$ and $\Phi_N \in \mathbb{S}_{m_N}^{(i)} \times \mathbb{S}_{n_N}$. Now, note that for all $(\mathbf{i},\mathbf{j})$ and $(\mathbf{i}',\mathbf{j}')$ elements of $\mathcal{S}_{N,(\{i\},\varnothing)}^{p,q}$, we can always find a permutation $\Phi_N \in \mathbb{S}_{m_N}^{(i)} \times \mathbb{S}_{n_N}$ such that $h(\Phi_N Y_{\mathbf{i},\mathbf{j}}) = h(Y_{\mathbf{i}',\mathbf{j}'})$. Thus, we have $\mathbb{E}[h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{F}_N(Y)] = \mathbb{E}[h(Y_{\mathbf{i}',\mathbf{j}'}) \mid \mathcal{F}_N(Y)]$. Hence, for any $(\mathbf{i},\mathbf{j}) \in \mathcal{S}_{N+1,(\{i\},\varnothing)}^{p,q}$, we deduce that

$$\begin{aligned}\mathbb{E}[\widehat{\mu}_N^{h,(i)}(Y) \mid \mathcal{F}_{N+1}(Y)] &= \mathbb{E}[h(Y_{(\mathbf{i},\mathbf{j})}) \mid \mathcal{F}_{N+1}(Y)] \\ &= \mathbb{E}[\widehat{\mu}_{N+1}^{h,(i)}(Y) \mid \mathcal{F}_{N+1}(Y)] \\ &= \widehat{\mu}_{N+1}^{h,(i)}(Y).\end{aligned}$$

Therefore, $\widehat{\mu}_N^{h,(i)}(Y)$ is a backward martingale with respect to $\mathcal{F}_N(Y)$ (see Appendix 3.A). By Theorem 3.A.3, we have that $\widehat{\mu}_N^{h,(i)}(Y) \xrightarrow[N\to\infty]{a.s.,L_1} \mathbb{E}[\widehat{\mu}_1^{h,(i)}(Y) \mid \mathcal{F}_\infty(Y)]$, where $\mathcal{F}_\infty(Y) = \bigcap_{N=1}^{\infty} \mathcal{F}_N(Y)$.

Finally, $\mathcal{F}_\infty(Y) = \sigma(\xi_i)$ so Proposition 3.4.2 implies that $\mathbb{E}[\widehat{\mu}_1^{h,(i)}(Y) \mid \mathcal{F}_\infty(Y)] = \psi_{(\{i\},\varnothing)}^{1,0} h(Y)$ and thus, $\widehat{\mu}_N^{h,(i)}(Y) \xrightarrow[N\to\infty]{a.s.,L_1} \psi_{(\{i\},\varnothing)}^{1,0} h(Y)$. □

3.4.3. Estimation of V^h

Finally, since we have defined estimators for $\psi_{(\{i\},\varnothing)}^{1,0} h$ and $\psi_{(\varnothing,\{j\})}^{0,1} h$, natural estimators for $v_h^{1,0} = \mathbb{V}[\psi_{(\{1\},\varnothing)}^{1,0} h]$ and $v_h^{0,1} = \mathbb{V}[\psi_{(\varnothing,\{j\})}^{0,1} h]$ can be given by :

$$\widehat{v}_N^{h;1,0} = \binom{m_N}{2}^{-1} \sum_{1 \le i_1 < i_2 \le m_N} \frac{(\widehat{\mu}_N^{h,(i_1)} - \widehat{\mu}_N^{h,(i_2)})^2}{2}$$

$$\widehat{v}_N^{h;0,1} = \binom{n_N}{2}^{-1} \sum_{1 \le j_1 < j_2 \le n_N} \frac{(\widehat{\nu}_N^{h,(j_1)} - \widehat{\nu}_N^{h,(j_2)})^2}{2}$$

Then, an estimator for V^h is

$$\widehat{V}_N^h := \frac{p^2}{\rho} \widehat{v}_N^{h;1,0} + \frac{q^2}{1-\rho} \widehat{v}_N^{h;0,1}.$$

The following theorem shows that $\widehat{V}_N^h$ is a consistent estimator for V^h.

Theorem 3.4.4. *We have* $\widehat{v}_N^{h;1,0} \xrightarrow[N\to\infty]{\mathbb{P}} v_h^{1,0}$ *and* $\widehat{v}_N^{h;0,1} \xrightarrow[N\to\infty]{\mathbb{P}} v_h^{0,1}$. *As a consequence,* $\widehat{V}_N^h \xrightarrow[N\to\infty]{\mathbb{P}} V^h$.

Before proving this theorem, we first need to highlight some properties of $\widehat{v}_N^{h;1,0}$ and $\widehat{v}_N^{h;0,1}$, which are proven in Appendix 3.D.

Proposition 3.4.5. *We have* $\mathbb{E}[\widehat{v}_N^{h;1,0}] = v_h^{1,0} + O\left(N^{-1}\right)$ *and* $\mathbb{E}[\widehat{v}_N^{h;0,1}] = v_h^{0,1} + O\left(N^{-1}\right)$. *As a consequence,* $\widehat{v}_N^{h;1,0}$ *and* $\widehat{v}_N^{h;0,1}$ *are asymptotically unbiased estimators for* $v_h^{1,0}$ *and* $v_h^{0,1}$.

Proposition 3.4.6. *We have* $\mathbb{V}[\widehat{v}_N^{h;1,0}] = O\left(N^{-1}\right)$ *and* $\mathbb{V}[\widehat{v}_N^{h;0,1}] = O\left(N^{-1}\right)$.

Proof of Theorem 3.4.4. For some $\epsilon > 0$, it follows from Proposition 3.4.5 that for large enough values of N, $\left|\mathbb{E}[\widehat{v}_N^{h;1,0}] - v_h^{1,0}\right| < \epsilon$. The triangular inequality and Chebyshev's inequality states that

$$\mathbb{P}\left(\left|\widehat{v}_N^{h;1,0} - v_h^{1,0}\right| > \epsilon\right) \leq \mathbb{P}\left(\left|\widehat{v}_N^{h;1,0} - \mathbb{E}[\widehat{v}_N^{h;1,0}]\right| \geq \epsilon - \left|\mathbb{E}[\widehat{v}_N^{h;1,0}] - v_h^{1,0}\right|\right) \leq \frac{\mathbb{V}[\widehat{v}_N^{h;1,0}]}{\left(\epsilon - \left|\mathbb{E}[\widehat{v}_N^{h;1,0}] - v_h^{1,0}\right|\right)^2}.$$

Applying Propositions 3.4.5 and 3.4.6 to the right-hand side of the inequality ensures that $\mathbb{P}\left(\left|\widehat{v}_N^{h;1,0} - v_h^{1,0}\right| > \epsilon\right) \xrightarrow[N\to\infty]{} 0$ which concludes the proof. □

With Theorem 3.4.4, it is possible to use $\widehat{V}_N^h$ for statistical inference tasks when plugged-in in place of V^h, a asymptotic normality result similar to Theorem 3.3.1 holds.

Corollary 3.4.7. *If* $V^h > 0$, *then*

$$\sqrt{\frac{N}{\widehat{V}_N^h}}\left(U_N^h - U_\infty^h\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

Proof. First, Theorem 3.3.1 ensures that $\sqrt{\frac{N}{\widehat{V}_N^h}}(U_N^h - U_\infty^h) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$. Second, it can be derived from Theorem 3.4.4 that $\sqrt{\frac{V^h}{\widehat{V}_N^h}} \xrightarrow[N\to\infty]{\mathbb{P}} 1$.

Then, applying Slutsky's theorem, we get $\sqrt{\frac{N}{\widehat{V}_N^h}}(U_N^h - U_\infty^h) = \sqrt{\frac{V^h}{\widehat{V}_N^h}} \times \sqrt{\frac{N}{V^h}}(U_N^h - U_\infty^h) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$. □

3.4.4. Calculation of the estimator

In practice, the computation of the estimators $(\widehat{\mu}_N^{h,(i)})_{1\leq i\leq m_N}$ and $(\widehat{\nu}_N^{h,(j)})_{1\leq j\leq n_N}$ using their definition is pretty tedious, as they are sums of $O(N^{p+q-1})$ terms. The computation cost of $\widehat{V}_N^h$ is then $O(N^{p+q})$ which is of same order as the computation of U_N^h when naively applying the kernel function to all the $p \times q$ submatrices and averaging.

However, one would actually try to avoid to compute U_N^h in that way if possible, as for simple kernels, it is possible to write U_N^h in the form of operation on matrices, which are more optimized, say $O(N^{a+b})$ with $a \leq p$ and $b \leq q$.

In these cases, it might be time-efficient to compute $\widehat{V}_N^h = \frac{p^2}{\rho}\widehat{v}_N^{h;1,0} + \frac{q^2}{1-\rho}\widehat{v}_N^{h;0,1}$ with the following alternative form.

Proposition 3.4.8. *An alternative form for $\widehat{v}_N^{h;1,0}$ and $\widehat{v}_N^{h;0,1}$ is given by*

$$\widehat{v}_N^{h;1,0} = \frac{(m_N - p)^2}{p^2(m_N - 1)} \sum_{i=1}^{m_N} \left(U_N^h - U_N^{h,(-i,\varnothing)}\right)^2$$

and

$$\widehat{v}_N^{h;0,1} = \frac{(n_N - q)^2}{q^2(n_N - 1)} \sum_{j=1}^{n_N} \left(U_N^h - U_N^{h,(\varnothing,-j)}\right)^2,$$

where

$$U_N^{h,(-i,\varnothing)} := \left[\binom{m_N - 1}{p}\binom{n_N}{q}\right]^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_p(\llbracket m_N \rrbracket \setminus \{i\}) \\ \mathbf{j} \in \mathcal{P}_q(\llbracket n_N \rrbracket)}} h(Y_{\mathbf{i},\mathbf{j}})$$

and

$$U_N^{h,(\varnothing,-j)} := \left[\binom{m_N}{p}\binom{n_N - 1}{q}\right]^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_p(\llbracket m_N \rrbracket) \\ \mathbf{j} \in \mathcal{P}_q(\llbracket n_N \rrbracket \setminus \{j\})}} h(Y_{\mathbf{i},\mathbf{j}}).$$

Proof. The relations are obtained from the definition of $\widehat{v}_N^{h;1,0}$ and $\widehat{v}_N^{h;0,1}$ and noticing that

$$\binom{m_N - 1}{p - 1}\binom{n_N}{q}\widehat{\mu}_N^{h,(i)} = \binom{m_N}{p}\binom{n_N}{q}U_N^h - \binom{m_N - 1}{p}\binom{n_N}{q}U_N^{h,(-i,\varnothing)}$$

and

$$\binom{m_N}{p}\binom{n_N - 1}{q - 1}\widehat{\nu}_N^{h,(j)} = \binom{m_N}{p}\binom{n_N}{q}U_N^h - \binom{m_N}{p}\binom{n_N - 1}{q}U_N^{h,(\varnothing,-j)}.$$

□

Remark. This is one alternative method to compute $\widehat{V}_N^h$, but not necessarily the optimal one. With this form, the computational cost of $\widehat{V}_N^h$ is $O(N^{a+b+1})$. This can be outperformed by the naive method in some specific cases, e.g. when $a = p$ and $b = q$. This method is particularly fit when $a + b + 1 < p + q$.

Remark. The alternative form of $\widehat{v}_N^{h;1,0}$ and $\widehat{v}_N^{h;0,1}$ is reminiscent of the jackknife estimator for the variance of U-statistics of one-dimensional arrays (Arvesen, 1969), but the two are well distinct. In the case where Y is a one-dimensional array, the U-statistic associated to the kernel $h : \mathbb{R}^p \to \mathbb{R}$ is

$$U_N^h = \binom{N}{p}^{-1} \sum_{\mathbf{i} \in \mathcal{P}_p(\llbracket N \rrbracket)} h(X_{\mathbf{i}}).$$

The jackknife estimator of the asymptotic variance of this U-statistic is

$$\widehat{V}_N^{h,J} = (N-1)\sum_{i=1}^{N}(U_N^h - U_N^{h,(-i)})^2$$

where

$$U_N^{h,(-i)} = \binom{N}{p}^{-1} \sum_{\mathbf{i}\in\mathcal{P}_p([\![N]\!]\setminus\{i\})} h(X_{\mathbf{i}}).$$

In fact, our estimator is closer to Sen's estimator of the asymptotic variance (Sen, 1960, 1977), which depends on the kernel size and is related to the jackknife estimator $\widehat{V}_N^{h,S} = (N-k)^2/(N-1)^2\widehat{V}_N^{h,J}$. However, it is unclear how these estimators could be translated in a two-dimensional setup where Y is a matrix instead of a vector, especially how to define the analog of $U_N^{h,(-i)}$.

3.4.5. Extension to functions of U-statistics

In the case of a function of U-statistics $g(U_N^{h_1},...,U_N^{h_D})$, Corollary 3.3.6 applies and the asymptotic variance to be estimated is $V^\delta = \nabla g(\theta)^T \Sigma^{h_1,...,h_D} \nabla g(\theta) \neq 0$, where $\theta = (U_\infty^{h_1},...,U_\infty^{h_D})$. Similar to the estimator for the asymptotic variance of V^h, we suggest an estimator for the covariance matrix $\Sigma^{h_1,...,h_D} = \left(C^{h_k,h_\ell}\right)_{1\leq k,\ell\leq D}$.

For each kernel h_k, $1 \leq k \leq D$, let $\widehat{\mu}_N^{h_k,(i)}$ and $\widehat{\nu}_N^{h_k,(j)}$ be the respective estimators of the conditional expectations $\psi^{1,0}_{(\{i\},\varnothing)}h_k$ and $\psi^{0,1}_{(\varnothing,\{j\})}h_k$, as defined in equations (3.8) and (3.9).

Now define

$$\widehat{c}_N^{h_k,h_\ell;1,0} := \binom{m_N}{2}^{-1} \sum_{1\leq i_1<i_2\leq m_N} \frac{(\widehat{\mu}_N^{h_k,(i_1)} - \widehat{\mu}_N^{h_k,(i_2)})(\widehat{\mu}_N^{h_\ell,(i_1)} - \widehat{\mu}_N^{h_\ell,(i_2)})}{2}$$

and

$$\widehat{c}_N^{h_k,h_\ell;0,1} := \binom{n_N}{2}^{-1} \sum_{1\leq j_1<j_2\leq n_N} \frac{(\widehat{\nu}_N^{h_k,(j_1)} - \widehat{\nu}_N^{h_k,(j_2)})(\widehat{\nu}_N^{h_\ell,(j_1)} - \widehat{\nu}_N^{h_\ell,(j_2)})}{2}.$$

Then, for two kernels h_k and h_ℓ,

$$\widehat{C}_N^{h_k,h_\ell} := \frac{p^2}{\rho}\widehat{c}_N^{h_k,h_\ell;1,0} + \frac{q^2}{1-\rho}\widehat{c}_N^{h_k,h_\ell;0,1},$$

is an estimator of the asymptotic covariance term C^{h_k,h_ℓ}.

With a similar proof as for Theorem 3.4.4, the following theorem ensures the consistency of this estimator.

Theorem 3.4.9. *For two kernel functions h_k and h_ℓ, we have $\widehat{c}_N^{h_k,h_\ell;1,0} \xrightarrow[N\to\infty]{\mathbb{P}} c_{h_k,h_\ell}^{1,0}$ and $\widehat{c}_N^{h_k,h_\ell;0,1} \xrightarrow[N\to\infty]{\mathbb{P}} c_{h_k,h_\ell}^{0,1}$. As a consequence, $\widehat{C}_N^{h_k,h_\ell} \xrightarrow[N\to\infty]{\mathbb{P}} C^{h_k,h_\ell}$.*

Thus, for linearly independent kernel functions $(h_1, ..., h_D)$, the entries of the matrix $\widehat{\Sigma}_N^{h_1,...,h_D} := \left(\widehat{C}_N^{h_k,h_\ell}\right)_{1\le k,\ell\le D}$ converge to the entries of $\Sigma^{h_1,...,h_D}$. Set $\widehat{V}_N^\delta := \nabla g(U_N^{h_1}, ..., U_N^{h_D})^T \widehat{\Sigma}_N^{h_1,...,h_D} \nabla g(U_N^{h_1}, ..., U_N^{h_D})$, then we have the following straightforward result.

Corollary 3.4.10. *We have* $\widehat{V}_N^\delta \xrightarrow[N\to\infty]{\mathbb{P}} V^\delta$.

Finally, the plug-in corollary also stands.

Corollary 3.4.11. *If* $V^\delta > 0$, *then*

$$\sqrt{\frac{N}{\widehat{V}_N^\delta}}\left(g(U_N^{h_1}, ..., U_N^{h_D}) - g(\theta)\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

3.5. RCE models, kernel functions and network comparison

3.5.1. Examples of RCE models

Bipartite Expected Degree Distribution model The Bipartite Expected Degree Distribution (BEDD) model, suggested by Ouadah et al. (2022), is a binary graph model characterised by two distributions from which the row and column nodes draw a weight. The probability of a connection between two nodes is fully determined by the corresponding row and column weight distributions. The distribution of a graph following a BEDD model can be written using latent variables $(\xi_i)_{i\ge1}$ and $(\eta_j)_{j\ge1}$ corresponding to the row and column nodes of the graph :

$$\begin{aligned} \xi_i, \eta_j &\overset{iid}{\sim} \mathcal{U}[0,1] \\ Y_{ij} \;\mid\; \xi_i, \eta_j &\sim \mathcal{B}(\lambda f(\xi_i) g(\eta_j)). \end{aligned} \tag{3.10}$$

where

- f and g are positive, càdlàg, nondecreasing, bounded and normalized ($\int f = \int g = 1$) real functions $[0,1] \to \mathbb{R}_+$,
- λ is a positive real number such that $\lambda \le \|f\|_\infty^{-1} \|g\|_\infty^{-1}$,
- $\mathcal{B}$ is the Bernoulli distribution.

The BEDD model is a RCE version of the Expected Degree Sequence model of Chung and Lu (2002) but in the BEDD, the row weights $f(\xi_i)$ and column weights $g(\eta_j)$ are exchangeable. Indeed, f and g characterize the weight distributions of the row and column nodes whereas the weights are fixed in the Expected Degree Sequence model. The triplet (λ, f, g) is called the BEDD parameters.

Latent Block model The Latent Block model (LBM) (Govaert and Nadif, 2003) is a binary graph model characterised by a partition of row and column nodes in several groups. It can be considered as a bipartite extension of the Stochastic Block model (Nowicki and Snijders, 2001). The probability of interaction between two nodes is fully determined by the groups to which they belong. All the nodes have the same probability to belong to each group. The distribution of a LBM is most commonly written using independent latent variables for the node attribution in a group $(Z_i)_{i\geq 1}$ and $(W_j)_{j\geq 1}$ corresponding to the row and column nodes of the graph :

$$\begin{aligned} Z_i &\overset{iid}{\sim} \mathcal{M}(1;\boldsymbol{\alpha}) \\ W_j &\overset{iid}{\sim} \mathcal{M}(1;\boldsymbol{\beta}) \\ Y_{ij} \mid Z_i = k, W_j = \ell &\sim \mathcal{B}(\pi_{k\ell}), \end{aligned} \tag{3.11}$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_K)$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_L)$ are the probability vectors of the rows and the columns and $\boldsymbol{\pi} = (\pi_{k\ell})_{1\leq k\leq K, 1\leq \ell\leq L} \in [0,1]^{KL}$ is a matrix of probabilities.

The LBM is a RCE model, since the group attribution variables of the nodes are exchangeable.

W-graph model Let w be a function of $[0,1]^2 \to [0,1]$. The W-graph model associated to w is defined by

$$\begin{aligned} \xi_i, \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1] \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{B}\left(w(\xi_i, \eta_j)\right), \end{aligned} \tag{3.12}$$

w is sometimes referred to as a graphon. For identification reasons, we assume $\int w(\cdot, \eta) d\eta$ and $\int w(\xi, \cdot) d\xi$ to be càdlàg, nondecreasing and bounded.

Any RCE model can be written as a W-graph model (Diaconis and Janson, 2008) so can the BEDD model and the LBM be expressed with a graphon. For the BEDD model, $w(\xi_i, \eta_j) = \lambda f(\xi_i) g(\eta_j)$ where f and g are those of Equation (3.10). For the LBM, $w(\xi_i, \eta_j) = \sum_{k,\ell} \pi_{k\ell} \mathbb{1}\{s(\xi_i) = k\} \mathbb{1}\{t(\eta_j) = \ell\}$ where $s(\xi_i) = 1 + \sum_{k=1}^{K} \mathbb{1}\{\xi_i > \sum_{k'=1}^{k} \alpha_{k'}\}$, $t(\eta_j) = 1 + \sum_{\ell=1}^{L} \mathbb{1}\{\eta_j > \sum_{\ell'=1}^{\ell} \beta_{\ell'}\}$ and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\pi}$ are those of Equation (3.11).

Extension to weighted graphs All the models presented above are defined for binary graphs. However, one can extend it to weighted graphs by switching the Bernoulli distributions with another. In particular, the Poisson distribution is particularly fitted for count values in $\mathbb{N}$:

$$\begin{aligned} \xi_i, \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1] \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{P}\left(w(\xi_i, \eta_j)\right), \end{aligned} \tag{3.13}$$

where $w : [0,1]^2 \to \mathbb{R}_+$ is the graphon of this model. Similarly, one can define a Poisson-BEDD model or Poisson-LBM.

3.5.2. Examples of kernel functions

We present three examples of statistics which are U-statistics or functions of U-statistics. Sometimes, a kernel function h has a long expression, especially when it is the symmetric version of some simpler function h^0, as in (3.3). In this case, we introduce the kernels of interest with h^0 instead of h, but the U-statistics and asymptotic results always apply to the symmetric version h.

Motif counts Motifs are the name given to small subgraphs. Their occurrences in the complete network can be counted. Motif counts are useful statistics for random graphs as they provide information on the network local structure (Shen-Orr et al., 2002; Kashtan et al., 2004). Many random network models hinge on motif frequencies. In the Exponential Random Graph Model (Frank and Strauss, 1986), motif frequencies are sufficient statistics. The *dk*-random graph model (Orsini et al., 2015) also largely relies on motif frequencies.

Their asymptotic properties are widely studied and a large numbers of studies use motif counts to perform statistical tests (Reinert and Röllin, 2010; Bickel et al., 2011; Bhattacharyya and Bickel, 2015; Coulson et al., 2016; Gao and Lafferty, 2017; Maugis et al., 2020; Naulet et al., 2021; Ouadah et al., 2022). Our framework is particularly well adapted to the use of motif counts for statistical tests as frequencies are in fact U-statistics with kernel functions of the same size as the motifs.

For many applications, motifs are considered as elementary building blocks which can be specifically interpreted. This is the case in molecular biology (Shen-Orr et al., 2002; Pržulj et al., 2004; Ali et al., 2014), neurology (Zhao et al., 2011), sociology (Bearman et al., 2004), evolution (Przytycka, 2006) and ecology (Stouffer et al., 2007; Baker et al., 2015). In particular, Figure 3 of Simmons et al. (2019) lists all the bipartite motifs consisting of from 2 to 6 nodes. For example, their motif 6 represents a 2×2 clique, where every node is connected to all others (Figure 3.1). Their motif 14 represents a path between two column nodes, passing through another column node and two row nodes (Figure 3.1). The latter indicates an indirect interaction between the row nodes, hinging on the middle row node. Lanuza et al. (2023) found that motif 6 is over-represented in plant-pollinator interaction networks, while motif 14 is under-represented, compared to Erdös-Rényi graphs of the same density.

These motifs can be counted with $U_N^{h_6}$ and $U_N^{h_{14}}$ the U-statistics associated to the kernels h_6 and h_{14}, that we introduce as the symmetric version of the following functions :

$$\begin{aligned} h_6^0(Y_{(i_1,i_2;j_1,j_2)}) &= Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_1}Y_{i_2j_2}, \\ h_{14}^0(Y_{(i_1,i_2;j_1,j_2,j_3)}) &= Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_2}Y_{i_2j_3}(1-Y_{i_2j_1})(1-Y_{i_1j_3}). \end{aligned} \tag{3.14}$$

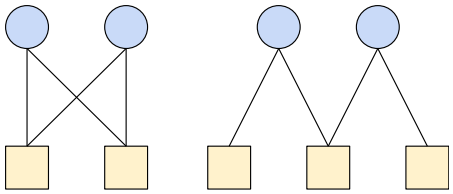

Figure 3.1 – Motifs 6 (left) and 14 (right), counted by $U_N^{h_6}$ and $U_N^{h_{14}}$

Remark. With this example, it is very apparent that introducing simple, non-symmetric kernel functions and then symmetrizing them is way simpler than introducing symmetric kernel functions. h_6^0 is already symmetric, so $h_6 = h_6^0$. However, h_{14}^0 is not symmetric and one can sum over all the permutations of the indices to make it symmetric using formula (3.3). Because of the automorphisms of motif 14, it involves only 6 permutations instead of 2!3! = 12.

$$\begin{aligned}
&h_{14}(Y_{(i_1,i_2;j_1,j_2,j_3)}) \\
&\quad = \frac{1}{6}Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_2}Y_{i_2j_3}(1-Y_{i_2j_1})(1-Y_{i_1j_3}) + \frac{1}{6}Y_{i_1j_2}Y_{i_1j_3}Y_{i_2j_3}Y_{i_2j_1}(1-Y_{i_1j_1})(1-Y_{i_2j_2}) \\
&\quad + \frac{1}{6}Y_{i_1j_3}Y_{i_1j_1}Y_{i_2j_1}Y_{i_2j_2}(1-Y_{i_1j_2})(1-Y_{i_2j_3}) + \frac{1}{6}Y_{i_2j_1}Y_{i_2j_2}Y_{i_1j_2}Y_{i_1j_3}(1-Y_{i_1j_1})(1-Y_{i_2j_3}) \\
&\quad + \frac{1}{6}Y_{i_2j_2}Y_{i_2j_3}Y_{i_1j_3}Y_{i_1j_1}(1-Y_{i_2j_1})(1-Y_{i_1j_2}) + \frac{1}{6}Y_{i_2j_3}Y_{i_2j_1}Y_{i_1j_1}Y_{i_1j_2}(1-Y_{i_2j_2})(1-Y_{i_1j_3}).
\end{aligned}$$

Using Corollary 3.4.7, the following studentized statistics converge to a standard normal distribution

$$Z_N^6 := \sqrt{\frac{N}{\widehat{V}_N^{h_6}}}\left(U_N^{h_6} - U_\infty^{h_6}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1),$$

$$Z_N^{14} := \sqrt{\frac{N}{\widehat{V}_N^{h_{14}}}}\left(U_N^{h_{14}} - U_\infty^{h_{14}}\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

Product graphon The W-graph model encompasses all the dissociated RCE models. One can make use of this model to compare the form of the graphon that has generated a graph to a known form. One interesting form is the product form, which corresponds to an absence of specific interaction between row nodes and column nodes. For example, one can normalize the graphon in the Poisson W-graph model described by (3.13) such that

$$Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{P}(\lambda\bar{w}(\xi_i, \eta_j)),$$

where $\lambda > 0$ and $\iint \bar{w} = 1$. Call $\bar{w}$ a normalized graphon. Define f and g as the marginals of $\bar{w}$, i.e.

$$f = \int \bar{w}(\cdot,\eta)d\eta \qquad\qquad g = \int \bar{w}(\xi,\cdot)d\xi. \tag{3.15}$$

A density-free dissimilarity measure between a graphon and its corresponding product form could be the quantity $d(\bar{w})$ defined as

$$d(\bar{w}) := \|\bar{w} - fg\|_2^2 = \iint (\bar{w}(\xi,\eta) - f(\xi)g(\eta))^2 \, d\xi d\eta. \tag{3.16}$$

Observe that $d(\bar{w}) = 0$ if and only if $\bar{w}$ is of product form, e.g. $\bar{w}(\xi,\eta) = f(\xi)g(\eta)$ almost everywhere in $[0,1]^2$. In this case, the W-graph model (3.5.2) is a BEDD model and f and g have the same role as in (3.10).

We suggest an estimator for $d(\bar{w})$. Let h_A, h_B, h_C and h_D be kernel functions of respective size 2×2, 1×2, 2×1 and 1×1 defined as in Table 3.1. These kernel functions are linearly independent, so Corollary 3.3.5 applies to the associated U-statistics $(U_N^{h_A}, U_N^{h_B}, U_N^{h_C}, U_N^{h_D})$, and they are jointly asymptotically normal with asymptotic covariance matrix Σ^{h_A,h_B,h_C,h_D}.

h	$h^0(Y_{(i_1,\dots,i_r;j_1,\dots,j_c)})$	$\mathbb{E}[h(Y_{(i_1,\dots,i_r;j_1,\dots,j_c)})]$
h_A	$h^0_{A,1}(Y_{(i_1,i_2;j_1,j_2)}) - 2h^0_{A,2}(Y_{(i_1,i_2;j_1,j_2)})$	$\lambda^3 \iint \bar{w}(\xi,\eta)(\bar{w}(\xi,\eta) - 2f(\xi)g(\eta))d\xi d\eta$
$h_{A,1}$	$Y_{i_1j_1}(Y_{i_1j_1} - 1)Y_{i_2j_2}$	$\lambda^3 \iint \bar{w}(\xi,\eta)^2 d\xi d\eta$
$h_{A,2}$	$Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_2}$	$\lambda^3 \iint \bar{w}(\xi,\eta)f(\xi)g(\eta)d\xi d\eta$
h_B	$Y_{i_1j_1}Y_{i_1j_2}$	$\lambda^2 \int f(\xi)^2 d\xi$
h_C	$Y_{i_1j_1}Y_{i_2j_1}$	$\lambda^2 \int g(\eta)^2 d\eta$
h_D	$Y_{i_1j_1}$	λ

Table 3.1 – Kernel functions used for the estimation of $d(\bar{w})$. The expectations are given by Lemma 3.E.1 in Appendix.

We define the estimator of $d(\bar{w})$ as $\widehat{d}_N = t(U_N^{h_A}, U_N^{h_B}, U_N^{h_C}, U_N^{h_D}) := U_N^{h_A}/(U_N^{h_D})^3 + U_N^{h_B}U_N^{h_C}/(U_N^{h_D})^4$. We have $t(U_\infty^{h_A}, U_\infty^{h_B}, U_\infty^{h_C}, U_\infty^{h_D}) = \iint \bar{w}(\xi,\eta)(\bar{w}(\xi,\eta) - 2f(\xi)g(\eta))d\xi d\eta + \int f(\xi)^2 d\xi \times \int g(\eta)^2 d\eta = d(\bar{w})$. Then, Corollary 3.4.11 ensures that the studentized statistic converge to a standard normal distribution

$$Z_N^d := \sqrt{\frac{N}{\widehat{V}_N^d}} \left(\widehat{d}_N - d(\bar{w})\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

where $\widehat{V}_N^d = \nabla t(U_N^{h_A}, U_N^{h_B}, U_N^{h_C}, U_N^{h_D})^T \widehat{\Sigma}_N^{h_A,h_B,h_C,h_D} \nabla t(U_N^{h_A}, U_N^{h_B}, U_N^{h_C}, U_N^{h_D})$ and $\widehat{\Sigma}_N^{h_A,h_B,h_C,h_D}$ is defined as in Corollary 3.4.11.

Heterogeneity of the rows of a network The degree distributions of a network hold significant information about its topology. For a binary network, denote $D_i := \sum_j Y_{ij}$ the degree of the i-th row, then $\mathbb{E}[D_i \mid \xi_i] = f(\xi_i)\mathbb{E}[D_i]$, where f is given by (3.15). Therefore, the marginals f and g accounts for the expected relative degree distributions of the binary network, which is known to characterize networks. In the BEDD model, these distributions (and the density of the network) fully characterize the model.

Although the sum of the edge weights and the number of edges stemming from a nodes are equivalent for binary networks, for weighted networks, they are two different quantities. The sum of the edge weights stemming from a node is sometimes called its strength (Barrat et al., 2004), which do not depend entirely on the number of edges. In a network with weighted edges, $f(\xi_i)$ corresponds to expected relative strength of the i-th row node, instead of its expected relative degree.

One way to characterize the degree/strength distributions of a network is to calculate their variance. In particular, the variance of the degree distribution of a network quantifies its heterogeneity (i.e. the unbalance between strongly interacting nodes and the others) and can be used as an index to characterize networks Snijders (1981). In our framework with random graph models, rather than directly study the empirical distribution of row degrees/strengths, one would like to retrieve information on f and g, the distributions of the expected degrees/strengths specified by the model. For the BEDD model, f and g are directly given by the model. The variance of the row expected relative degree/strength distribution is $F_2 - 1$, where $F_2 := \int f^2(\xi) d\xi$.

F_2 can be estimated using the estimator $\widehat{F}_{2,N} := \kappa(U_N^{h_1}, U_N^{h_2}) = U_N^{h_1}/U_N^{h_2}$, using the U-statistics associated to the submatrix kernels functions h_1 and h_2 defined as

$$h_1(Y_{(i;j_1,j_2)}) = Y_{ij_1}Y_{ij_2}$$

and

$$h_2(Y_{(i_1,i_2;j_1,j_2)}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_2j_2} + Y_{i_1j_2}Y_{i_2j_1}).$$

By Lemma 3.E.2 in Appendix, we have $U_\infty^{h_1} = \lambda^2 F_2$ and $U_\infty^{h_2} = \lambda^2$. Given that $\nabla\kappa(U_N^{h_1}, U_N^{h_2}) = (1/U_N^{h_2}, -U_N^{h_1}/(U_N^{h_2})^2)$, the application of Corollary 3.4.11 yields

$$Z_N^{F_2} := \sqrt{\frac{N}{\widehat{V}_N^{F_2}}}\left(\widehat{F}_{2,N} - F_2\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1), \tag{3.17}$$

where

$$\widehat{V}_N^{F_2} = \frac{1}{(U_N^{h_2})^2}\widehat{V}_N^{h_1} - \frac{2\widehat{F}_{2,N}}{(U_N^{h_2})^2}\widehat{C}_N^{h_1,h_2} + \frac{(\widehat{F}_{2,N})^2}{(U_N^{h_2})^2}\widehat{V}_N^{h_2}.$$

In a similar way, one can define the analogous estimator $\widehat{G}_{2,N}$ for $G_2 := \int g^2(\eta)d\eta$ and its associated variance estimator to quantify the heterogeneity of the columns of networks.

3.5.3. Network comparison

Network comparison has a long history in network analysis literature (Emmert-Streib et al., 2016; Tantardini et al., 2019). Few comparison methods use random network models (Asta and Shalizi, 2014; Maugis et al., 2020; Le Minh, 2023). However, there are several advantages to using model-based approaches. Indeed, random models define a probability distribution of networks which can be used to derive statistical guarantees. Also, models can be used to control the sources of heterogeneity in networks. The comparison can be made with respect to a quantity of interest, which makes it easier to interpret. Now, we show how to define a test statistic for model-based network comparison with our framework. Naturally, U-statistics define network statistics characterizing a network. They can be used to analyze a single network, but their use is easily extended to compare different networks.

Let Y^A and Y^B be two independent networks of respective sizes $m_N^A \times n_N^A$ and $m_N^B \times n_N^B$. Define a network quantity of interest θ. These two networks are generated by two models, leading to different values θ^A and θ^B of this quantity for the two models. Let $\widehat{\theta}_N$ be an estimator for this quantity of interest. If $\widehat{\theta}_N$ is a U-statistic or a function of U-statistics, then as previously, we have both

$$\sqrt{N}(\widehat{\theta}_N(Y^A) - \theta^A) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V^A)$$

and

$$\sqrt{N}(\widehat{\theta}_N(Y^B) - \theta^B) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V^B),$$

where V^A and V^B are the asymptotic variances.

To compare θ^A and θ^B, we can confront the test hypotheses $\mathcal{H}_0 : \theta^A = \theta^B$ and $\mathcal{H}_1 : \theta^A \neq \theta^B$. Because Y^A and Y^B are independent, the previous convergence results give

$$\sqrt{\frac{N}{V^A + V^B}}\left(\widehat{\delta}_N(Y^A, Y^B) - (\theta^A - \theta^B)\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1),$$

where $\widehat{\delta}_N(Y^A, Y^B) := \widehat{\theta}_N(Y^A) - \widehat{\theta}_N(Y^B)$. Therefore, using the consistent estimators $\widehat{V}_N^A$ and $\widehat{V}_N^B$ of the asymptotic variances, the test statistic

$$Z_N(Y^A, Y^B) := \sqrt{\frac{N}{\widehat{V}_N^A + \widehat{V}_N^B}}\widehat{\delta}_N(Y^A, Y^B)$$

admits the convergence result

$$Z_N(Y^A, Y^B) - \sqrt{\frac{N}{\widehat{V}_N^A + \widehat{V}_N^B}}(\theta^A - \theta^B) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

Under $\mathcal{H}_0$, we have $\theta^A - \theta^B = 0$, so $Z_N(Y^A, Y^B)$ converges in distribution to a standard Gaussian variable.

3.6. Simulations

In this section, we illustrate our theoretical results, in particular Corollaries 3.4.7 and 3.4.11, using the previous examples of graph models and kernel functions. For each example of kernel function, we check the asymptotic normality of the studentized statistic for networks simulated under different configurations (different models, values of N and ρ). We give the resulting Q-Q plots. We also examine the coverage probabilities of the confidence intervals built with the estimates or the standard deviations of the statistics, depending on whether they are fitted for estimation or statistical testing.

3.6.1. Motif counts

Model I This graph model is a LBM with 2 row groups and 2 columns groups of equal proportion. Using the notations of (3.11), this means $\boldsymbol{\alpha} = \boldsymbol{\beta} = (0.5, 0.5)$. The probability matrix $\boldsymbol{\pi}$ has size 2×2. We set $\pi_{k\ell} = 0.5$ for all $1 \le k, \ell \le 2$ except $\pi_{11} = 0.95$.

Under Model I, for each value $N \in \{2^{k/2} : 10 \le k \le 22\}$ and $\rho \in \{1/8, 1/2\}$, we have simulated $K = 500$ networks of size $m_N \times n_N$ where $m_N = \lfloor \rho N \rfloor$ and $n_N = N - m_N$. The relative frequencies of motifs 6 and 10 of Simmons et al. (2019) are respectively given by $U_N^{h_6}$ and $U_N^{h_{14}}$ where h_6 and h_{14} were defined by (3.14).

The Q-Q plots of the studentized statistic associated to h_6 and h_{14} are given in Figures 3.2 and 3.3. For $\rho = 1/2$, we observe that the empirical distribution of both statistics converge and become close to a normal distribution as $N \gtrsim 128$. Figure 3.4 gives the respective coverage probabilities for $U_\infty^{h_6}$ and $U_\infty^{h_14}$ of the intervals $\left[U_N^{h_6} - \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{h_6}}}, U_N^{h_6} + \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{h_6}}}\right]$ and $\left[U_N^{h_{14}} - \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{h_{14}}}}, U_N^{h_{14}} + \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{h_{14}}}}\right]$ respectively, where Φ is the quantile function of the standard normal distribution. The coverage probabilities converge to α but with different speeds depending on the motif. We also observe that a larger number of nodes is needed to reach the target coverage probabilities with $\rho = 1/8$ (rectangular matrix) than $\rho = 1/2$ (square matrix).

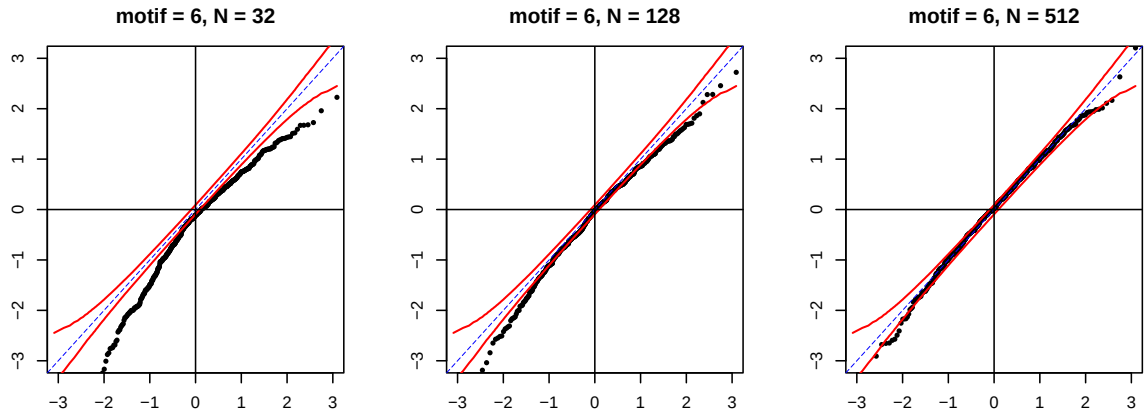

Figure 3.2 – Q-Q plots for Z_N^6 the studentized statistic associated with $U_N^{h_6}$, with $\rho = 0.5$.

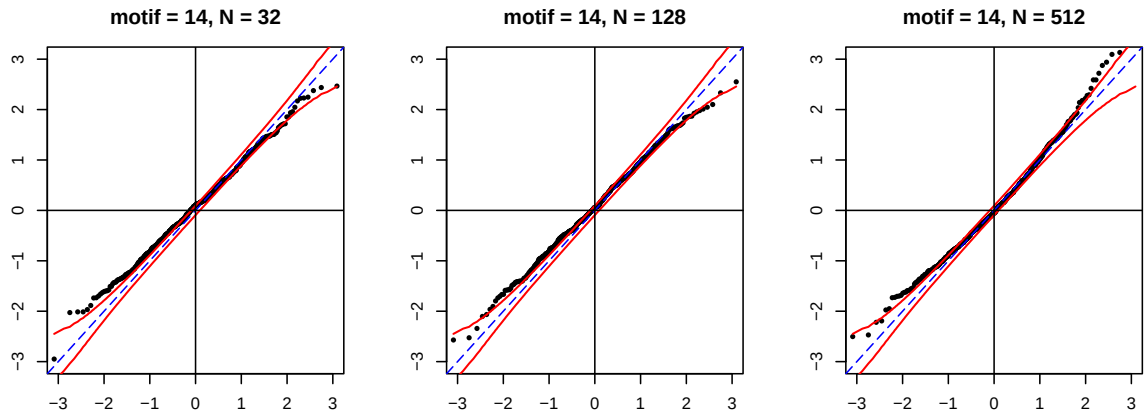

Figure 3.3 – Q-Q plots for Z_N^{14} the studentized statistic associated with $U_N^{h_{14}}$, with $\rho = 0.5$.

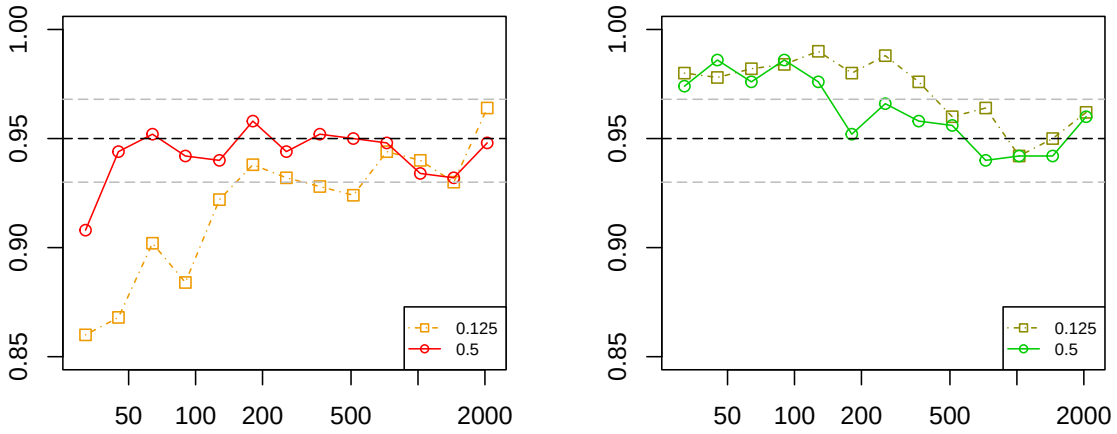

Figure 3.4 – Empirical coverage probabilities for the asymptotic confidence intervals at level $\alpha = 0.95$ of $U_N^{h_6}$ (left) and $U_N^{h_{14}}$ (right) for different values of N (x-axis), $\rho \in \{1/8, 1/2\}$. Grey dashed lines represent the confidence interval at level 0.95 of the frequency $Z = X/K$, if X follows the binomial distribution with parameters K and $\alpha = 0.95$.

3.6.2. Graphon product distance

Model II(ϵ) We consider a Poisson-LBM. As this is a weighted graph model, $\boldsymbol{\pi}$ is a matrix of weights rather than probabilities, i.e. $(\pi_{k\ell})_{1\leq k\leq K, 1\leq \ell\leq L}$ are real non-negative numbers. We consider 2 row groups and 2 column groups of equal proportion. Let $\boldsymbol{\pi}^0$ be a weight matrix. We set $\pi^0_{11} = 4$, $\pi^0_{12} = \pi^0_{21} = 2$ and $\pi^0_{22} = 1$. The LBM with parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\pi}$ is also a BEDD model. Indeed, its graphon has a product form and can be written $w_0(\xi,\eta) = \lambda\bar{w}_0(\xi,\eta) = \lambda f_0(\xi)g_0(\eta)$ where $\lambda = 9/4$, $f_0 = g_0$ both take values $4/3$ on $[0, 0.5]$, $2/3$ on $]0.5, 1]$.

Now, let $\boldsymbol{\tau}$ be the 2×2 matrix where $\tau_{11} = \tau_{22} = 2$ and $\tau_{12} = \tau_{21} = 0$. For $\epsilon \geq 0$, we define Model II(ϵ) as a LBM with group probabilities $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and with weight matrix $\boldsymbol{\pi}^\epsilon = \frac{\lambda}{\lambda+\epsilon}(\boldsymbol{\pi}^0 + \epsilon\boldsymbol{\tau})$. Thus, Model II(0) is a LBM with product form (BEDD model) and for $\epsilon > 0$, Model II(ϵ) with graphon $w_\epsilon = \lambda\bar{w}_\epsilon$ is a perturbed version and rescaled to preserve the same density λ. As ϵ grows, the graphon of Model II(ϵ) strays further from a BEDD model. Indeed, one can show that $d(\bar{w}_\epsilon) = 64\epsilon^2(5+2\epsilon)^2/(9+4\epsilon)^4$, which is an increasing function for $\epsilon \geq 0$.

Under Model II(ϵ), for each value $N \in \{2^{k/2} : 10 \leq k \leq 22\}$, $\rho \in \{1/8, 1/2\}$, $\epsilon \in \{0.5, 1, 1.5, 2, 2.5, 3\}$, we have simulated $K = 500$ networks of size $m_N \times n_N$ where $m_N = \lfloor \rho N \rfloor$ and $n_N = N - m_N$. For each network, we have computed $\widehat{d}_N$ and $\widehat{V}^d_N$ as estimates of $d(\bar{w}_\epsilon)$ and V^d respectively. The Q-Q plots of the studentized statistic Z^d_N are given in Figures 3.5. It is apparent that Z^d_N is not centered, especially for $N \lesssim 512$. This is due to the fact that $\widehat{d}_N$ is obtained via the delta method, so it is a biased estimator of $d(\bar{w}_\epsilon)$ for finite values of N. However, the bias converges to 0 when N grows, so we find that the statistic achieves normality for $N \gtrsim 2048$. Figure 3.6 gives for different values of ϵ, the coverage probability for $d(\bar{w}_\epsilon)$ of the interval $\left[\widehat{d}_N - \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}^d_N}}, \widehat{d}_N + \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}^d_N}}\right]$ where Φ is the quantile function of the standard normal distribution. We observe that convergence is fastest when $\rho = 1/2$ but also when ϵ is larger, which seems to indicate that estimation of $d(\bar{w})$ is more precise for square matrices and when w is further away from a product model.

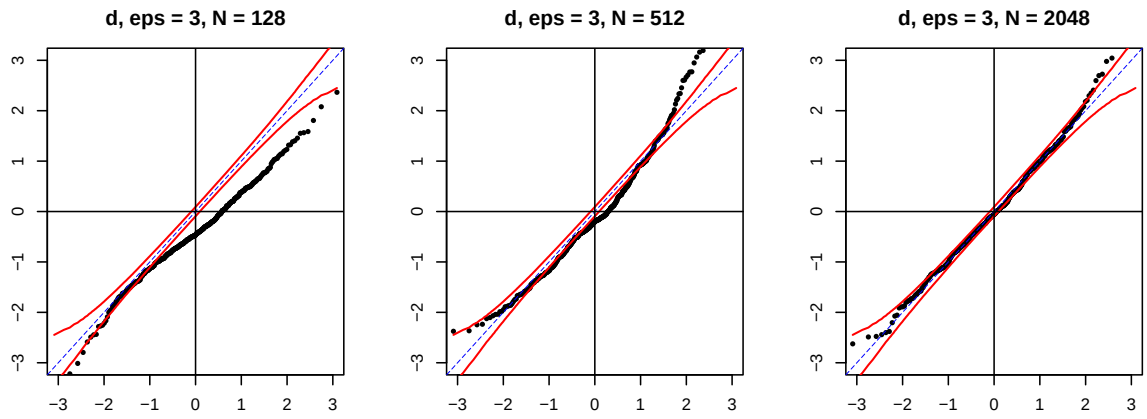

Figure 3.5 – Q-Q plot for Z_N^d the studentized statistic associated with $\widehat{d}_N$, $\epsilon = 3$, $\rho = 0.5$

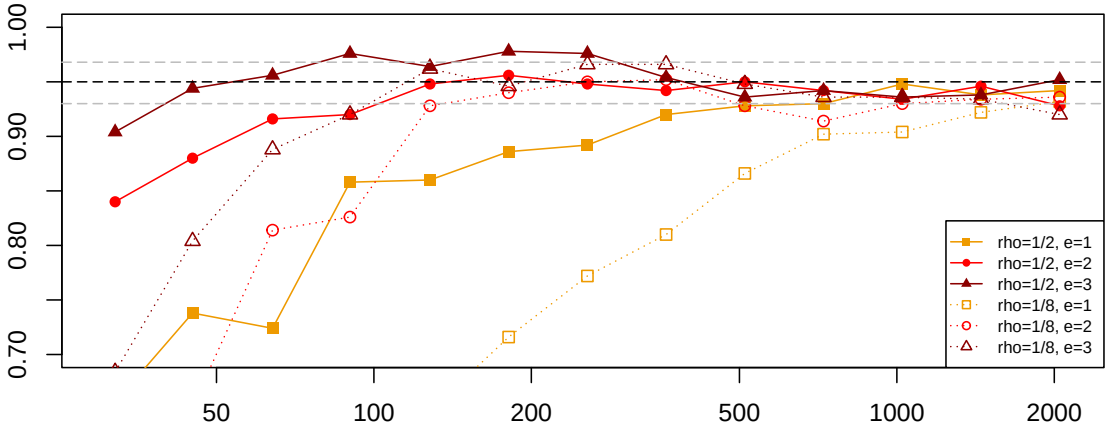

Figure 3.6 – Empirical coverage probabilities for the asymptotic confidence intervals at level $\alpha = 0.95$ of $\widehat{d}_N$ for different values of N (x-axis), $\rho \in \{1/8, 1/2\}$, $\epsilon \in \{1, 2, 3\}$. Grey dashed lines represent the confidence interval at level 0.95 of the frequency $Z = X/K$, if X follows the binomial distribution with parameters K and $\alpha = 0.95$.

3.6.3. Heterogeneity in the row weights of a network

Model III In this example, we consider a weighted BEDD model with power-law strength distributions, i.e. the marginals f and g have the form $f(\xi) = (\alpha_f + 1)\xi^{\alpha_f}$ and $g(\eta) = (\alpha_g + 1)\eta^{\alpha_g}$, where α_f and α_g are real non-negative numbers. α_f is directly related to $F_2 = \int f(\xi)^2 d\xi = (\alpha_f + 1)^2/(2\alpha_f + 1)$.

Under Model III, for each value $N \in \{2^{k/2} : 10 \leq k \leq 22\}$ and $\rho \in \{1/8, 1/2, 7/8\}$, we have simulated $K = 500$ networks of size $m_N \times n_N$ where $m_N = \lfloor \rho N \rfloor$ and $n_N = N - m_N$. We estimate F_2 and G_2 using $\widehat{F}_{2,N}$ and its symmetric counterpart $\widehat{G}_{2,N}$. We

also compute $\widehat{V}_N^{F_2}$ and $\widehat{V}_N^{G_2}$ to obtain the studentized statistics $Z_N^{F_2}$ and $Z_N^{G_2}$. Q-Q plots of $Z_N^{F_2}$ are given in Figures 3.7. Figure 3.8 gives the respective coverage probabilities for F_2 and G_2 of the intervals $\left[\widehat{F}_{2,N} - \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{F_2}}}, \widehat{F}_{2,N} + \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{F_2}}}\right]$ and $\left[\widehat{G}_{2,N} - \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{G_2}}}, \widehat{G}_{2,N} + \Phi(1-\frac{\alpha}{2})\sqrt{\frac{N}{\widehat{V}_N^{G_2}}}\right]$ respectively, where Φ is the quantile function of the standard normal distribution. Despite the bias due to the delta method, apparent for $N = 64$ in Figure 3.7, we find that the coverage probabilities fall in the confidence intervals even for small values of N. We would expect that for F_2, the worst case correspond to $\rho = 1/8$ (there are less rows) and that for G_2, the worst case correspond to $\rho = 7/8$ (there are less columns), but Figure 3.8 does not show a clear difference between the three values of ρ.

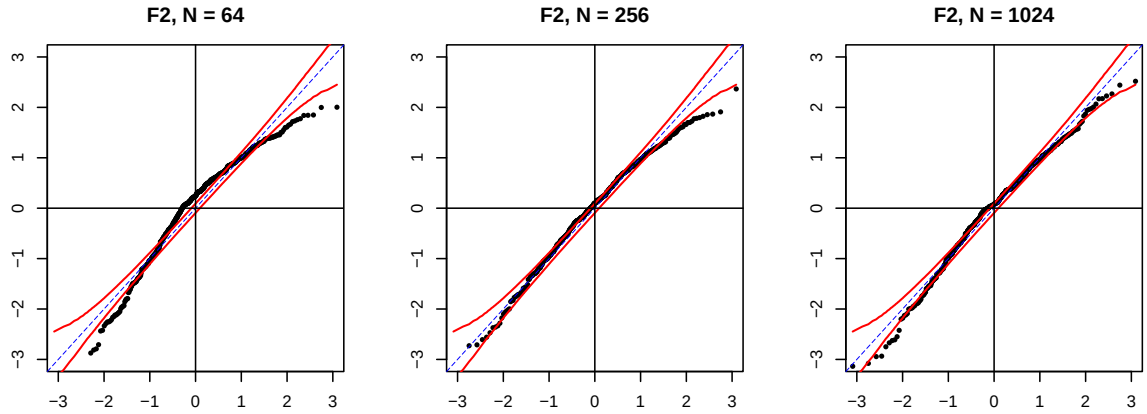

Figure 3.7 – Q-Q plots for $Z_N^{F_2}$ the studentized statistic associated with $\widehat{F}_{2,N}$, $\rho = 0.5$

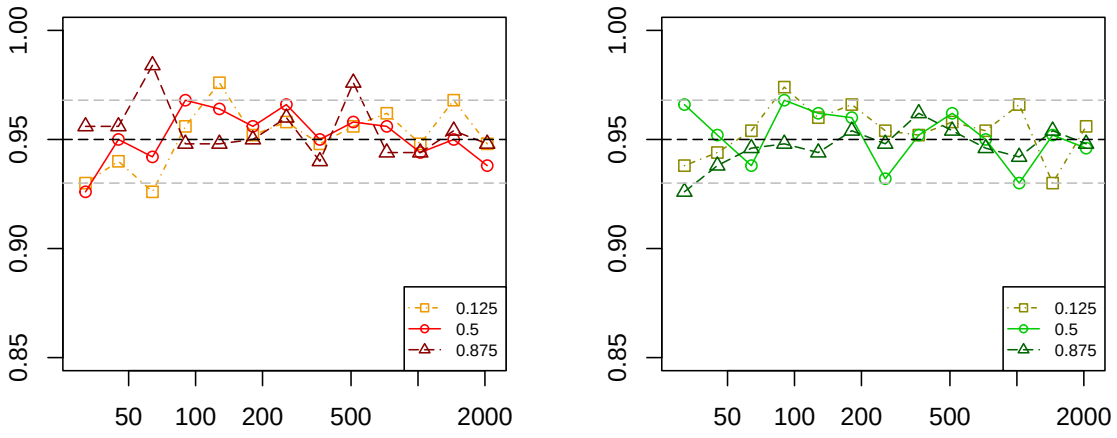

Figure 3.8 – Empirical coverage probabilities for the asymptotic confidence intervals at level $\alpha = 0.95$ of $Z_N^{F_2}$ (left) and $Z_N^{G_2}$ (right) for different values of N (x-axis), $\rho \in \{1/8, 1/2, 7/8\}$. Grey dashed lines represent the confidence interval at level 0.95 of the frequency $Z = X/K$, if X follows the binomial distribution with parameters K and $\alpha = 0.95$.

3.7. Illustrations

To illustrate the use and interpretation of some of the U-statistics introduced in this paper, we considered the set of law-makers networks compiled by Michalska-Smith and Allesina (2019). The database contains networks arising from different fields (ecology, social sciences, life sciences). We focused on the subset of so-called 'legislature' networks both because of their sizes and because network comparison is of interest for this dataset.

Data description. Four law-maker assemblies were considered: the European Parliament ('EP'), the General Assembly of the United Nations ('UN'), the US House of Representatives ('USH') and the US Senate ('USS'). One network has been recorded each year for each parliament; we considered the 26 years from 1979 to 2004, for which the data are available for all the four assemblies. The network recorded for a given assembly in a given year consists of the votes (yes or no) of the different members (rows of the adjacency matrix) for the different proposed laws (columns of the adjacency matrix).
Figure 3.9 gives the dimensions and densities of the 26 networks collected in each assembly: the main difference is that the European Parliament is both larger (both in terms of members and laws) and sparser than the three others.

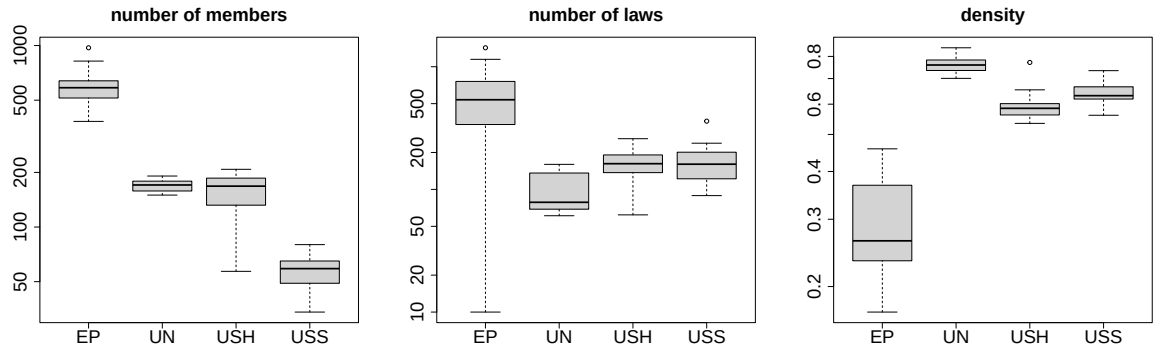

Figure 3.9 – Distribution of the number of members (left), number of laws (center) and density (right) of the four lawmakers networks across the 26 years (in log-scale).

Degree imbalance. We then focused on the degree of imbalance among the rows (resp. columns), which, under the weighted BEDD model defined in Equation (3.10), can be measured by the U-statistic F_2 (resp. G_2). Figure 3.10 gives the evolution of each of the two indicators along the years for each parliament. We observe that, for each of them, both F_2 and G_2 remain above 1, all along the period: as expected no uniformity exist, neither among the members

($F_2 > 1$), nor among the laws ($G_2 > 1$). Regarding the US networks (USH and USS), the imbalanced if more marked among the laws than between the members. As expected also, the confidence intervals are narrower for the largest networks (EP). No systematic pattern is observed, except the shift in the imbalance among resolutions voted at the General Assembly of the United Nations (UN) that is observed in 1992 (and which happens to coincide with the UN membership of former soviet republics).

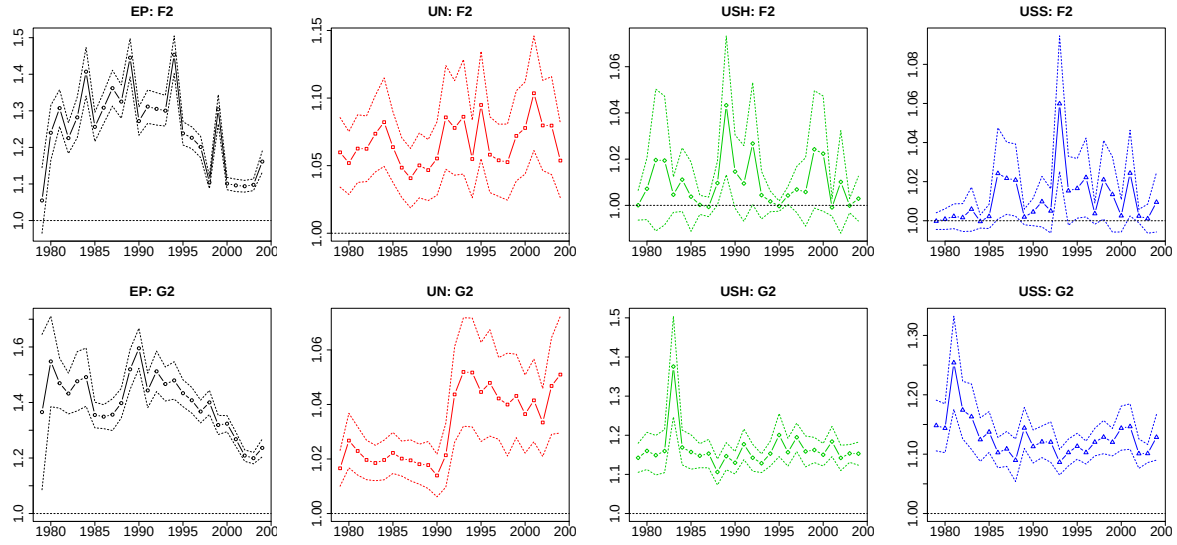

Figure 3.10 – Evolution of degree imbalance in each assembly along the years. Top: F_2 U-statistics, bottom: G_2. From left to right: European Parliament, General Assembly of the United Nations, US House of Representatives and US Senate. Solid line: U-statistic as an estimate of F_2 (resp. G_2). Dotted line: 95%-confidence interval for F_2 (resp. G_2).

Network comparison. For each available year, we then compared the networks of the four assemblies in terms of degree imbalance (F_2 and G_2) and frequency of topological motifs 6 (as given in Figure 3.1). We chose this motif as it constitutes a clique, characterizing a group behavior, in which close members are in favor of the same laws. For each of these parameters, we use the comparison test procedure described in Section 3.5.3.

Figure 3.11 display the results. We observe no significant difference between the two US assemblies, which are also the smallest ones: the absence of significant differences can therefore result from a weak power of the tests when considering small networks. We also observe a higher heterogeneity among the members of the European Parliament with respect to all other assemblies, as well as a higher heterogeneity among the member of the United Nations assembly, with respect to the two US chambers. A different picture is obtained for the heterogeneity among the laws, which is significantly higher in the UN assembly and significantly lower in the EP

assembly.

The frequency of motif 6 is interesting, as it may reveal a specific socio-political behavior. To this respect, the group structure turns out to be much stronger in the UN than in the EP, the members of which represent both different political orientations and different nations.

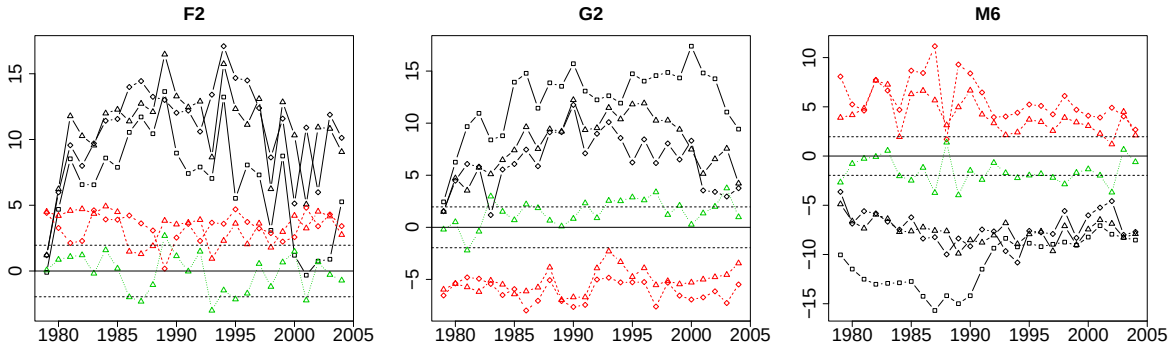

Figure 3.11 – Network comparison for the U-statistics of F_2, G_2 and for the count of the motif 6 (M6). EP-UN= —□—, EP-USS= —◇—, EP-USH= —△—, UN-USS= - -◇- -, UN-USH= - -△- -, USS-USH= ··· △ ···. Horizontal lines = standard normal quantiles with level .025 and .975.

Appendix 3.A Backward martingales

Here, we present the backward martingales and their convergence theorem, which is used to prove the convergence of some estimators. The proof of Theorem 3.A.3 can be found in Doob (1953), Section 7, Theorem 4.2. We recall beforehand the definition of a decreasing filtration.

Definition 3.A.1. *A decreasing filtration is a decreasing sequence of σ-fields $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$, i.e. such that for all $n \geq 1$, $\mathcal{F}_{n+1} \subset \mathcal{F}_n$.*

Definition 3.A.2. *Let $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$ be a decreasing filtration and $M = (M_n)_{n\geq 1}$ a sequence of integrable random variables adapted to $\mathcal{F}$. $(M_n, \mathcal{F}_n)_{n\geq 1}$ is a backward martingale if and only if for all $n \geq 1$, $\mathbb{E}[M_n \mid \mathcal{F}_{n+1}] = M_{n+1}$.*

Theorem 3.A.3. *Let $(M_n, \mathcal{F}_n)_{n\geq 1}$ be a backward martingale. Then, $(M_n)_{n\geq 1}$ is uniformly integrable, and, denoting $M_\infty = \mathbb{E}[M_1 \mid \mathcal{F}_\infty]$ where $\mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \mathcal{F}_n$, we have*

$$M_n \xrightarrow[n\to\infty]{a.s., L_1} M_\infty.$$

Appendix 3.B Proofs of the results presented in Section 3.2

Proof of Lemma 3.2.2. We prove this lemma by induction on (r,c) the sizes of $\mathbf{i}$ and $\mathbf{j}$. For $(r,c) = (1,0)$ and $(r,c) = (0,1)$, we have

$$\mathbb{E}[p^{1,0}h(Y_{\mathbf{i},\varnothing}) \mid \mathcal{A}_{\varnothing,\varnothing}] = \mathbb{E}[\psi^{1,0}h(Y_{\mathbf{i},\varnothing})] - \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})] = 0$$

and

$$\mathbb{E}[p^{0,1}h(Y_{\varnothing,\mathbf{j}}) \mid \mathcal{A}_{\varnothing,\varnothing}] = \mathbb{E}[\psi^{0,1}h(Y_{\varnothing,\mathbf{i}})] - \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})] = 0.$$

Suppose that the lemma is true for all $(0,0) < (r',c') < (r,c)$. Let $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_r(\mathbb{N}) \times \mathcal{P}_c(\mathbb{N})$, $\underline{\mathbf{i}} \subset \mathbf{i}$ and $\underline{\mathbf{j}} \subset \mathbf{j}$. Denote $\underline{r} = \mathrm{Card}(\underline{\mathbf{i}})$ and $\underline{c} = \mathrm{Card}(\underline{\mathbf{j}})$. We can write

$$\begin{aligned}
\mathbb{E}[p^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] &= \mathbb{E}[\psi^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] - \mathbb{E}[p^{\underline{r},\underline{c}}h(Y_{\underline{\mathbf{i}},\underline{\mathbf{j}}}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] \\
&\quad - \sum_{(0,0)<(r',c')<(r,c)} \sum_{\substack{\mathbf{i}' \in \mathcal{P}_{r'}(\mathbf{i}), \mathbf{j}' \in \mathcal{P}_{c'}(\mathbf{j}) \\ (\mathbf{i}',\mathbf{j}') \neq (\underline{\mathbf{i}},\underline{\mathbf{j}})}} \mathbb{E}[p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] \\
&= \sum_{(0,0)\leq(r',c')<(\underline{r},\underline{c})} \sum_{\mathbf{i}' \in \mathcal{P}_{r'}(\underline{\mathbf{i}}), \mathbf{j}' \in \mathcal{P}_{c'}(\underline{\mathbf{j}})} \mathbb{E}[p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] \\
&\quad - \sum_{(0,0)<(r',c')<(r,c)} \sum_{\substack{\mathbf{i}' \in \mathcal{P}_{r'}(\mathbf{i}), \mathbf{j}' \in \mathcal{P}_{c'}(\mathbf{j}) \\ (\mathbf{i}',\mathbf{j}') \neq (\underline{\mathbf{i}},\underline{\mathbf{j}})}} \mathbb{E}[p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] \\
&= - \sum_{(0,0)<(r',c')<(r,c)} \sum_{\substack{\mathbf{i}' \in \mathcal{P}_{r'}(\mathbf{i}), \mathbf{j}' \in \mathcal{P}_{c'}(\mathbf{j}) \\ \mathbf{i}' \not\subset \underline{\mathbf{i}}, \mathbf{j}' \not\subset \underline{\mathbf{j}}}} \mathbb{E}[p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'}) \mid \mathcal{A}_{\mathbf{i}' \cap \underline{\mathbf{i}}, \mathbf{j}' \cap \underline{\mathbf{j}}}].
\end{aligned}$$

where we have used the fact that the $p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'})$ are measurable by their respective $\mathcal{A}_{\mathbf{i}',\mathbf{j}'}$ so that $\mathbb{E}[p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] = \mathbb{E}[p^{r',c'}h(Y_{\mathbf{i}',\mathbf{j}'}) \mid \mathcal{A}_{\mathbf{i}' \cap \underline{\mathbf{i}}, \mathbf{j}' \cap \underline{\mathbf{j}}}]$.

Since the last sum excludes the case $\mathbf{i}' = \underline{\mathbf{i}}$ and $\mathbf{j}' = \underline{\mathbf{j}}$, then the induction hypothesis ensures that all the terms are equal to 0, so $\mathbb{E}[p^{r,c}h(Y_{\mathbf{i},\mathbf{j}}) \mid \mathcal{A}_{\underline{\mathbf{i}},\underline{\mathbf{j}}}] = 0$, which concludes the proof by induction. □

Proof of Corollary 3.2.3. First, we see that for some (r_1,c_1) and (r_2,c_2),

$$\begin{aligned}
&\mathrm{Cov}(P_{m,n}^{r_1,c_1}h_1(Y), P_{m,n}^{r_2,c_2}h_2(Y)) \\
&\quad = \left[\binom{m}{r_1}\binom{m}{r_2}\binom{n}{c_1}\binom{n}{c_2}\right]^{-1} \sum_{\substack{\mathbf{i}_1 \in \mathcal{P}_{r_1}([\![m]\!]) \\ \mathbf{j}_1 \in \mathcal{P}_{c_1}([\![n]\!])}} \sum_{\substack{\mathbf{i}_2 \in \mathcal{P}_{r_2}([\![m]\!]) \\ \mathbf{j}_2 \in \mathcal{P}_{c_2}([\![n]\!])}} \mathrm{Cov}(p^{r_1,c_1}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1}), p^{r_2,c_2}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2})).
\end{aligned}$$

If $(r_1,c_1) \neq (r_2,c_2)$, then from Proposition 3.2.1, all the covariance terms are equal to 0, so $\mathrm{Cov}(P_{m,n}^{r_1,c_1}h_1(Y), P_{m,n}^{r_2,c_2}h_2(Y)) = 0$ and that is the first part of the corollary.

If $(r_1, c_1) = (r_2, c_2) = (r, c)$, then from Proposition 3.2.1, the covariance terms

$$\mathrm{Cov}(p^{r,c}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1}), p^{r,c}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2})) = 0$$

if $(\mathbf{i}_1, \mathbf{j}_1) \neq (\mathbf{i}_2, \mathbf{j}_2)$. Using this fact and the exchangeability of Y, we have

$$\begin{aligned}
&\mathrm{Cov}(P^{r,c}_{m,n}h_1(Y), P^{r,c}_{m,n}h_2(Y)) \\
&\qquad = \left[\binom{m}{r}\binom{n}{c}\right]^{-2} \sum_{\substack{\mathbf{i}_1 \in \mathcal{P}_r([\![m]\!]) \\ \mathbf{j}_1 \in \mathcal{P}_c([\![n]\!])}} \sum_{\substack{\mathbf{i}_2 \in \mathcal{P}_r([\![m]\!]) \\ \mathbf{j}_2 \in \mathcal{P}_c([\![n]\!])}} \mathrm{Cov}(p^{r_1,c_1}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1}), p^{r_2,c_2}h_2(Y_{\mathbf{i}_2,\mathbf{j}_2})) \\
&\qquad = \left[\binom{m}{r}\binom{n}{c}\right]^{-2} \sum_{\substack{\mathbf{i}_1 \in \mathcal{P}_r([\![m]\!]) \\ \mathbf{j}_1 \in \mathcal{P}_c([\![n]\!])}} \mathrm{Cov}(p^{r_1,c_1}h_1(Y_{\mathbf{i}_1,\mathbf{j}_1}), p^{r_2,c_2}h_2(Y_{\mathbf{i}_1,\mathbf{j}_1})) \\
&\qquad = \left[\binom{m}{r}\binom{n}{c}\right]^{-1} \mathrm{Cov}(p^{r,c}h_1(Y_{[\![r]\!],[\![c]\!]}), p^{r,c}h_2(Y_{[\![r]\!],[\![c]\!]})),
\end{aligned}$$

which proves the second part of the corollary. □

Appendix 3.C Proofs of the results presented in Section 3.3

Proof of Lemma 3.3.2. We only prove the first convergence result, since the second can be deduced by analogy.

By definition,

$$p^{1,0}_{(\{i\},\varnothing)}h = \psi^{1,0}_{(\{i\},\varnothing)}h - \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})].$$

First, we see that the $p^{1,0}_{(\{i\},\varnothing)}h(Y)$ only depends on ξ_i, so all the $p^{1,0}_{(\{i\},\varnothing)}h(Y)$ are independent. Because h is symmetric and Y is RCE, they are also identically distributed.

By the tower rule, we also have $\mathbb{E}[\psi^{1,0}_{(\{i\},\varnothing)}h] = \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})]$ so $\mathbb{E}[p^{1,0}_{(\{i\},\varnothing)}h] = 0$.

Finally, we have, from the Aldous-Hoover representation theorem, then Cauchy-Schwarz's inequality, and the fact that $\mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2] < \infty$,

$$\begin{aligned}
\mathbb{V}[p^{1,0}_{(\{i\},\varnothing)}h] &= v^{1,0}_h \\
&= \mathbb{V}[\mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)}) \mid \xi_i]] \\
&= \mathrm{Cov}(h(Y_{(1,\dots,p;1,\dots,q)}), h(Y_{(1,p+1,\dots,2p-1;1,\dots,q)})) \\
&< \mathbb{V}[h(Y_{(1,\dots,p;1,\dots,q)})] \\
&< \infty.
\end{aligned}$$

The classical CLT gives the desired result. □

Proof of Lemma 3.3.3. For all $\mathbf{i} \in \mathcal{P}_r(\llbracket m \rrbracket)$ and $\mathbf{j} \in \mathcal{P}_c(\llbracket n \rrbracket)$ and $(r,c) \in \mathbb{N}^2$, we have $\mathbb{E}[\psi^{r,c}_{(\mathbf{i},\mathbf{j})} h] = \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})]$.

By recursion, we have $\mathbb{E}[A_N] = \mathbb{E}[P^{r,c}_N h] = 0$ for all $(r,c) > (0,0)$ since

$$\begin{aligned}\mathbb{E}[P^{r,c}_N h] &= \mathbb{E}[p^{r,c}_{(\mathbf{i},\mathbf{j})} h] \\ &= \mathbb{E}[\psi^{r,c}_{(\mathbf{i},\mathbf{j})} h] - \sum_{(0,0)\le(r',c')<(r,c)} \sum_{\substack{\mathbf{i}'\in\mathcal{P}_{r'}(\mathbf{i}) \\ \mathbf{j}'\in\mathcal{P}_{c'}(\mathbf{j})}} \mathbb{E}[p^{r',c'}_{(\mathbf{i}',\mathbf{j}')} h].\end{aligned}$$

Now, Corollary 3.2.4 imply that $\mathrm{Cov}(P^{r,c}_N h, P^{r',c'}_N h) = 0$ unless $(r,c) = (r',c')$. So

$$\begin{aligned}\mathbb{V}[A_N] &= N \sum_{\substack{(0,0)<(r,c)\le(p,q) \\ (r,c)\neq(1,0)\neq(0,1)}} \sum_{\substack{(0,0)<(r',c')\le(p,q) \\ (r',c')\neq(1,0)\neq(0,1)}} \binom{p}{r}\binom{p}{r'}\binom{q}{c}\binom{q}{c'} \mathrm{Cov}(P^{r,c}_N h, P^{r',c'}_N h) \\ &= N \sum_{\substack{(0,0)<(r,c)\le(p,q) \\ (r,c)\neq(1,0)\neq(0,1)}} \binom{p}{r}^2\binom{q}{c}^2 \mathbb{V}[P^{r,c}_N h] \\ &= N \sum_{\substack{(0,0)<(r,c)\le(p,q) \\ (r,c)\neq(1,0)\neq(0,1)}} \binom{p}{r}^2\binom{q}{c}^2 \binom{m_N}{r}^{-1}\binom{n_N}{c}^{-1} \mathbb{V}[p^{r,c}_{(\llbracket r\rrbracket,\llbracket c\rrbracket)} h].\end{aligned} \tag{3.18}$$

From equation 3.5, we see that $h(Y_{(1,\dots p,1,\dots,q)})$ is a linear combination of all the $(p^{r,c}_{(\mathbf{i},\mathbf{j})} h)_{\substack{0\le r\le p, 0\le c\le q \\ \mathbf{i}\in\mathcal{P}_r(\llbracket p\rrbracket),\mathbf{j}\in\mathcal{P}_c(\llbracket q\rrbracket)}}$. So $\mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2] < \infty$ ensures that $\mathbb{V}[p^{r,c}_{(\llbracket r\rrbracket,\llbracket c\rrbracket)} h] < \infty$. Therefore $\mathbb{V}[A_N] = O(N m_N^{-r} n_N^{-c}) = O(N^{1-r-c})$.

Thus, we deduce from Markov's inequality that $A_N \xrightarrow[N\to\infty]{\mathbb{P}} 0$. □

Lemma 3.C.1. *1.* $U^{\tilde{h}}_N = U^h_N$,

2. $\mathbb{E}[\tilde{h}(Y_{(1,\dots,p';1,\dots,q')})] = \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})]$,

3. $\mathbb{E}[\tilde{h}(Y_{(1,\dots,p';1,\dots,q')})^2] \le \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2]$.

Proof of Lemma 3.C.1. The two first properties are straightforward. The third property stems

from

$$\begin{aligned}
\mathbb{E}[\tilde{h}(Y_{(1,\dots,p';1,\dots,q')})^2] &= \left[\binom{p'}{p}\binom{q'}{q}\right]^{-2} \sum_{\substack{\mathbf{i}_1\subset\{i_1,\dots,i_{p'}\}\\ \mathbf{j}_1\subset\{j_1,\dots,j_{q'}\}}} \sum_{\substack{\mathbf{i}_2\subset\{i_1,\dots,i_{p'}\}\\ \mathbf{j}_2\subset\{j_1,\dots,j_{q'}\}}} \mathbb{E}[h(Y_{\mathbf{i}_1,\mathbf{j}_1})h(Y_{\mathbf{i}_2,\mathbf{j}_2})] \\
&= \left[\binom{p'}{p}\binom{q'}{q}\right]^{-2} \sum_{\substack{\mathbf{i}_1\subset\{i_1,\dots,i_{p'}\}\\ \mathbf{j}_1\subset\{j_1,\dots,j_{q'}\}}} \sum_{\substack{\mathbf{i}_2\subset\{i_1,\dots,i_{p'}\}\\ \mathbf{j}_2\subset\{j_1,\dots,j_{q'}\}}} \Big(\mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2] \\
&\quad -\frac{1}{2}\mathbb{E}[(h(Y_{\mathbf{i}_1,\mathbf{j}_1}) - h(Y_{\mathbf{i}_2,\mathbf{j}_2}))^2]\Big) \\
&\leq \left[\binom{p'}{p}\binom{q'}{q}\right]^{-2} \sum_{\substack{\mathbf{i}_1\subset\{i_1,\dots,i_{p'}\}\\ \mathbf{j}_1\subset\{j_1,\dots,j_{q'}\}}} \sum_{\substack{\mathbf{i}_2\subset\{i_1,\dots,i_{p'}\}\\ \mathbf{j}_2\subset\{j_1,\dots,j_{q'}\}}} \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2] \\
&\leq \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2].
\end{aligned}$$

□

Proof of Corollary 3.3.5. First, we show that $\lim_{N\to+\infty} N\mathrm{Cov}(U_N^{h_k}, U_N^{h_\ell})$

Let $(Z^{h_k})_{1\leq k\leq D}$ be a vector of random variables following a centered multivariate Gaussian distribution with covariance matrix $\Sigma^{h_1,\dots,h_D}$ defined in the theorem. Then $Z^{h_k} \sim \mathcal{N}(0, V^{h_k})$ for all $1 \leq k \leq D$ and $\mathrm{Cov}(Z^{h_k}, Z^{h_\ell}) = C^{h_k,h_\ell}$ for all $1 \leq k \leq D$ and $1 \leq \ell \leq D$.

Denote $p' := \max_k(p_k)$ and $q' := \max_k(q_k)$. For some $t = (t_1, t_2, ..., t_n) \in \mathbb{R}^n$, we set $\tilde{h}_t := t_1\tilde{h}_1 + t_2\tilde{h}_2 + ... + t_n\tilde{h}_D$. $\tilde{h}_t$ is a kernel function of size $p' \times q'$.

First, assume that $t \neq (0, ..., 0)$. Then by hypothesis, $\tilde{h}_t \not\equiv 0$, therefore Proposition 3.C.1 implies that $\sum_{k=1}^D t_k U_N^{h_k} = \sum_{k=1}^D t_k U_N^{\tilde{h}_k} = U_N^{\tilde{h}_t}$, the U-statistic with quadruplet kernel $\tilde{h}_t$ (of size $p' \times q'$). Using Cauchy-Schwarz inequality and the fact that from Proposition 3.C.1, $\mathbb{E}[\tilde{h}_k(Y_{(1,\dots,p';1,\dots,q')})^2] \leq \mathbb{E}[h_k(Y_{(1,\dots,p;1,\dots,q)})^2] < \infty$ for all $1 \leq k \leq D$, we have furthermore

$$\begin{aligned}
\mathbb{E}[\tilde{h}_t(Y_{(1,\dots,p';1,\dots,q')})^2] &= \sum_{k=1}^n t_k^2 \mathbb{E}[\tilde{h}_k(Y_{(1,\dots,p';1,\dots,q')})^2] \\
&\quad + 2\sum_{1\leq k\neq \ell\leq D} t_k t_\ell \mathbb{E}[\tilde{h}_k(Y_{(1,\dots,p';1,\dots,q')})\tilde{h}_\ell(Y_{(1,\dots,p';1,\dots,q')})], \\
&\leq \sum_{k=1}^n t_k^2 \mathbb{E}[\tilde{h}_k(Y_{(1,\dots,p';1,\dots,q')})^2] \\
&\quad + 2\sum_{1\leq k\neq \ell\leq D} t_k t_\ell \sqrt{\mathbb{E}[\tilde{h}_k(Y_{(1,\dots,p';1,\dots,q')})^2]\mathbb{E}[\tilde{h}_\ell(Y_{(1,\dots,p';1,\dots,q')})^2]}, \\
&< \infty.
\end{aligned}$$

Therefore, Theorem 3.3.1 also applies for $U_N^{\tilde{h}_t}$ and $\sqrt{N}(U_N^{\tilde{h}_t} - U_\infty^{\tilde{h}_t}) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V^{\tilde{h}_t})$, where :

- $U_\infty^{\tilde{h}_t} = \sum_{k=1}^D t_k U_\infty^{h_k}$,

- $V^{\tilde{h}_t} = t^T \Sigma^{h_1,\dots,h_D} t$.

The second point comes from the fact that $V^{\tilde{h}_t} = \lim_{N\to+\infty} N \sum_{k=1}^{D}\sum_{\ell=1}^{D} t_k t_\ell \mathrm{Cov}(U_N^{h_k}, U_N^{h_\ell})$ and by Corollary 3.2.4, $\lim_{N\to+\infty} N\mathrm{Cov}(U_N^{h_k}, U_N^{h_\ell}) = \frac{p^2}{c} c^{1,0}_{h_k,h_\ell} + \frac{q^2}{1-c} c^{0,1}_{h_k,h_\ell} = \Sigma^{h_1,\dots,h_D}_{k\ell}$. Therefore, we can conclude that $\sqrt{N}\sum_{k=1}^{D} t_k (U_N^{h_k} - U_\infty^{h_k}) = \sqrt{N}(U_N^{\tilde{h}_t} - U_\infty^{\tilde{h}_t}) \xrightarrow[N\to\infty]{\mathcal{D}} \sum_{k=1}^{D} t_k Z^{h_k}$.

Now assume that $t = (0,\dots,0)$. Then $\tilde{h}_t \equiv 0$ so $U_N^{\tilde{h}_t} = 0 = \sum_{k=1}^{D} t_k Z^{h_k}$. Therefore, $\sqrt{N}\sum_{k=1}^{D} t_k (U_N^{h_k} - U_\infty^{h_k}) = \sqrt{N}(U_N^{\tilde{h}_t} - U_\infty^{\tilde{h}_t}) \xrightarrow[N\to\infty]{\mathcal{D}} \sum_{k=1}^{D} t_k Z^{h_k}$ is still true.

We have proven that $\sqrt{N}\sum_{k=1}^{D} t_k (U_N^{h_k} - U_\infty^{h_k}) \xrightarrow[N\to\infty]{\mathcal{D}} \sum_{k=1}^{D} t_k Z^{h_k}$ for all $t \in \mathbb{R}^n$, so we can finally apply the Cramér-Wold theorem (Theorem 29.4 of Billingsley, 1995) which states that $\sqrt{N}\left(U_N^{h_k} - U_\infty^{h_k}\right)_{1\le k\le D}$ converges jointy in distribution to $(Z^{h_k})_{1\le k\le D}$, which is a centered multivariate Gaussian with covariance matrix $\Sigma^{h_1,\dots,h_D}$. □

Appendix 3.D Proofs of the results presented in Section 3.4

Lemma 3.D.1. *Let $k \in \mathbb{N}$. Then as $n \to \infty$,*

$$\binom{n}{k} = \frac{n^k}{k!} + O(n^{k-1})$$

Proof of Lemma 3.4.1. The sum $T_N^{p,q}(I_K, J_K)$ defined by equation (3.6) is a sum over T_N expectation terms where

$$T_N := \mathrm{Card}(\mathcal{T}^{p,q}_{N,(I_K,J_K)}) = \prod_{k=1}^{K} \mathrm{Card}(\mathcal{S}^{p,q}_{N,(\underline{\mathbf{i}}_k,\underline{\mathbf{j}}_k)}) = \prod_{k=1}^{K} \binom{m_N - \underline{p}_k}{p - \underline{p}_k}\binom{n_N - \underline{q}_k}{q - \underline{q}_k}. \tag{3.19}$$

These expectation terms $\mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1}\dots X_{\mathbf{i}_K,\mathbf{j}_K}]$ only depend on the number of times elements appear in pairwise intersections between the $(\mathbf{i}_k, \mathbf{j}_k)$, $1 \le k \le K$. In particular, $\mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1}\dots X_{\mathbf{i}_K,\mathbf{j}_K}] = \alpha(I_K, J_K)$ when all the $\mathbf{i}_k$ contains the elements of $\underline{\mathbf{i}}_k$ and all the other elements do not appear in any other $\underline{\mathbf{i}}_{k'}$ (see equation 3.7). Denote A_N the number of terms of $T_N^{p,q}(I_K, J_K)$ where $\mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1}\dots X_{\mathbf{i}_K,\mathbf{j}_K}] = \alpha(I_K, J_K)$, then

$$T_N^{p,q}(I_K, J_K) = \frac{A_N}{T_N}\alpha(I_K, J_K) + \frac{1}{T_N} \sum_{\substack{\mathcal{T}^{p,q}_{N,(I_K,J_K)} \\ \mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1}\dots X_{\mathbf{i}_K,\mathbf{j}_K}] \neq \alpha(I_K,J_K)}} \mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1} X_{\mathbf{i}_2,\mathbf{j}_2}\dots X_{\mathbf{i}_K,\mathbf{j}_K}]. \tag{3.20}$$

Using Jensen's inequality, Hölder's inequality and the exchangeability of the submatrices, we have for all the expectation terms,

$$0 \le |\mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1}\dots X_{\mathbf{i}_K,\mathbf{j}_K}]| \le \mathbb{E}[|X_{\mathbf{i}_1,\mathbf{j}_1}\dots X_{\mathbf{i}_K,\mathbf{j}_K}|] \le \prod_{k=1}^{K} \mathbb{E}[|X_{\mathbf{i}_k,\mathbf{j}_k}|^K]^{\frac{1}{K}} = \mathbb{E}[|X_{(1,\dots,p;1,\dots,q)}|^K].$$

In particular, this holds for all $(T_N - A_N)$ terms of the remaining sum in Equation (3.20) when $\mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1}...X_{\mathbf{i}_K,\mathbf{j}_K}] \neq \alpha(I_K, J_K)$, so

$$0 \leq |T_N^{p,q}(I_K, J_K) - \frac{A_N}{T_N}\alpha(I_K, J_K)| \leq \left(1 - \frac{A_N}{T_N}\right)\mathbb{E}[|X_{(1,...,p;1,...,q)}|^K]. \tag{3.21}$$

We need to calculate T_N and A_N to use this inequality and conclude the proof. We organize the rest of the proof in 3 parts.

1. We find an expression for T_N. Since for all $1 \leq k \leq K$, from Lemma 3.D.1,

$$\binom{m_N - \underline{p}_k}{p - \underline{p}_k}\binom{n_N - \underline{q}_k}{q - \underline{q}_k} = \frac{m_N^{p-\underline{p}_k} n_N^{q-\underline{q}_k}}{(p-\underline{p}_k)!(q-\underline{q}_k)!}\left(1 + O\left(m_N^{-1} + n_N^{-1}\right)\right),$$

we have from Equation (3.19)

$$T_N = \left(\prod_{k=1}^{K}(p-\underline{p}_k)!(q-\underline{q}_k)!\right)^{-1} m_N^{Kp-P} n_N^{Kq-Q}\left(1 + O\left(m_N^{-1} + n_N^{-1}\right)\right). \tag{3.22}$$

2. We find an expression for A_N. The summation over $(\mathbf{i}_k, \mathbf{j}_k) \in \mathcal{S}^{p,q}_{N,(\underline{\mathbf{i}}_k,\underline{\mathbf{j}}_k)}$ is in fact a sum over the $p - \underline{p}_k$ elements of $\mathbf{i}_k$ and $q - \underline{q}_k$ elements of $\mathbf{j}_k$ that are not restricted by $\underline{\mathbf{i}}_k$ and $\underline{\mathbf{j}}_k$.
   - We can pick the first $\mathbf{i}_1$ choosing the $p - \underline{p}_1$ unrestricted indices among the $m_N - \bar{p}$ values of $\{1, ..., m_N\}$ excluding $\cup_{k=1}^{K}\underline{\mathbf{i}}_k$. The same follows for the pick of $\mathbf{j}_1$, so there are $\binom{m_N-\bar{p}}{p-\underline{p}_1}\binom{n_N-\bar{q}}{q-\underline{q}_1}$ possible picks for $(\mathbf{i}_1, \mathbf{j}_1)$.
   - The pick of $\mathbf{i}_2$ consists in choosing the $p - \underline{p}_2$ unrestricted indices among the $m_N - \bar{p} - (p - \underline{p}_1)$ values of $\{1, ..., m_N\}$ excluding $\cup_{k=1}^{K}\underline{\mathbf{i}}_k$ and the elements already taken by $\mathbf{i}_1$. We deduce that there are $\binom{m_N-\bar{p}-(p-\underline{p}_1)}{p-\underline{p}_2}\binom{n_N-\bar{q}-(q-\underline{q}_1)}{q-\underline{q}_2}$ possible picks for $(\mathbf{i}_2, \mathbf{j}_2)$.
   - Iteratively, for all $1 \leq k \leq K$, we find that there are $\binom{m_N-\bar{p}-\sum_{k'=1}^{k-1}(p-\underline{p}_{k'})}{p-\underline{p}_k}\binom{n_N-\bar{q}-\sum_{k'=1}^{k-1}(q-\underline{q}_{k'})}{q-\underline{q}_k}$ possible picks for $(\mathbf{i}_k, \mathbf{j}_k)$.

   We deduce that the number of possible picks for all the $(\mathbf{i}_k, \mathbf{j}_k)$, $1 \leq k \leq K$ so that $\mathbb{E}[X_{\mathbf{i}_1,\mathbf{j}_1}...X_{\mathbf{i}_K,\mathbf{j}_K}] = \alpha(I_K, J_K)$ is

$$A_N = \prod_{k=1}^{K}\binom{m_N - \bar{p} - \sum_{k'=1}^{k-1}(p-\underline{p}_{k'})}{p-\underline{p}_k}\binom{n_N - \bar{q} - \sum_{k'=1}^{k-1}(q-\underline{q}_{k'})}{q-\underline{q}_k}.$$

   But we see that from Lemma 3.D.1, for $1 \leq k \leq K$,

$$\binom{m_N - \bar{p} - \sum_{k'=1}^{k-1}(p-\underline{p}_{k'})}{p-\underline{p}_k}\binom{n_N - \bar{q} - \sum_{k'=1}^{k-1}(q-\underline{q}_{k'})}{q-\underline{q}_k} = \frac{m_N^{p-\underline{p}_k} n_N^{q-\underline{q}_k}}{(p-\underline{p}_k)!(q-\underline{q}_k)!}\left(1 + O\left(m_N^{-1} + n_N^{-1}\right)\right).$$

   So we have

$$A_N = \left(\prod_{k=1}^{K}(p-\underline{p}_k)!(q-\underline{q}_k)!\right)^{-1} m_N^{Kp-P} n_N^{Kq-Q}\left(1 + O\left(m_N^{-1} + n_N^{-1}\right)\right).$$

3. Now with the expressions of T_N and A_N, we can deduce that

$$T_N - A_N = O\left(m_N^{Kp-\underline{P}} n_N^{Kq-\underline{Q}} (m_N^{-1} + n_N^{-1})\right),$$

so

$$1 - \frac{A_N}{T_N} = \frac{O\left(m_N^{Kp-\underline{P}} n_N^{Kq-\underline{Q}} (m_N^{-1} + n_N^{-1})\right)}{\left(\prod_{k=1}^{K} (p - \underline{p}_k)!(q - \underline{q}_k)!\right)^{-1} m_N^{Kp-\underline{P}} n_N^{Kq-\underline{Q}} \left(1 + O\left(m_N^{-1} + n_N^{-1}\right)\right)} = O\left(m_N^{-1} + n_N^{-1}\right).$$

Therefore, we can finally conclude using Equation (3.21),

$$\begin{aligned} T_N^{p,q}(I_K, J_K) &= \left(1 + O\left(m_N^{-1} + n_N^{-1}\right)\right) \alpha(I_K, J_K) + O\left(m_N^{-1} + n_N^{-1}\right) \\ &= \alpha(I_K, J_K) + O\left(m_N^{-1} + n_N^{-1}\right). \end{aligned}$$

□

Proof of Proposition 3.4.5. In this proof, for the estimator defined by equation (3.8), we write $\widehat{\mu}_N^{(i)}$ instead of $\widehat{\mu}_N^{h,(i)}$ to simplify the notation without ambiguity.

Notice that

$$\begin{aligned} \mathbb{E}[\widehat{v}_N^{h;1,0}] &= \frac{1}{2}\mathbb{E}[(\widehat{\mu}_N^{(1)} - \widehat{\mu}_N^{(2)})^2] \\ &= \mathbb{E}[(\widehat{\mu}_N^{(1)})^2] - \mathbb{E}[\widehat{\mu}_N^{(1)}\widehat{\mu}_N^{(2)}]. \end{aligned} \tag{3.23}$$

- First, we calculate $\mathbb{E}[(\widehat{\mu}_N^{(1)})^2]$:

$$\begin{aligned} \mathbb{E}[\widehat{\mu}_N^{(1)}(Y)^2] &= \binom{m_N - 1}{p-1}^{-2} \binom{n_N}{q}^{-2} \sum_{(\mathbf{i}_1,\mathbf{j}_1) \in \mathcal{S}_{N,(\{1\},\varnothing)}^{p,q}} \sum_{(\mathbf{i}_2,\mathbf{j}_2) \in \mathcal{S}_{N,(\{1\},\varnothing)}^{p,q}} \mathbb{E}[h(Y_{\mathbf{i}_1,\mathbf{j}_1}) h(Y_{\mathbf{i}_2,\mathbf{j}_2})] \\ &= T_N^{p,q}(I_2, J_2) \end{aligned}$$

where $I_2 = (\{1\}, \{1\})$ and $J_2 = (\varnothing, \varnothing)$. Applying Lemma 3.4.1, with $\underline{P} = \mathrm{Card}(\{1\}) + \mathrm{Card}(\{1\}) = 2$, $\underline{Q} = \mathrm{Card}(\varnothing) + \mathrm{Card}(\varnothing) = 0$ and

$$\begin{aligned} \alpha(I_2, J_2) &= \mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)}) h(Y_{(1,p+1,\ldots,2p-1;q+1,\ldots,2q)})] \\ &= \mathbb{E}[\mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)}) h(Y_{(1,p+1,\ldots,2p-1;q+1,\ldots,2q)}) \mid \xi_1]] \\ &= \mathbb{E}[\mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)}) \mid \xi_1]^2] \\ &= \mathbb{E}[\psi_{(\{1\},\varnothing)}^{1,0} h^2]. \end{aligned}$$

we find

$$\mathbb{E}[(\widehat{\mu}_N^{(1)})^2] = \mathbb{E}[\psi_{(\{1\},\varnothing)}^{1,0} h^2] + O\left(N^{-1}\right). \tag{3.24}$$

- Next, we calculate $\mathbb{E}[\widehat{\mu}_N^{(1)}\widehat{\mu}_N^{(2)}]$:

$$\begin{aligned}\mathbb{E}[\widehat{\mu}_N^{(1)}\widehat{\mu}_N^{(2)}] &= \binom{m_N-1}{p-1}^{-2}\binom{n_N}{q}^{-2}\sum_{(\mathbf{i}_1,\mathbf{j}_1)\in\mathcal{S}^{p,q}_{N,(\{1\},\varnothing)}}\sum_{(\mathbf{i}_2,\mathbf{j}_2)\in\mathcal{S}^{p,q}_{N,(\{2\},\varnothing)}}\mathbb{E}[h(Y_{\mathbf{i}_1,\mathbf{j}_1})h(Y_{\mathbf{i}_2,\mathbf{j}_2})]\\ &= T_N^{p,q}(I_2',J_2')\end{aligned}$$

where $I_2' = (\{1\},\{2\})$ and $J_2' = (\varnothing,\varnothing)$. Applying Lemma 3.4.1 with

$$\begin{aligned}\alpha(I_2',J_2') &= \mathbb{E}[h(Y_{(1,3,\dots,p+1;1,\dots,q)})h(Y_{(2,p+2,\dots,2p;q+1,\dots,2q)})]\\ &= \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})]^2\\ &= \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h]^2,\end{aligned}$$

we find

$$\mathbb{E}[\widehat{\mu}_N^{(1)}\widehat{\mu}_N^{(2)}] = \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h]^2 + O\left(N^{-1}\right). \tag{3.25}$$

Finally, we can combine Equations (3.23), (3.24) and (3.25) to obtain

$$\mathbb{E}[\widehat{v}_N^{h;1,0}] = \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h^2] - \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h]^2 + O\left(N^{-1}\right) = v_h^{1,0} + O\left(N^{-1}\right),$$

which proves the theorem. □

Proof of Proposition 3.4.6. In this proof, we write $\widehat{\mu}_N^{(i)}$ instead of $\widehat{\mu}_N^{h,(i)}$ to simplify the notation without ambiguity.

Notice that

$$\begin{aligned}\mathbb{E}[(\widehat{v}_N^{h;1,0})^2] &= \binom{m_N}{2}^{-2}\sum_{1\le i_1<i_2\le m_N}\sum_{1\le i_1'<i_2'\le m_N}\mathbb{E}\left[\frac{(\widehat{\mu}_N^{(i_1)}-\widehat{\mu}_N^{(i_2)})^2(\widehat{\mu}_N^{(i_1')}-\widehat{\mu}_N^{(i_2')})^2}{4}\right]\\ &= \binom{m_N}{2}^{-1}\sum_{1\le i_1<i_2\le m_N}\frac{1}{4}\mathbb{E}[(\widehat{\mu}_N^{(1)}-\widehat{\mu}_N^{(2)})^2(\widehat{\mu}_N^{(3)}-\widehat{\mu}_N^{(4)})^2] + O\left(N^{-1}\right)\\ &= \mathbb{E}[(\widehat{\mu}_N^{(1)})^2(\widehat{\mu}_N^{(2)})^2] - 2\mathbb{E}[(\widehat{\mu}_N^{(1)})^2\widehat{\mu}_N^{(2)}\widehat{\mu}_N^{(3)}] + \mathbb{E}[\widehat{\mu}_N^{(1)}\widehat{\mu}_N^{(2)}\widehat{\mu}_N^{(3)}\widehat{\mu}_N^{(4)}]\\ &\quad + O\left(N^{-1}\right).\end{aligned} \tag{3.26}$$

Now, we calculate each of the three expectation terms in this equation.

- First, we calculate $\mathbb{E}[(\widehat{\mu}_N^{(1)})^2(\widehat{\mu}_N^{(2)})^2]$:

$$\begin{aligned}&\mathbb{E}[(\widehat{\mu}_N^{(1)})^2(\widehat{\mu}_N^{(2)})^2]\\ &\quad= \binom{m_N-1}{p-1}^{-4}\binom{n_N}{q}^{-4}\sum_{(\mathbf{i}_1,\mathbf{j}_1)\in\mathcal{S}^{p,q}_{N,(\{1\},\varnothing)}}\sum_{(\mathbf{i}_2,\mathbf{j}_2)\in\mathcal{S}^{p,q}_{N,(\{1\},\varnothing)}}\sum_{(\mathbf{i}_3,\mathbf{j}_3)\in\mathcal{S}^{p,q}_{N,(\{2\},\varnothing)}}\sum_{(\mathbf{i}_4,\mathbf{j}_4)\in\mathcal{S}^{p,q}_{N,(\{2\},\varnothing)}}\\ &\qquad\mathbb{E}[h(Y_{\mathbf{i}_1,\mathbf{j}_1})h(Y_{\mathbf{i}_2,\mathbf{j}_2})h(Y_{\mathbf{i}_3,\mathbf{j}_3})h(Y_{\mathbf{i}_4,\mathbf{j}_4})]\\ &\quad= T_N^{p,q}(I_4,J_4)\end{aligned}$$

where $I_4 = (\{1\},\{1\},\{2\},\{2\})$ and $J_4 = (\varnothing,\varnothing,\varnothing,\varnothing)$. Applying Lemma 3.4.1 with

$$\begin{aligned}\alpha(I_4,J_4) &= \mathbb{E}[h(Y_{(1,3,\ldots,p+1;1,\ldots,q)})h(Y_{(1,p+2,\ldots,2p;q+1,\ldots,2q)}) \\ &\quad \times h(Y_{(2,2p+1,\ldots,3p-1;2q+1,\ldots,3q)})h(Y_{(2,3p,\ldots,4p-2;3q+1,\ldots,4q)})] \\ &= \mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)})h(Y_{(1,p+1,\ldots,2p-1;q+1,\ldots,2q)})]^2 \\ &= \mathbb{E}[\mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)})h(Y_{(1,p+1,\ldots,2p-1;q+1,\ldots,2q)}) \mid \xi_1]]^2 \\ &= \mathbb{E}[\mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)}) \mid \xi_1]^2]^2 \\ &= \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h^2]^2.\end{aligned}$$

we find

$$\mathbb{E}[(\widehat{\mu}_N^{(1)})^2(\widehat{\mu}_N^{(2)})^2] = \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h^2]^2 + O\left(N^{-1}\right). \tag{3.27}$$

- Next, we calculate $\mathbb{E}[(\widehat{\mu}_N^{(1)})^2\widehat{\mu}_N^{(2)}\widehat{\mu}_N^{(3)}]$:

$$\begin{aligned}&\mathbb{E}[(\widehat{\mu}_N^{(1)})^2\widehat{\mu}_N^{(2)}\widehat{\mu}_N^{(3)}] \\ &\quad = \binom{m_N-1}{p-1}^{-4}\binom{n_N}{q}^{-4}\sum_{(\mathbf{i}_1,\mathbf{j}_1)\in\mathcal{S}^{p,q}_{N,(\{1\},\varnothing)}}\sum_{(\mathbf{i}_2,\mathbf{j}_2)\in\mathcal{S}^{p,q}_{N,(\{1\},\varnothing)}}\sum_{(\mathbf{i}_3,\mathbf{j}_3)\in\mathcal{S}^{p,q}_{N,(\{2\},\varnothing)}}\sum_{(\mathbf{i}_4,\mathbf{j}_4)\in\mathcal{S}^{p,q}_{N,(\{3\},\varnothing)}} \\ &\qquad \mathbb{E}[h(Y_{\mathbf{i}_1,\mathbf{j}_1})h(Y_{\mathbf{i}_2,\mathbf{j}_2})h(Y_{\mathbf{i}_3,\mathbf{j}_3})h(Y_{\mathbf{i}_4,\mathbf{j}_4})] \\ &\quad = T_N^{p,q}(I_4',J_4')\end{aligned}$$

where $I_4' = (\{1\},\{1\},\{2\},\{3\})$ and $J_4' = (\varnothing,\varnothing,\varnothing,\varnothing)$. Applying Lemma 3.4.1 with

$$\begin{aligned}\alpha(I_4',J_4') &= \mathbb{E}[h(Y_{(1,4,\ldots,p+2;1,\ldots,q)})h(Y_{(1,p+3,\ldots,2p+1;q+1,\ldots,2q)}) \\ &\quad \times h(Y_{(2,2p+2,\ldots,3p;2q+1,\ldots,3q)})h(Y_{(3,3p+1,\ldots,4p-1;3q+1,\ldots,4q)})] \\ &= \mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)})h(Y_{(1,p+1,\ldots,2p-1;q+1,\ldots,2q)})]^2\mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)})]^2 \\ &= \mathbb{E}[\mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)})h(Y_{(1,p+1,\ldots,2p-1;q+1,\ldots,2q)}) \mid \xi_1]]\mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)]^2 \\ &= \mathbb{E}[\mathbb{E}[h(Y_{(1,\ldots,p;1,\ldots,q)}) \mid \xi_1]^2]\mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h]^2 \\ &= \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)^2]\mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h]^2.\end{aligned}$$

we find

$$\mathbb{E}[(\widehat{\mu}_N^{(1)}(Y))^2\widehat{\mu}_N^{(2)}(Y)\widehat{\mu}_N^{(3)}(Y)] = \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)^2]\mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)]^2 + O\left(N^{-1}\right). \tag{3.28}$$

- Now, we calculate $\mathbb{E}[\widehat{\mu}_N^{(1)}(Y)\widehat{\mu}_N^{(2)}(Y)\widehat{\mu}_N^{(3)}(Y)\widehat{\mu}_N^{(4)}(Y)]$:

$$\begin{aligned}&\mathbb{E}[\widehat{\mu}_N^{(1)}(Y)\widehat{\mu}_N^{(2)}(Y)\widehat{\mu}_N^{(3)}(Y)\widehat{\mu}_N^{(4)}(Y)] \\ &\quad = \binom{m_N-1}{p-1}^{-4}\binom{n_N}{q}^{-4}\sum_{(\mathbf{i}_1,\mathbf{j}_1)\in\mathcal{S}^{p,q}_{N,(\{1\},\varnothing)}}\sum_{(\mathbf{i}_2,\mathbf{j}_2)\in\mathcal{S}^{p,q}_{N,(\{2\},\varnothing)}}\sum_{(\mathbf{i}_3,\mathbf{j}_3)\in\mathcal{S}^{p,q}_{N,(\{3\},\varnothing)}}\sum_{(\mathbf{i}_4,\mathbf{j}_4)\in\mathcal{S}^{p,q}_{N,(\{4\},\varnothing)}} \\ &\qquad \mathbb{E}[h(Y_{\mathbf{i}_1,\mathbf{j}_1})h(Y_{\mathbf{i}_2,\mathbf{j}_2})h(Y_{\mathbf{i}_3,\mathbf{j}_3})h(Y_{\mathbf{i}_4,\mathbf{j}_4})] \\ &\quad = T_N^{p,q}(I_4'',J_4'')\end{aligned}$$

where $I_4' = (\{1\},\{2\},\{3\},\{4\})$ and $J_4' = (\varnothing,\varnothing,\varnothing,\varnothing)$. Applying Lemma 3.4.1 with

$$\begin{aligned}\alpha(I_4'', J_4'') &= \mathbb{E}[h(Y_{(1,5,\dots,p+3;1,\dots,q)})h(Y_{(2,p+4,\dots,2p+2;q+1,\dots,2q)}) \\ &\quad \times h(Y_{(3,2p+3,\dots,3p+1;2q+1,\dots,3q)})h(Y_{(4,3p+2,\dots,4p;3q+1,\dots,4q)})] \\ &= \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})]^4 \\ &= \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)]^4.\end{aligned}$$

we find

$$\mathbb{E}[\widehat{\mu}_N^{(1)}(Y)\widehat{\mu}_N^{(2)}(Y)\widehat{\mu}_N^{(3)}(Y)\widehat{\mu}_N^{(4)}(Y)] = \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)]^4 + O\left(N^{-1}\right). \tag{3.29}$$

Finally, injecting the calculated expressions (3.27), (3.28) and (3.29) in (3.26), we obtain

$$\begin{aligned}\mathbb{E}[(\widehat{v}_N^{h;1,0})^2] &= \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)^2]^2 - 2\mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)^2]\mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)]^2 \\ &\quad + \mathbb{E}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)]^4 + O\left(N^{-1}\right) \\ &= \mathbb{V}[\psi^{1,0}_{(\{1\},\varnothing)}h(Y)^2]^2 + O\left(N^{-1}\right) \\ &= (v_h^{1,0})^2 + O\left(N^{-1}\right) \\ &= \mathbb{E}[\widehat{v}_N^{h;1,0}]^2 + O\left(N^{-1}\right)\end{aligned}$$

where we have applied Proposition 3.4.5 in the last step. This proves that $\mathbb{V}[\widehat{v}_N^{h;1,0}] = O\left(N^{-1}\right)$, concluding the proof. □

Appendix 3.E Proofs of the results presented in Section 3.5

Lemma 3.E.1. *Let Y be a matrix sampled from a Poisson W-graph model. Let $\bar{w}$, f and g be defined as in Section 3.5.2. For the kernel functions defined in Table 3.1, we have*

- $\mathbb{E}[h_{A,1}(Y_{(i_1,i_2;j_1,j_2)})] = \lambda^3 \iint \bar{w}(\xi,\eta)^2 d\xi d\eta$,
- $\mathbb{E}[h_{A,2}(Y_{(i_1,i_2;j_1,j_2)})] = \lambda^3 \iint \bar{w}(\xi,\eta) f(\xi) g(\eta) d\xi d\eta$,
- $\mathbb{E}[h_B(Y_{(i_1;j_1,j_2)})] = \lambda^2 \int f(\xi)^2 d\xi$,
- $\mathbb{E}[h_C(Y_{(i_1,i_2;j_1)})] = \lambda^2 \int g(\eta)^2 d\eta$,
- $\mathbb{E}[h_D(Y_{(i_1;j_1)})] = \lambda$.

Proof. The result for h_D is straightforward. For the other kernel functions :

$$\begin{aligned}\mathbb{E}[h_{A,1}(Y_{(i_1,i_2;j_1,j_2)})] &= \mathbb{E}[\mathbb{E}[Y_{i_1j_1}(Y_{i_1j_1}-1)Y_{i_2j_2} \mid \boldsymbol{\xi},\boldsymbol{\eta}]] \\ &= \mathbb{E}[\mathbb{E}[Y_{i_1j_1}(Y_{i_1j_1}-1) \mid \boldsymbol{\xi},\boldsymbol{\eta}]\mathbb{E}[Y_{i_2j_2} \mid \boldsymbol{\xi},\boldsymbol{\eta}]] \\ &= \mathbb{E}[\lambda^2 \bar{w}(\xi_{i_1},\eta_{j_1})^2 \times \lambda] \\ &= \lambda^3 \iint \bar{w}(\xi,\eta)^2 d\xi d\eta.\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[h_{A,2}(Y_{(i_1,i_2;j_1,j_2)})] &= \mathbb{E}[\mathbb{E}[Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_2} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\mathbb{E}[Y_{i_1j_1} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]\mathbb{E}[Y_{i_1j_2} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]\mathbb{E}[Y_{i_2j_2} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\lambda\bar{w}(\xi_{i_1}, \eta_{j_1}) \times \lambda\bar{w}(\xi_{i_1}, \eta_{j_2}) \times \lambda\bar{w}(\xi_{i_2}, \eta_{j_2})] \\
&= \lambda^3 \iint \left[\bar{w}(\xi_{i_1}, \eta_{j_2}) \left(\int \bar{w}(\xi_{i_1}, \eta_{j_1}) d\eta_{j_1}\right) \left(\int \bar{w}(\xi_{i_2}, \eta_{j_2}) d\xi_{i_2}\right)\right] d\xi_{i_1} d\eta_{j_2} \\
&= \lambda^3 \iint \bar{w}(\xi, \eta) f(\xi) g(\eta) d\xi d\eta.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[h_B(Y_{(i_1;j_1,j_2)})] &= \mathbb{E}[\mathbb{E}[Y_{i_1j_1}Y_{i_1j_2} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\mathbb{E}[Y_{i_1j_1} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]\mathbb{E}[Y_{i_1j_2} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\lambda\bar{w}(\xi_{i_1}, \eta_{j_1}) \times \lambda\bar{w}(\xi_{i_1}, \eta_{j_2})] \\
&= \lambda^2 \int \left[\left(\int \bar{w}(\xi_{i_1}, \eta_{j_1}) d\eta_{j_1}\right) \left(\int \bar{w}(\xi_{i_1}, \eta_{j_2}) d\eta_{j_2}\right)\right] d\xi_{i_1} \\
&= \lambda^2 \int f(\xi)^2 d\xi.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[h_C(Y_{(i_1,i_2;j_1)})] &= \mathbb{E}[\mathbb{E}[Y_{i_1j_1}Y_{i_2j_1} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\mathbb{E}[Y_{i_1j_1} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]\mathbb{E}[Y_{i_2j_1} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\lambda\bar{w}(\xi_{i_1}, \eta_{j_1}) \times \lambda\bar{w}(\xi_{i_2}, \eta_{j_1})] \\
&= \lambda^2 \int \left[\left(\int \bar{w}(\xi_{i_1}, \eta_{j_1}) d\xi_{i_1}\right) \left(\int \bar{w}(\xi_{i_2}, \eta_{j_1}) d\xi_{i_2}\right)\right] d\eta_{j_1} \\
&= \lambda^2 \int g(\eta)^2 d\eta.
\end{aligned}$$

□

Lemma 3.E.2. *Let Y be a matrix sampled from a W-graph model. Let h_1 and h_2 be the kernel functions defined as in Section 3.5.2. We have*

- $\mathbb{E}[h_1(Y_{(i_1;j_1,j_2)})] = \lambda^2 \int f(\xi)^2 d\xi = \lambda^2 F_2$,
- $\mathbb{E}[h_2(Y_{(i_1,i_2;j_1,j_2)})] = \lambda^2$.

Proof. The proof for h_1 is identical to that for h_B in the proof of Lemma 3.E.1. For h_2,

$$\begin{aligned}
\mathbb{E}[h_2(Y_{(i_1,i_2;j_1,j_2)})] &= \mathbb{E}[\mathbb{E}[Y_{i_1j_1}Y_{i_2j_2} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\mathbb{E}[Y_{i_1j_1} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]\mathbb{E}[Y_{i_2j_2} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \mathbb{E}[\lambda \times \lambda] \\
&= \lambda^2.
\end{aligned}$$

□

Chapter 4

Asymptotic distribution of degenerate U-statistics on bipartite networks

This chapter corresponds to a work in progress, aimed to be submitted later.

Contents

Introduction

Degenerate U-statistics have a completely different asymptotic behavior from that of non-degenerate ones. To explain when degeneracy arises, we recall the weak convergence result for U-statistics of row-column exchangeable (RCE) matrices found in the previous chapters. It is of the form $\sqrt{N}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V)$. However, when $V = 0$, we notice that the limit distribution becomes trivial. In this case, such a result is not much informative and cannot be exploited to perform statistical inference.

The previous chapters have proposed a complete method to investigate non-degenerate U-statistics. In the examples, we have always assumed that the U-statistics are not degenerate. Yet, many U-statistics of interest are degenerate. The first objective of this chapter is to derive weak convergence results for degenerate U-statistics, similar to those valid in the non-degenerate case.

However, even if we obtain such limit theorems, the degeneracy of a U-statistic is not always known *a priori*. In this case, we would not know which results to apply. Calculations are needed to identify whether a U-statistic is degenerate or not. If it is degenerate, the order of degeneracy has also to be found, which is even more tedious. The second objective of this chapter is to explore ways and alternatives to identify degenerate U-statistics and their order of degeneracy.

Before that, we start with an example of degenerate U-statistic, to give some context. We will see that degenerate U-statistics have a faster rate of convergence than non-degenerate U-statistics, which is actually desirable. A possible approach to tackle degenerate U-statistics is the use of so-called incomplete U-statistics, but we will see that they lose their advantage when tackling the rate of convergence.

4.1. A curse or a blessing?

U-statistics of RCE matrices can be degenerate when their kernel h of size $p \times q$ is centered $\mathbb{E}[h(Y_{[\![p]\!],[\![q]\!]})] = 0$. Actually, it is common to see centered kernels, especially when the U-statistic is used as a test statistic. The degeneracy leads to a faster rate of convergence, which means more powerful tests. Let us start with an example to understand how.

4.1.1. An example

Let $(\mathcal{L}(\mu))_{\mu \geq 0}$ be a family of probability distributions and $Y \sim \mathcal{L} - \mathrm{BEDD}(\Theta)$, where $\Theta = (\lambda, f, g)$ are BEDD parameters as defined in Chapter 2. We remind that the BEDD model is defined by

$$\begin{aligned} \xi_i &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall i \geq 1, \\ \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall j \geq 1, \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{L}(\lambda f(\xi_i) g(\eta_j)), && \forall i \geq 1, j \geq 1. \end{aligned}$$

The $\mathcal{L} - BEDD$ has been shown to be fully characterized by λ, $(F_k)_{k \geq 1}$ and $(G_k)_{k \geq 1}$. In two previous examples in Chapters 2 and 3, we have estimated F_2, which is interesting as it quantifies the heterogeneity of the row degrees. Another question of interest is to test if the row degrees are homogeneous, which means $f \equiv 1$. Let us define the null hypothesis $\mathcal{H}_0 : f \equiv 1$ and confront it to $\mathcal{H}_1 : f \not\equiv 1$.

Under $\mathcal{H}_0$, we have $F_2 = 1$ and otherwise $F_2 > 1$. We can use the same U-statistic kernels as in the previous chapters, defined by

$$h_1(Y_{\{i_1,i_2\},\{j_1,j_2\}}) = \frac{1}{2}(Y_{i_1 j_1} Y_{i_1 j_2} + Y_{i_2 j_1} Y_{i_2 j_2}),$$

and

$$h_2(Y_{\{i_1,i_2\},\{j_1,j_2\}}) = \frac{1}{2}(Y_{i_1 j_1} Y_{i_2 j_2} + Y_{i_2 j_1} Y_{i_1 j_2}).$$

For the Poisson-BEDD model, these kernels have expectations $\mathbb{E}[h_1(Y_{\{i_1,i_2\},\{j_1,j_2\}})] = \lambda^2 F_2$ and $\mathbb{E}[h_2(Y_{\{i_1,i_2\},\{j_1,j_2\}})] = \lambda^2$.

We can consider our usual asymptotic framework with growing $m_N \to \infty$, $n_N \to \infty$ and $m_N/N \to \rho \in]0,1[$. $U_N := U_{m_N,n_N}$ is the U-statistic with kernel $h = h_1 - h_2$. We have $\mathbb{E}[U_N] = \lambda^2(F_2 - 1)$ and under $\mathcal{H}_0$, it becomes $\mathbb{E}[U_N] = 0$.

However, we can show that this U-statistic is degenerate under $\mathcal{H}_0$. Indeed, we have both

$$\begin{aligned}\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} - Y_{11}Y_{22} - Y_{21}Y_{12} \mid \xi_1] \\ &= \frac{\lambda^2}{2}(f(\xi_1)^2 + F_2 - 2f(\xi_1)),\end{aligned} \tag{4.1}$$

and

$$\begin{aligned}\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \eta_1] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} - Y_{11}Y_{22} - Y_{21}Y_{12} \mid \eta_1] \\ &= \frac{\lambda^2}{2}(2F_2 - 2)g(\eta_1).\end{aligned} \tag{4.2}$$

Under $\mathcal{H}_0$, these two conditional expectations are equal to 0. Therefore, the asymptotic variance of the previous limit theorems is

$$V = \frac{4}{\rho}\mathbb{V}[\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1]] + \frac{4}{1-\rho}\mathbb{V}[\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \eta_1]] = 0.$$

Thus, these limit theorems are only little informative about the limit distribution of U_N.

Nevertheless, remember that degenerate U-statistics of i.i.d. observations have a faster rate of convergence than non-degenerate U-statistics, due to their variance decreasing faster. If we manage to identify their limit distributions, degenerate U-statistics of row-column exchangeable matrices will also have a much faster rate of convergence than their non-degenerate counterpart.

Figure 4.1 shows the variation of $\log \mathbb{V}[U_N]$ as a function of $\log N$ for simulated networks following a BEDD model with $f \equiv 1$. If U_N is non-degenerate, then $\mathbb{V}[U_N] = V/N + o(N^{-1})$ so the slope of the curve should converge to -1 as $\log N$ increases. Here, we observe that the slope is closer to -3, which seems to indicate degeneracy and some relation of the form $\mathbb{V}[U_N] = V_3/N^3 + o(N^{-3})$, where V_3 is a constant. The example of this section will be investigated in more detail in Section 4.5.

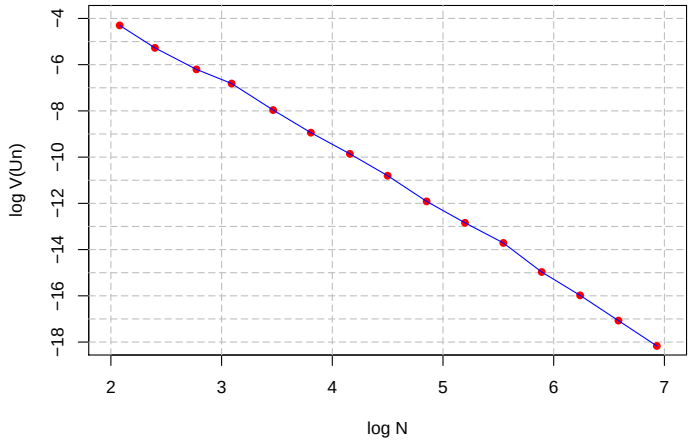

Figure 4.1 – The decreasing empirical variance of U_N. For each value of $N \in \{2^{k/2} : 6 \leq k \leq 20\}$, we have simulated 500 networks of size $\lfloor N/2 \rfloor \times \lfloor N/2 \rfloor$, following a Poisson-BEDD model with $\lambda = 1$, $f \equiv 1$ and a power-law form for g, i.e. $g(\eta) = (\alpha_g + 1)\eta^{\alpha_g}$, and α_g chosen such that $G_2 = 2$. The plot shows the logarithm of the empirical variance $\log \widehat{\mathbb{V}}[U_N]$ as a function of $\log N$.

4.1.2. Incomplete U-statistics

One common approach to tackling the issue of degeneracy in the i.i.d. case is the use of incomplete U-statistics. Consider $(X_1, X_2, ...)$ i.i.d. variables and the kernel function h of k variables. The U-statistic associated to h is

$$U_n = \binom{n}{k}^{-1} \sum_{\mathbf{i} \in \mathcal{P}_k([\![n]\!])} h(X_{\mathbf{i}}).$$

An incomplete U-statistic is defined by

$$U_n^{\mathcal{D}} = \frac{1}{\delta_n} \sum_{\mathbf{i} \in \mathcal{D}_n} h(X_{\mathbf{i}}),$$

where $\mathcal{D} = (\mathcal{D}_n)_{n \geq 1}$ is such that for all $n \geq 1$, $\mathcal{D}_n \in \mathcal{P}_k([\![n]\!])$ and $\delta_n = |\mathcal{D}_n|$. $\mathcal{D}$ is called the design of the incomplete U-statistic. Because incomplete U-statistics average a smaller number of terms than "complete" U-statistics, their variance is greater. Indeed, we have $0 \leq \mathbb{V}[U_n^{\mathcal{D}} - U_n] = \mathbb{V}[U_n^{\mathcal{D}}] - 2\mathrm{Cov}(U_n^{\mathcal{D}}, U_n) + \mathbb{V}[U_n] = \mathbb{V}[U_n^{\mathcal{D}}] - \mathbb{V}[U_n]$.

More especially, denote $f_n^{(c)}$ the number of pairs of sets in $\mathcal{D}_n$ that have c elements in common. A simple calculation can show that the variance of $U_n^{\mathcal{D}}$ is

$$\mathbb{V}[U_n^{\mathcal{D}}] = \delta_n^{-2} \sum_{c=1}^{k} f_n^{(c)} v_c,$$

where $v = \mathrm{Cov}(X_{\mathbf{i}}, X_{\mathbf{i}'})$ where $\mathbf{i}$ and $\mathbf{i}'$ have c elements in common.

Incomplete U-statistics have two main uses. First, the computation of the complete U-statistic requires to average $\binom{n}{k} = O(n^k)$ terms. This becomes costly when n or k are large. By symmetry, one hopes that omitting terms only affects the variance to an acceptable degree. The second use of incomplete U-statistics is that if the design is chosen properly, it can partially solve the problem of degeneracy. Indeed, the following theorem is a slightly reworked version by Lee (1990) of a result by Brown and Kildea (1978). It ensures that, under conditions on the design, incomplete U-statistics always converge to a Gaussian distribution, no matter the degeneracy of the kernel.

Theorem 4.1.1. *Let $\mathcal{D}$ be a design such that $\delta_n/n \to 0$ and for all $1 \le c \le k$, we have $f_n^{(c)}/\delta_n^2 = O(n^{-1})$. Then*

$$\sqrt{\delta_n}(U_n^{\mathcal{D}} - U_\infty) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, v_k).$$

The use of incomplete U-statistics can force the limit distribution to be Gaussian, but it comes to the cost of a way slower rate of convergence. Indeed, we have $\delta_n \le n$ whereas, for degenerate U-statistic, we would have hoped to have the convergence of $n^{d/2}(U_n - U_\infty)$ where $d-1$ is the order of degeneracy.

In conclusion, because of the rate of convergence of degenerate U-statistics, degeneracy would be a desirable property for U-statistics, but only when the form of the limit distribution can be identified. The use of incomplete U-statistics would certainly be possible for our U-statistics. We have not tried to transpose their convergence theorem to U-statistics of row-column exchangeable matrices. However, it is expected that their use cancels the faster rate of convergence. In the next section, we will see that it is possible to identify the distribution of degenerate U-statistics, without reducing their rate of convergence.

4.2. Another orthogonal decomposition for U-statistics

In order to do this, we need to define a new system of Hoeffding-type decomposition, related but different from the one defined in the previous chapter. This new decomposition, as well as the following theoretical developments, are transposed from the theory of generalized U-statistics of Janson and Nowicki (1991) to our setup with bipartite exchangeable networks.

4.2.1. Aldous-Hoover-Kallenberg representation of RCE matrices

We recall the Aldous-Hoover-Kallenberg (AHK) representation for RCE matrices and its consequence on U-statistics. If Y is a dissociated RCE matrix, there exists $(\xi_i)_{i\ge1}$, $(\eta_j)_{j\ge1}$

and $(\zeta_{ij})_{i,j\geq 1}$ arrays of i.i.d. random variables with uniform distribution over $[0,1]$ and a real measurable function ϕ such that for all $1 \leq i, j < \infty$,

$$Y_{ij} \overset{a.s.}{=} \phi(\xi_i, \eta_j, \zeta_{ij}). \tag{4.3}$$

A function of entries of Y can be written with the AHK variables. In particular, the kernel $h(Y_{\mathbf{i},\mathbf{j}})$, where $\mathbf{i} \in \mathcal{P}_p(\mathbb{N})$ and $\mathbf{j} \in \mathcal{P}_q(\mathbb{N})$, can be written as

$$h(Y_{\mathbf{i},\mathbf{j}}) = h\Big((\phi(\xi_i, \eta_j, \zeta_{ij}))_{i\in\mathbf{i}, j\in\mathbf{j}}\Big) =: h_\phi((\xi_i)_{i\in\mathbf{i}}, (\eta_j)_{j\in\mathbf{j}}, (\zeta_{ij})_{i\in\mathbf{i},j\in\mathbf{j}}),$$

and $h_\phi : [0,1]^{p+q+pq} \to \mathbb{R}$ is a symmetric function. The U-statistic with kernel h can be rewritten with h_ϕ as follows

$$U_n = \left[\binom{m}{p}\binom{n}{q}\right]^{-1} \sum_{\substack{\mathbf{i}\in\mathcal{P}_p([\![m]\!]) \\ \mathbf{j}\in\mathcal{P}_q([\![n]\!])}} h_\phi((\xi_i)_{i\in\mathbf{i}}, (\eta_j)_{j\in\mathbf{j}}, (\zeta_{ij})_{i\in\mathbf{i},j\in\mathbf{j}}).$$

Denoting $h_{\mathbf{i},\mathbf{j}} \coloneqq h_\phi((\xi_i)_{i\in\mathbf{i}}, (\eta_j)_{j\in\mathbf{j}}, (\zeta_{ij})_{i\in\mathbf{i},j\in\mathbf{j}})$, we can write

$$U_n = \left[\binom{m}{p}\binom{n}{q}\right]^{-1} \sum_{\substack{\mathbf{i}\in\mathcal{P}_p([\![m]\!]) \\ \mathbf{j}\in\mathcal{P}_q([\![n]\!])}} h_{\mathbf{i},\mathbf{j}}.$$

4.2.2. Graph subsets of AHK variables

The idea behind the new decomposition of a U-statistic is to find orthogonal projections first for $h_{\mathbf{i},\mathbf{j}}$, for all $\mathbf{i}$ and $\mathbf{j}$ first, and then use the previous expression to derive the decomposition for the U-statistic. In order to define the projections for $h_{\mathbf{i},\mathbf{j}}$, we have to define the relevant subspaces for these projections. These subspaces, defined in the next section, are generated by subsets of AHK variables. In order to denote these subsets, we will be using a notation involving bipartite graphs. These graphs have no direct link with the network data, they are just a formalism to define subsets of AHK variables.

Notations for bipartite graphs

We recall that a bipartite graph G is denoted $G = (V_1(G), V_2(G), E(G))$. A subgraph $F \subseteq G$ is such that $V_1(F) \subseteq V_1(G)$, $V_2(F) \subseteq V_2(G)$ and $E(F) \subseteq (V_1(G) \times V_2(G)) \cap E(G)$. We write $F \subset G$ if $F \subseteq G$ and $F \neq G$.

Let $E = \{e_i : i \in I\}$ be a countable set indexed by I and σ some mapping $\sigma : I \to I$. We denote the action of σ on E by $\sigma E = \{e_{\sigma(i)} : i \in I\}$. Let G be a bipartite graph. Suppose that

$V_1(G)$ is indexed by the set I and $V_2(G)$ by the set J. The action of a couple of mappings $\Phi = (\sigma_1, \sigma_2)$ on G, where $\sigma_1 : I \to I$ and $\sigma_1 : J \to J$, is denoted

$$\Phi G := (\sigma_1 V_1(G), \sigma_2 V_2(G), \Phi E(G)), \tag{4.4}$$

where $\Phi E(G) = \{(x_{\sigma_1(i)}, y_{\sigma_2(j)}) : (x_i, y_j) \in E(G), (i,j) \in I \times J\}$. Among these mappings, the bijective ones are called permutations. The group of permutations of $[\![n]\!]$ is denoted $\mathbb{S}_n$.

For two bipartite graphs G_1 and G_2 with same number of row nodes $r = v_1(G_1) = v_1(G_2)$ and column nodes $c = v_2(G_1) = v_2(G_2)$, we say that they are isomorphic if and only if there exists a couple of permutations $\Phi = (\sigma_1, \sigma_2) \in \mathbb{S}_r \times \mathbb{S}_c$ such that $\Phi G_1 = G_2$. In this case, we write $G_1 \sim G_2$. The number of elements Φ of $\mathbb{S}_r \times \mathbb{S}_c$ such that $\Phi G = G$ is the number of automorphisms of G, denoted $|\mathrm{Aut}(G)|$.

We define $K_{\mathbf{i},\mathbf{j}} = (\mathbf{i}, \mathbf{j}, \mathbf{i} \times \mathbf{j})$ the fully connected bipartite graph with row node set $\mathbf{i}$ and column node set $\mathbf{j}$. For $p \geq 0$ and $q \geq 0$, we denote $K_{p,q} = K_{[\![p]\!],[\![q]\!]}$.

For $r \geq 0$ and $c \geq 0$, we can define a minimal set $\Gamma_{r,c}$ of all subgraphs of $K_{r,c}$ with r row nodes and c column nodes, such that every graph G with the same numbers of nodes is isomorphic to exactly one element of $\Gamma_{r,c}$. Denote $\Gamma^-_{p,q} = \cup_{(0,0)<(r,c)\leq(p,q)} \Gamma_{r,c}$. As a reminder, $(0,0) < (r,c) \leq (p,q)$ means $0 \leq r \leq p$, $0 \leq c \leq q$ and $(r,c) \neq (0,0)$. Every non-empty graph G with $v_1(G) \leq p$ and $v_2(G) \leq q$ is isomorphic to exactly one element of $\Gamma^-_{p,q}$.

Definition of graph subsets

Let G be a bipartite graph. We can define the set $H(G)$ of AHK variables associated to G as

$$H(G) = ((\xi_i)_{i \in V_1(G)}, (\eta_j)_{j \in V_2(G)}, (\zeta_{ij})_{(i,j) \in E(G)}).$$

We see that $h_{\mathbf{i},\mathbf{j}} = h_\phi((\xi_i)_{i \in \mathbf{i}}, (\eta_j)_{j \in \mathbf{j}}, (\zeta_{ij})_{i \in \mathbf{i}, j \in \mathbf{j}}) = h_\phi(H(K_{\mathbf{i},\mathbf{j}}))$. In other words, $h_{\mathbf{i},\mathbf{j}}$ belongs to some functional probability space generated by the AHK variables $H(K_{\mathbf{i},\mathbf{j}})$. The subspaces on which $h_{\mathbf{i},\mathbf{j}}$ will be decomposed are generated by subsets of $H(K_{\mathbf{i},\mathbf{j}})$, which are of the form $H(G)$, where $G \subset K_{\mathbf{i},\mathbf{j}}$, as shown in Figure 4.2.

In the following section, we define more rigorously these subspaces and we exhibit some of their properties.

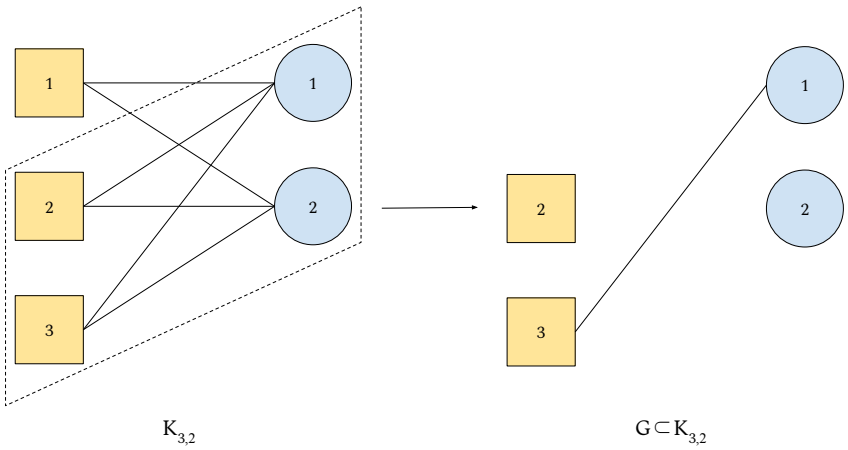

Figure 4.2 – A bipartite graph and one subgraph. For each graph, the row nodes are on the right and the column nodes are on the left. Left: the graph $K_{3,2}$. Right: a subgraph G extracted from the row nodes $\{2,3\}$ and the column nodes $\{1,2\}$ of $K_{3,2}$. Here, G only keeps one edge among the four allowed between the row nodes $\{2,3\}$ and the column nodes $\{1,2\}$. G defines the subset $H(G) = ((\xi_2,\xi_3),(\eta_1,\eta_2),(\zeta_{13}))$.

4.2.3. Decomposition of the probability space

Let G be a bipartite graph and denote $L_2(G)$ the space of all square-integrable random variables measurable with respect to $\sigma(H(G))$. $L_2(G)$ is a Hilbert space with inner product $\langle X,Y\rangle = \mathbb{E}[XY]$. We investigate the following decomposition for $X \in L_2(G)$

$$X = \sum_{F\subseteq G} p^F(X), \tag{4.5}$$

where the $p^F(X)$ are defined by recursion with $p^{\varnothing}(X) = \mathbb{E}[X]$ and for all F,

$$p^F(X) = \mathbb{E}[X \mid H(F)] - \sum_{F'\subset F} p^{F'}(X).$$

Now, we define $L_2^*(G) \subset L_2(G)$ as follows

$$L_2^*(G) = \{X \in L_2(G) : \mathbb{E}[X \mid H(F)] = 0, \forall F \subset G\}. \tag{4.6}$$

These subspaces are linked to the decomposition (4.5). First, we show that each term of the decomposition belongs indeed to one of these spaces, which shows that the decomposition is a decomposition on these subspaces. The following proposition can be shown by induction, as indicated in Appendix 4.A.

Proposition 4.2.1. *For two bipartite graphs $F \subseteq G$ and $X \in L_2(G)$, $p^F(X) \in L_2^*(F)$.*

Now, we prove the most important property of this decomposition. A Hoeffding-type decomposition is an orthogonal decomposition. The following proposition shows that it is the case.

Proposition 4.2.2. *For all bipartite graph G, $L_2(G)$ is the orthogonal direct sum $L_2(G) \oplus^{\perp}_{F \subseteq G} L_2^*(F)$.*

Proof. Equation (4.5) and Proposition 4.2.1 already show that $L_2(G) \oplus_{F \subseteq G} L_2^*(F)$. We only have to show that for any two distinct bipartite graphs G_1 and G_2, we have $L_2^*(G_1) \perp L_2^*(G_2)$. Let $X_1 \in L_2^*(G_1)$ and $X_2 \in L_2^*(G_2)$. Let $\overline{G} = G_1 \cap G_2$. Since G_1 and G_2 are distinct, then at least one of the affirmations $\overline{G} \subset G_1$ and $\overline{G} \subset G_2$ is true. Assume that $\overline{G} \subset G_1$. Therefore, $\mathbb{E}[X_1 X_2] = \mathbb{E}[\mathbb{E}[X_1 X_2 \mid H(G_1)]] = \mathbb{E}[X_1 \mathbb{E}[X_2 \mid H(\overline{G})]] = 0$, so $L_2^*(G_1) \perp L_2^*(G_2)$. □

4.2.4. Decomposition of U-statistics

For all $(0,0) \leq (p,q) \leq (m,n)$, $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p([\![m]\!]) \times \mathcal{P}_q([\![n]\!])$, $G \subseteq K_{\mathbf{i},\mathbf{j}}$, we can apply the decomposition (4.5) on $h_{\mathbf{i},\mathbf{j}} \in L_2(K_{\mathbf{i},\mathbf{j}})$.

$$p^G(h_{\mathbf{i},\mathbf{j}}) = \mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid H(G)] - \sum_{F \subset G} p^F(h_{\mathbf{i},\mathbf{j}}),$$

where $p^{\varnothing}(h_{\mathbf{i},\mathbf{j}}) = \mathbb{E}[h_{\mathbf{i},\mathbf{j}}] = \mathbb{E}[h_{[\![p]\!],[\![q]\!]}]$.

For all $G \subseteq K_{\mathbf{i},\mathbf{j}}$, we remind that $V_1(G) \subseteq \mathbf{i}$ and $V_2(G) \subseteq \mathbf{j}$. Define $\overline{V_1(G)}$ and $\overline{V_2(G)}$ the complements of respectively $V_1(G)$ and $V_2(G)$ in respectively $\mathbf{i}$ and $\mathbf{j}$. In fact, the term $p^G(h_{\mathbf{i},\mathbf{j}})$ does not depend on the elements of $\overline{V_1(G)}$ and $\overline{V_2(G)}$, i.e. even if $(\mathbf{i}_1,\mathbf{j}_1) \neq (\mathbf{i}_2,\mathbf{j}_2)$, as long as $G \subset K_{\mathbf{i}_1,\mathbf{j}_1} \cap K_{\mathbf{i}_2,\mathbf{j}_2}$, we have $p^G(h_{\mathbf{i}_1,\mathbf{j}_1}) = p^G(h_{\mathbf{i}_2,\mathbf{j}_2})$. Therefore, we use the notation $p^G := p^G(h_{\mathbf{i},\mathbf{j}})$, for all $G \in K_{\mathbf{i},\mathbf{j}}$. Finally,

$$h_{\mathbf{i},\mathbf{j}} = \sum_{G \subseteq K_{\mathbf{i},\mathbf{j}}} p^G,$$

and U-statistic $U_{m,n}$ can be rewritten

$$\begin{aligned} U_{m,n} &= \binom{m}{p}^{-1}\binom{n}{q}^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j} \in \mathcal{P}_q([\![n]\!])}} \sum_{G \subseteq K_{\mathbf{i},\mathbf{j}}} p^G \\ &= \binom{m}{p}^{-1}\binom{n}{q}^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j} \in \mathcal{P}_q([\![n]\!])}} \sum_{(0,0) \leq (r,c) \leq (p,q)} \sum_{\substack{G \subseteq K_{\mathbf{i},\mathbf{j}} \\ (v_1(G),v_2(G)) = (r,c)}} p^G \\ &= \sum_{(0,0) \leq (r,c) \leq (p,q)} P_{m,n}^{r,c}, \end{aligned}$$

where $P_{m,n}^{r,c} = \binom{m}{p}^{-1}\binom{n}{r}^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j} \in \mathcal{P}_q([\![n]\!])}} \sum_{\substack{G \subseteq K_{\mathbf{i},\mathbf{j}} \\ (v_1(G),v_2(G)) = (r,c)}} p^G$.

Note that in general, for $G \subseteq K_{\mathbf{i},\mathbf{j}}$, p^G is not symmetric, that means $p^G(h_{\sigma_1\mathbf{i},\sigma_2\mathbf{j}}) \neq p^G(h_{\mathbf{i},\mathbf{j}})$ for a couple of permutations $(\sigma_1,\sigma_2) \in \mathbb{S}_p \times \mathbb{S}_q$. We define $\bar{p}^G$ the symmetrized version of p^G as

$$\bar{p}^G = \sum_{(\sigma_1,\sigma_2)\in\mathbb{S}_p\times\mathbb{S}_q} p^G(h_{\sigma_1\mathbf{i},\sigma_2\mathbf{j}}) = \sum_{\Phi\in\mathbb{S}_p\times\mathbb{S}_q} p^{\Phi G} = \sum_{\substack{G'\subseteq K_{\mathbf{i},\mathbf{j}} \\ G'\sim G}} p^{G'}.$$

For two isomorphic subgraphs G_1 and G_2 of $K_{\mathbf{i},\mathbf{j}}$, we have $\bar{p}^{G_1} = \bar{p}^{G_2}$ by symmetry. There is exactly one element $G \in \Gamma_{r,c}$, where $r = v_1(G_1) = v_1(G_2)$ and $c = v_2(G_1) = v_2(G_2)$, which is isomorphic to both G_1 and G_2. Therefore, for all $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p([\![m]\!]) \times \mathcal{P}_q([\![n]\!])$, we can index these quantities with the graph $G \in \Gamma_{r,c}$ instead of $G \in K_{\mathbf{i},\mathbf{j}}$. Then, we denote

$$\tilde{p}^G_{\mathbf{i},\mathbf{j}} := \bar{p}^{G'},$$

where $G \in \Gamma_{r,c}$ and G' is any subgraph of $K_{\mathbf{i},\mathbf{j}}$ which is isomorphic to G. We can also denote $\tilde{p}^G$ the function $\tilde{p}^G : (\mathbf{i},\mathbf{j}) \longmapsto \tilde{p}^G_{\mathbf{i},\mathbf{j}}$.

Because there are $r!\binom{p}{r}c!\binom{q}{c}|\mathrm{Aut}(G)|^{-1}$ distinct subgraphs of $K_{\mathbf{i},\mathbf{j}}$ that are isomorphic to $G \in \Gamma_{r,c}$, we obtain the following decomposition

$$h_{\mathbf{i},\mathbf{j}} = (p!q!)^{-1} \sum_{G\subseteq K_{\mathbf{i},\mathbf{j}}} \tilde{p}^G = \sum_{0\le(r,c)\le(p,q)} \sum_{G\in\Gamma_{r,c}} \frac{1}{(p-r)!(q-c)!|\mathrm{Aut}(G)|} \tilde{p}^G_{\mathbf{i},\mathbf{j}}$$

and

$$P^{r,c}_{m,n} = \sum_{G\in\Gamma_{r,c}} \frac{1}{(p-r)!(q-c)!|\mathrm{Aut}(G)|} \tilde{P}^G_{m,n},$$

where for all $G \in \Gamma_{r,c}$, $\tilde{P}^G_{m,n} = \binom{m}{p}^{-1}\binom{n}{q}^{-1} \sum_{\substack{\mathbf{i}\in\mathcal{P}_p([\![m]\!]) \\ \mathbf{j}\in\mathcal{P}_q([\![n]\!])}} \tilde{p}^G_{\mathbf{i},\mathbf{j}}$ is the U-statistic of kernel $\tilde{p}^G$. Finally, the $U_{m,n}$ can be rewritten as

$$U_{m,n} = \sum_{0\le(r,c)\le(p,q)} \sum_{G\in\Gamma_{r,c}} \frac{1}{(p-r)!(q-c)!|\mathrm{Aut}(G)|} \tilde{P}^G_{m,n}.$$

This decomposition is related to the one used in Chapter 3. The latter consists in an orthogonal projection of $h_{\mathbf{i},\mathbf{j}} \in L_2(K_{\mathbf{i},\mathbf{j}})$ on the subspaces $(\underline{L}_2(K_{\mathbf{i}',\mathbf{j}'}))_{\mathbf{i}'\subseteq\mathbf{i},\mathbf{j}'\subseteq\mathbf{j}}$, where

$$\underline{L}_2(K_{\mathbf{i},\mathbf{j}}) = \{X \in L_2(K_{\mathbf{i},\mathbf{j}}) : \mathbb{E}[X \mid H(K_{\mathbf{i}',\mathbf{j}'})] = 0, \forall \mathbf{i}' \subseteq \mathbf{i}, \mathbf{j}' \subseteq \mathbf{j}\}.$$

Comparing this with the subspaces (4.6), we see that the decomposition on the subspaces of the form (4.2.4) is coarser, as they only consist in subspaces generated by graphs of the form $K_{\mathbf{i},\mathbf{j}}$. For this reason, it does not capture the subtleties determining the limit distribution of degenerate U-statistics. We will see that the new decomposition is able to fill this gap, at the cost of being more complex.

4.2.5. Variance

Just like the classical Hoeffding decomposition of U-statistics of i.i.d. observations, our new decomposition can also be used to decompose the variance of U-statistics of row-column exchangeable matrices. The following two results come from the orthogonality of the projections. The first expression links $\mathbb{V}[U_{m,n}]$ to the variance of the projections $\mathbb{V}[p^G] = \mathbb{E}[(p^G)^2]$. It is obtained by direct calculation, as shown in Appendix 4.B.

Proposition 4.2.3.

$$\mathbb{V}[U_{m,n}] = \sum_{(0,0)<(r,c)\leq(p,q)} \frac{(m-r)!}{m!}\frac{(n-c)!}{n!} V^{(r,c)},$$

where for all $(0,0) < (r,c) \leq (p,q)$,

$$V^{(r,c)} = \frac{p!^2 q!^2}{(p-r)!^2(q-c)!^2} \sum_{G\in\Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1}\mathbb{E}[(p^G)^2].$$

The second expression links $\mathbb{V}[U_{m,n}]$ to the variance of the U-statistics $\tilde{P}^G_{m,n}$ associated to the symmetrized projections $\tilde{p}^G$.

Corollary 4.2.4.

$$\mathbb{V}[U_{m,n}] = \sum_{0<(r,c)\leq(p,q)} \sum_{G\in\Gamma_{r,c}} \left(\frac{1}{(p-r)!(q-c)!|\mathrm{Aut}(G)|}\right)^2 \mathbb{V}[\tilde{P}^G_{m,n}]$$

It can actually be naturally obtained from Proposition 4.2.3 using the following lemma.

Lemma 4.2.5.

$$\mathbb{V}[\tilde{P}^G_{m,n}] = \frac{(m-r)!}{m!}\frac{(n-c)!}{n!} p!^2 q!^2 |\mathrm{Aut}(G)|\mathbb{E}[(p^G)^2].$$

The proof of this lemma requires to handle the symmetrized projections, which can be tricky. In this regard, the next lemma is particularly helpful. For this reason, it will be used several times in this chapter. The proofs of both lemmas are given in Appendix 4.B.

Lemma 4.2.6. *Let G subgraph of $K_{p,q}$. Let $(G^1_{\mathbf{i},\mathbf{j}})_{\substack{\mathbf{i}\in\mathcal{P}_p([\![m]\!])\\ \mathbf{j}\in\mathcal{P}_q([\![n]\!])}}$ and $(G^2_{\mathbf{i},\mathbf{j}})_{\substack{\mathbf{i}\in\mathcal{P}_p([\![m]\!])\\ \mathbf{j}\in\mathcal{P}_q([\![n]\!])}}$ two families of graphs such that for all $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p([\![m]\!]) \times \mathcal{P}_q([\![n]\!])$, both $G^1_{\mathbf{i},\mathbf{j}}, G^2_{\mathbf{i},\mathbf{j}} \subseteq K_{\mathbf{i},\mathbf{j}}$ and are isomorphic to G. We have*

$$\sum_{\substack{\mathbf{i}_1,\mathbf{i}_2\in\mathcal{P}_p([\![m]\!])\\ \mathbf{j}_1,\mathbf{j}_2\in\mathcal{P}_q([\![n]\!])}} \sum_{\Phi_1,\Phi_2\in\mathbb{S}_p\times\mathbb{S}_q} \mathbb{1}(\Phi_1 G^1_{\mathbf{i}_1,\mathbf{j}_1} = \Phi_2 G^2_{\mathbf{i}_2,\mathbf{j}_2}) = \frac{m!(m-r)!}{(m-p)!^2}\frac{n!(n-c)!}{(n-q)!^2}|\mathrm{Aut}(G)|.$$

4.2.6. Principal part and support graphs

Let us use the usual asymptotic framework, with a sequence for network sizes (m_N, n_N) such that $m_N + n_N = N$ and $m_N/N \xrightarrow[N\to\infty]{} \rho$, for some $\rho \in]0,1[$. We denote $U_N := U_{m_N,n_N}$, $P_N^{r,c} := P_{m_N,n_N}^{r,c}$ and $\tilde{P}_N^G := \tilde{P}_{m_N,n_N}^G$. The kernel h is still a function of a submatrix of size $p \times q$.

In this asymptotic framework, we now define the principal part of U_N. Let

$$p^{(k)} := \sum_{\substack{G \in K_{p,q} \\ v_1(G)+v_2(G)=k}} p^G,$$

for $1 \le k \le p+q$. Let d be the smallest integer such that $p^{(d)} \neq 0$. We call $d-1$ the *degree of degeneracy* of U_N. Then we have $P_N^{r,c} = 0$ for all (r,c) such that $r+c < d$. We call $\sum_{\substack{(0,0)\le(r,c)\le(p,q) \\ r+c=d}} P_N^{r,c}$ the *principal part* of U_N and the couples (r,c) such that $r+c=d$ are the *principal degrees* of U_N. We call the *principal support graphs* of U_N the graphs $G \subseteq K_{m,n}$ such that

- $v_1(G) + v_2(G) = d$,
- $p^G \neq 0$.

Example. Let Y be a random matrix such that $Y_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Let h_1 and h_2 be the kernel functions defined by $h_1(Y_{\{1\},\{1,2\}}) = Y_{11}Y_{12}$ and $h_2(Y_{\{1,2\},\{1,2\}}) = (Y_{11}Y_{22} + Y_{12}Y_{21})/2$, and $U_N^{h_1}$ and $U_N^{h_2}$ are the U-statistics associated to these kernels.
Y admits a natural AHK representation, which is $Y_{ij} \overset{a.s.}{=} \phi(\xi_i, \eta_j, \zeta_{ij}) = \Phi^{-1}(\zeta_{ij})$, where Φ^{-1} is the inverse c.d.f. of the standard Gaussian distribution. Remarkably, Y_{ij} does not depend on the AHK variables ξ_i and η_j. We have $\mathbb{E}[Y_{ij}] = \mathbb{E}[Y_{ij} \mid \xi_i] = \mathbb{E}[Y_{ij} \mid \eta_j] = \mathbb{E}[Y_{ij} \mid \xi_i, \eta_j] = 0$ and $\mathbb{E}[Y_{ij} \mid \xi_i, \eta_j, \zeta_{ij}] = Y_{ij}$.

- For $U_N^{h_1}$, $\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$ if and only if $H(K_{1,2}) = (\xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}) \subseteq H(G)$. Indeed, we have for all $G \subset K_{1,2}$, $\mathbb{E}[h_1(Y_{\{1\},\{1,2\}}) \mid H(G)] = 0$ and $\mathbb{E}[h_1(Y_{\{1\},\{1,2\}}) \mid H(K_{1,2})] = Y_{11}Y_{12}$. Therefore, the only graph $G \subseteq K_{1,2}$ such that $p^G \neq 0$ is $G = K_{1,2}$. Thus, $U_N^{h_1}$ is degenerate of order 2 and the family of principal support graphs of $U_N^{h_1}$ is $(K_{\mathbf{i},\mathbf{j}})_{\mathbf{i}\in\mathcal{P}_1([\![m_N]\!]),\mathbf{j}\in\mathcal{P}_2([\![n_N]\!])}$ (Fig. 4.3).
- For $U_N^{h_2}$, $\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$ if and only if $(\xi_1, \xi_2, \eta_1, \eta_2, \zeta_{11}, \zeta_{22}) \subseteq H(G)$ or $(\xi_1, \xi_2, \eta_1, \eta_2, \zeta_{12}, \zeta_{21}) \subseteq H(G)$. Therefore, if $\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$, then $v_1(G) = 2$ or $v_2(G) = 2$, so $U_N^{h_2}$ is degenerate of order 3. The principal support graphs are the graphs which are isomorphic to one graph $G \subseteq \Gamma_{2,2}$ such that $\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$ (Fig. 4.3).

From Proposition 4.2.3,

$$\mathbb{V}[U_N] = \sum_{(0,0)<(r,c)\le(p,q)} \frac{(m_N - r)!}{m_N!} \frac{(n_N - c)!}{n_N!} V^{(r,c)}$$

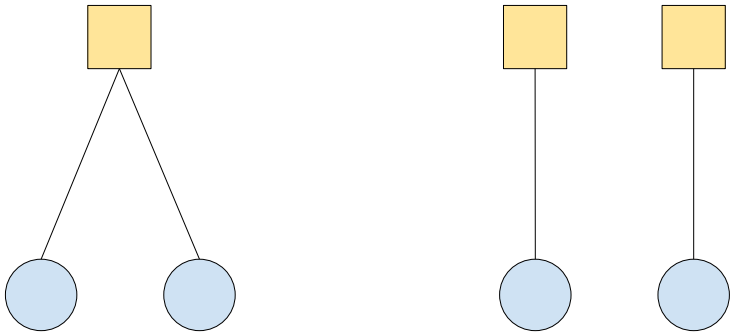

Figure 4.3 – Examples of principal support graphs for $U_N^{h_1}$ (left) and $U_N^{h_2}$ (right). The principal support graphs of $U_N^{h_1}$ are the graphs that are isomorphic to the left one. The principal support graphs of $U_N^{h_2}$ are the 2×2 graphs containing graphs that are isomorphic to the right one.

We see that $\mathbb{V}[U_N]$ is the sum of the $p \times q$ terms of the form $\frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!}V^{(r,c)}$. Each term behaves like $\frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!}V^{(r,c)} \asymp N^{-r-c}$. If for some (r,c), $\sum_{\substack{G \in K_{p,q} \\ (v_1(G),v_2(G))=(r,c)}} p^G = 0$, then $V^{(r,c)} = 0$. Therefore,

$$\begin{aligned}\mathbb{V}[U_N] &= N^{-d} \sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ r+c=d}} \rho^{-r}(1-\rho)^{-c}V^{(r,c)} + o(N^{-d}) \\ &= N^{-d}\sum_{r=0}^{d} \rho^{-r}(1-\rho)^{-d-r}V^{(r,d-r)} + o(N^{-d})\end{aligned}$$

This is a hint that the right normalization for the convergence in distribution of U_N is given by its principal degrees. The following theorem, proven in Appendix 4.C, confirms it.

Theorem 4.2.7. *There is a random variable W such that $N^{d/2}(U_N - p^{\varnothing}) \xrightarrow{\mathcal{D}} W$ if and only if $N^{d/2}\sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ r+c=d}} P_N^{r,c} \xrightarrow{\mathcal{D}} W$.*

This theorem says that the limit distribution of $U_N - p^{\varnothing}$ renormalized by $N^{d/2}$ is the same as that of its principal part $\sum_{\substack{(0,0)\leq(r,c)\leq(p,q) \\ r+c=d}} P_N^{r,c}$, renormalized by the same quantity. Therefore, we shall investigate the asymptotic behaviour of U_N by studying its principal part. Next, we identify the limit distribution for the $N^{d/2}(U_N - p^{\varnothing})$ using the properties of the principal part. In particular, we will see that the limit distribution depends more precisely on the form of the principal support graphs of U_N.

4.2.7. Gaussian case

Theorem 4.2.8. *If all principal support graphs of U_N are connected, then*

$$N^{d/2}(U_N - p^{\varnothing}) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \sum_{\substack{(0,0)<(r,c)\leq(p,q)\\ r+c=d}} \rho^{-r}(1-\rho)^{-c}V^{(r,c)}.$$

The proof of this theorem uses the fact that $N^{d/2}(U_N - p^{\varnothing})$ has the same limit as $N^{d/2}\sum_{\substack{(0,0)<(r,c)\leq(p,q)\\ r+c=d}} P_N^{r,c}$, where

$$P_N^{r,c} = \sum_{G\in\Gamma_{r,c}} \frac{1}{(p-r)!(q-c)!|\mathrm{Aut}(G)|}\tilde{P}_N^G.$$

The convergence of the terms $N^{d/2}\tilde{P}_N^G$ is proved by the methods of moments (Lem. 4.2.10). The calculation of the moments involve sums of terms of the form $\mathbb{E}[\prod_{k=1}^K p^{G_k}]$, the values of which depend on the configuration of the sequence of graphs $G_1,...,G_k$ (Lem. 4.2.9). Therefore, the moments are obtained by counting the frequency of the relevant configurations in these sums. Below, Lemmas 4.2.9 and 4.2.10 are established before completing the proof of Theorem 4.2.8.

Lemma 4.2.9. *Let $G_1,...,G_K$ be subgraphs of K_{m_N,n_N}. If $\mathbb{E}[\prod_{k=1}^K p^{G_k}] \neq 0$, then for all G_k, $1 \leq k \leq K$, each vertex of $V_1(G_k)$ or $V_2(G_k)$ or edge of $E(G_k)$ must also appear in another G_ℓ, $\ell \neq k$.*

Furthermore, if $G_1,...,G_K$ are connected and non-empty, then either $G_1,...,G_K$ coincide in $K/2$ pairs (and K is necessarily even), or some vertex belongs to at least three of them.

Proof. For some $\ell \in [\![K]\!]$, denote $G_{1:k}^{(-\ell)} = \cup_{\substack{i=1\\ i\neq\ell}}^k G_i$. We have

$$\begin{aligned}\mathbb{E}[\prod_{k=1}^K p^{G_k}] &= \mathbb{E}[\mathbb{E}[\prod_{k=1}^K p^{G_k} \mid H(G_{1:K}^{(-\ell)})]]\\ &= \prod_{\substack{k=1\\ k\neq\ell}}^K p^{G_k}\mathbb{E}[\mathbb{E}[p^{G_\ell} \mid H(G_{1:K}^{(-\ell)})]]\\ &= \prod_{\substack{k=1\\ k\neq\ell}}^K p^{G_k}\mathbb{E}[\mathbb{E}[p^{G_\ell} \mid H(G_\ell \cap G_{1:K}^{(-\ell)})]].\end{aligned} \tag{4.7}$$

Suppose there is a vertex or edge of a G_ℓ that does not belong to any other G_k, $k \neq \ell$. In this case, $G_\ell \cap G_{1:K}^{(-\ell)} \subset G_\ell$, so $\mathbb{E}[p^{G_\ell} \mid H(G_\ell \cap G_{1:K}^{(-\ell)})] = 0$, which proves the first result.

From that result, if $\mathbb{E}[\prod_{k=1}^K p^{G_k}] \neq 0$ and no vertex belongs to more than two of $G_1,...,G_K$, then each vertex and edge belongs to exactly two of them. That also means that every connected component must belong to exactly two of them. Therefore, if all graphs are connected, then these graphs coincide in pairs. □

Lemma 4.2.10. *Let $(G_k)_{1\le k\le K}$ be a sequence of distinct connected graphs of $\Gamma^-_{p,q}$, with $v_1(G_k) = r_k$ and $v_2(G_k) = c_k$ for $1 \le k \le K$. We have that*

$$(m_N^{r_k/2} n_N^{c_k/2} \tilde{P}_N^{G_k})_{1\le k\le K} \xrightarrow{\mathcal{D}} (W_k)_{1\le k\le K}, \tag{4.8}$$

where W_k are independent variables with respective distribution $\mathcal{N}(0, p!^2 q!^2 |\mathrm{Aut}(G_k)| \mathbb{E}[(p^{G_k})^2])$.

Proof. Let $\alpha = (\alpha_1, \alpha_2)$ where α_1 is a mapping from $[\![p]\!]$ to $[\![m_N]\!]$ and α_2 is a mapping from $[\![q]\!]$ to $[\![n_N]\!]$. Denote $\alpha(G)$

Let a_k be nonnegative integers. For all $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p([\![m_N]\!]) \times \mathcal{P}_q([\![n_N]\!])$, let $G_{k,\mathbf{i},\mathbf{j}}$ be a graph of $K_{\mathbf{i},\mathbf{j}}$ which is isomorphic to G_k.

$$\mathbb{E}\Big[\prod_{k=1}^K (m_N^{r_k/2} n_N^{c_k/2} \tilde{P}_N^{G_k})^{a_k}\Big]$$
$$= m_N^{\sum_{k=1}^K a_k r_k/2} \binom{m_N}{p}^{-\sum_{k=1}^K a_k} n_N^{\sum_{k=1}^K a_k c_k/2} \binom{n_N}{q}^{-\sum_{k=1}^K a_k} \mathbb{E}\left[\prod_{k=1}^K \Big(\sum_{\substack{\mathbf{i}_k\in\mathcal{P}_p([\![m_N]\!])\\ \mathbf{j}_k\in\mathcal{P}_q([\![n_N]\!])}} \tilde{p}_{\mathbf{i}_k,\mathbf{j}_k}^{G_k}\Big)^{a_k}\right]$$

where we can develop

$$\mathbb{E}\left[\prod_{k=1}^K \Big(\sum_{\substack{\mathbf{i}_k\in\mathcal{P}_p([\![m_N]\!])\\ \mathbf{j}_k\in\mathcal{P}_q([\![n_N]\!])}} \tilde{p}_{\mathbf{i}_k,\mathbf{j}_k}^{G_k}\Big)^{a_k}\right] = \sum_{\substack{\mathbf{i}_k^\ell\in\mathcal{P}_p([\![m_N]\!])\\ \mathbf{j}_k^\ell\in\mathcal{P}_q([\![n_N]\!])}} \mathbb{E}\left[\prod_{k=1}^K \prod_{\ell=1}^{a_k} \tilde{p}_{\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}^{G_k}\right]$$
$$= \sum_{\substack{\mathbf{i}_k^\ell\in\mathcal{P}_p([\![m_N]\!])\\ \mathbf{j}_k^\ell\in\mathcal{P}_q([\![n_N]\!])}} \sum_{\Phi_\ell^k\in\mathbb{S}_p\times\mathbb{S}_q} \mathbb{E}\left[\prod_{k=1}^K \prod_{\ell=1}^{a_k} p^{\Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}}\right]$$

Lemma 4.2.9 states that $\mathbb{E}[\prod_{k=1}^K \prod_{\ell=1}^{a_k} p^{\Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}}] \neq 0$ if and only if either all the $\Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}$ coincide in pairs (and only in pairs), or no vertex appears in exactly one of these graphs and at least one vertex appears in at least three.

In the second case, assume without loss of generality that a row node appears in three graphs. Then $G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)} := \cup_{k=1}^K \cup_{j=1}^{a_k} \Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}$ has $v_1(G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)})$ row nodes and $v_2(G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)})$ column nodes, where $\max r_k \le v_1(G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)}) \le \sum_{k=1}^K a_k r_k/2 - 1$ and $\max c_k \le v_2(G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)}) \le \sum_{k=1}^K a_k c_k/2 - 1$ (we have necessarily $\max r_k \le \sum_{k=1}^K a_k r_k/2 - 1$ and $\max c_k \le \sum_{k=1}^K a_k c_k/2 - 1$, else $\mathbb{E}[\prod_{k=1}^K \prod_{\ell=1}^{a_k} p^{\Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}}] = 0$).

Let $(\max r_k, \max c_k) \le (r*, c*) \le (p,q)$. Count the number of terms of the sum such that $v_1(G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)}) = r^*$ and $v_2(G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)}) = c^*$. There are exactly $\binom{m_N}{r^*}\binom{n_N}{c^*}$ ways to pick r^* row nodes and c^* nodes for $G^*_{(\mathbf{i}_k^\ell),(\mathbf{j}_k^\ell)}$. Now, for a specific set of r^* row nodes and c^* column nodes, for each $1 \le k \le K$, $1 \le \ell \le a_k$, there are $\binom{r^*}{r_k}\binom{c^*}{c_k}\binom{m_N-r^*}{p-r_k}\binom{n_N-c^*}{q-c_k}$ ways to pick $(\mathbf{i}_k^\ell, \mathbf{j}_k^\ell)$ such that the nodes of $G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}$ are contained in the r^* specific row nodes and c^* specific column nodes. Therefore, there are at most $p!q!\binom{r^*}{r_k}\binom{c^*}{c_k}\binom{m_N-r^*}{p-r_k}\binom{n_N-c^*}{q-c_k}$ picks for $(\mathbf{i}_k^\ell, \mathbf{j}_k^\ell)$ and Φ_k^ℓ. Finally, the number of terms is least than

$$\begin{aligned}
B_N^{r^*,c^*} &:= \binom{m_N}{r^*}\binom{n_N}{c^*}\prod_{k=1}^K\prod_{\ell=1}^{a_k} p!q!\binom{r^*}{r_k}\binom{m_N-r^*}{p-r_k}\binom{n_N-c^*}{q-c_k} \\
&= \binom{m_N}{r^*}\binom{n_N}{c^*}\prod_{k=1}^K\left[p!q!\binom{r^*}{r_k}\binom{m_N-r^*}{p-r_k}\binom{n_N-c^*}{q-c_k}\right]^{a_k} \\
&= O\left(m_N^{r^*}n_N^{c^*}\prod_{k=1}^K\left[m_N^{p-r_k}n_N^{q-c_k}\right]^{a_k}\right) \\
&= O\left(m_N^{r^*+\sum_{k=1}^K a_k(p-r_k)}n_N^{c^*+\sum_{k=1}^K a_k(q-c_k)}\right).
\end{aligned}$$

Now the total number of these terms is

$$\begin{aligned}
B_N &\le \sum_{(\max r_k,\max c_k)\le(r^*,c^*)\le(\sum_{k=1}^K a_k r_k/2-1,\sum_{k=1}^K a_k c_k/2)} B_N^{r^*,c^*} \\
&= O(B_N^{\sum_{k=1}^K a_k r_k/2-1,\sum_{k=1}^K a_k c_k/2}) \\
&= O\left(m_N^{\sum_{k=1}^K a_k(p-r_k/2)-1}n_N^{\sum_{k=1}^K a_k(q-c_k/2)}\right) \\
&= o\left(m_N^{\sum_{k=1}^K a_k(p-r_k/2)}n_N^{\sum_{k=1}^K a_k(q-c_k/2)}\right).
\end{aligned}$$

We notice that the contribution of these terms are $o(1)$ in equation (4.2.7).

Now, there remains the terms of the first case, where the $\Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}$ coincide in pairs. Note that since the G_k are non-isomorphic, only graphs arising for the permutations of a same graph G_k can coincide. Therefore, the a_k are necessarily even. Furthermore, for each k, there are $a_k/2$ different pairs of coinciding graphs $\Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}$. There are $\frac{a_k!}{2^{a_k/2}(a_k/2)!}$ ways to partition a set of a_k graphs into $a_k/2$ pairs.

Fix k, ℓ_1, ℓ_2. The number of picks for $\mathbf{i}_k^{\ell_1}, \mathbf{j}_k^{\ell_1}, \mathbf{i}_k^{\ell_2}, \mathbf{j}_k^{\ell_2}, \Phi^{\ell_1}, \Phi^{\ell_2}$ such that $\Phi_k^{\ell_1}G_{k,\mathbf{i}_k^{\ell_1},\mathbf{j}_k^{\ell_1}} = \Phi_k^{\ell_2}G_{k,\mathbf{i}_k^{\ell_2},\mathbf{j}_k^{\ell_2}}$ is given by Lemma 4.2.6. Accounting for all $a_k/2$ pairs of the type (ℓ_1,ℓ_2), there are

$$\frac{m_N!(m_N-r_k)!}{(m_N-p)!^2}\frac{n_N!(n_N-c_k)!}{(n_N-q)!^2}|\mathrm{Aut}(G)|.$$

Therefore, taking into account the number of possible pairings and the picks for all $1 \le k \le K$,

$1 \le \ell \le a_k$, there are

$$\begin{aligned} A_N &= \prod_{k=1}^{K} \frac{a_k!}{2^{a_k/2}(a_k/2)!} \left(\frac{m_N!(m_N - r_k)!}{(m_N - p)!^2} \frac{n_N!(n_N - c_k)!}{(n_N - q)!^2} |\mathrm{Aut}(G_k)| \right)^{a_k/2} \\ &= m_N^{\sum_{k=1}^K a_k(r_k/2-p)} n_N^{\sum_{k=1}^K a_k(c_k/2-q)} \prod_{k=1}^{K} \frac{a_k!}{2^{a_k/2}(a_k/2)!} |\mathrm{Aut}(G_k)|^{a_k/2} \\ &\quad + o\left(m_N^{\sum_{k=1}^K a_k(r_k/2-p)} n_N^{\sum_{k=1}^K a_k(c_k/2-q)} \right). \end{aligned}$$

Each of these A_N terms is equal to $\mathbb{E}\left[\prod_{k=1}^K \prod_{\ell=1}^{a_k} p^{\Phi_k^\ell G_{k,\mathbf{i}_k^\ell,\mathbf{j}_k^\ell}} \right] = \prod_{k=1}^K \mathbb{E}[(p^{G_k})^2]^{a_k/2}$.

In conclusion, if all the a_k are even, then

$$\begin{aligned} \mathbb{E}\left[\prod_{k=1}^{K} (m_N^{r_k/2} n_N^{c_k/2} \tilde{P}_N^{G_k})^{a_k} \right] &= m_N^{\sum_{k=1}^K a_k r_k/2} \binom{m_N}{p}^{-\sum_{k=1}^K a_k} n_N^{\sum_{k=1}^K a_k c_k/2} \binom{n_N}{q}^{-\sum_{k=1}^K a_k} \\ &\quad \times A_N \prod_{k=1}^{K} \mathbb{E}[(p^{G_k})^2]^{a_k/2} \\ &= (p!q!)^{\sum_{k=1}^K a_k} \prod_{k=1}^{K} \frac{a_k!}{2^{a_k/2}(a_k/2)!} |\mathrm{Aut}(G_k)|^{a_k/2} \mathbb{E}[(p^{G_k})^2]^{a_k/2} \\ &= \prod_{k=1}^{K} \frac{a_k!}{2^{a_k/2}(a_k/2)!} \left(p!^2 q!^2 |\mathrm{Aut}(G_k)| \mathbb{E}[(p^{G_k})^2] \right)^{a_k/2}, \end{aligned}$$

and in the general case,

$$\begin{aligned} &\mathbb{E}\Big[\prod_{k=1}^{K} (m_N^{r_k/2} n_N^{c_k/2} \tilde{P}_N^{G_k})^{a_k} \Big] \\ &\quad = \begin{cases} \prod_{k=1}^K \frac{a_k!}{2^{a_k/2}(a_k/2)!} \left(p!^2 q!^2 |\mathrm{Aut}(G_k)| \mathbb{E}[(p^{G_k})^2] \right)^{a_k/2} & \text{if all } a_k \text{ are even} \\ 0 & \text{if at least one } a_k \text{ is odd} \end{cases}. \end{aligned} \tag{4.9}$$

Else, if there is at least one odd a_k, we have $\mathbb{E}[\prod_{k=1}^K (m_N^{r_k/2} n_N^{c_k/2} \tilde{P}_N^{G_k})^{a_k}] = 0$.

We remind that the moment of order a of a gaussian variable X with mean 0 and variance σ^2 is

$$\mathbb{E}[X^a] = \begin{cases} \frac{a!}{2^{a/2}(a/2)!} \sigma^a & \text{if } a \text{ is even} \\ 0 & \text{if } a \text{ is odd} \end{cases}.$$

So the application of the methods of moments to equation (4.9) concludes the proof of this lemma. □

Proof of Theorem 4.2.8. Theorem 4.2.7 states that $N^{d/2}(U_N - p^{\varnothing})$ has the same limit as $N^{d/2} \sum_{\substack{(0,0)<(r,c)\le(p,q) \\ r+c=d}} P_N^{r,c}$.

For all $(0,0) < (r,c) \leq (p,q)$,

$$P_N^{r,c} = \sum_{G \in \Gamma_{r,c}} \frac{1}{(p-r)!(q-c)!|\mathrm{Aut}(G)|} \tilde{P}_N^G.$$

So

$$N^{d/2} \sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ r+c=d}} P_N^{r,c} = \sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ r+c=d}} N^{d/2} m_N^{-r/2} n_N^{-c/2} \sum_{G \in \Gamma_{r,c}} \frac{m_N^{r/2} n_N^{c/2} \tilde{P}_N^G}{(p-r)!(q-c)!|\mathrm{Aut}(G)|}.$$

By construction, $N^{d/2} m_N^{-r/2} n_N^{-c/2} \xrightarrow[N\to\infty]{} \rho^{-r/2}(1-\rho)^{-c/2}$. Therefore, by Lemma 4.2.10, $N^{d/2} \sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ r+c=d}} P_N^{r,c}$ converges in distribution to

$$Z = \sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ r+c=d}} \rho^{-r/2}(1-\rho)^{-c/2} \sum_{G \in \Gamma_{r,c}} W_G,$$

where for all (r,c), $G \in \Gamma_{r,c}$, W_G are independent gaussian variables with mean 0 and variance $\frac{p!^2 q!^2}{(p-r)!^2(q-c)!^2|\mathrm{Aut}(G)|} \mathbb{E}[(p^G)^2]$.

Finally, it follows that Z is a gaussian variable with mean 0 and variance $\sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ r+c=d}} \rho^{-r}(1-\rho)^{-c} V^{(r,c)}$ where

$$V^{(r,c)} = \frac{p!^2 q!^2}{(p-r)!^2(q-c)!^2} \sum_{G \in \Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1} \mathbb{E}[(p^G)^2].$$

□

Example (continued). Let Y be a random matrix such that $Y_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Let h_1 be the kernel function defined by $h_1(Y_{\{1\},\{1,2\}}) = Y_{11}Y_{12}$ and $U_N^{h_1}$ the U-statistic associated to this kernel. In Section 4.2.6, we have seen that $U_N^{h_1}$ is degenerate of order 2 and the family of principal support graphs of $U_N^{h_1}$ is $(K_{\mathbf{i},\mathbf{j}})_{\mathbf{i} \in \mathcal{P}_1([\![m_N]\!]), \mathbf{j} \in \mathcal{P}_2([\![n_N]\!])}$, which are all connected.
Therefore, Theorem 4.2.8 implies

$$N^{3/2} U_N^{h_1} \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma_1^2),$$

where $\sigma_1^2 = V^{(1,2)} = \frac{4}{\rho(1-\rho)^2} |\mathrm{Aut}(K_{1,2})|^{-1} \mathbb{E}[(p^{K_{1,2}})^2] = \frac{4}{\rho(1-\rho)^2} \frac{1}{2} \mathbb{E}[Y_{11}^2 Y_{12}^2] = \frac{2}{\rho(1-\rho)^2}$.

4.2.8. Other limit distributions

Denote b the maximum number of connected components in the principal support graph. Theorem 4.2.8 ensures that the limit distribution is gaussian when all principal support graphs

are connected, that means when $b = 1$. In the theory of Janson and Nowicki (1991), the form of the limit distribution of generalized U-statistics is is determined by the number b. More precisely, the limit distribution is a polynomial of independent Gaussian variables, of degree b. Although we have not proved a similar result for our U-statistics of RCE matrices, it is natural to wonder if the result of Janson and Nowicki (1991) in the case $b > 1$ can be generalized the same way as in the Gaussian case ($b = 1$). Therefore, we risk a conjecture, as an attempt to transpose Theorem 3 of Janson and Nowicki (1991) to our U-statistics. It will remain unproven in this thesis.

Before giving the conjecture, let us give some definitions and notations. Let F and G be two bipartite graphs, with $V_1(F) = [\![v_1(F)]\!]$, $V_2(F) = [\![v_2(F)]\!]$, $V_1(G) = [\![v_1(G)]\!]$ and $V_2(G) = [\![v_2(G)]\!]$. Let $\alpha_F(G)$ be the transformation of the graph G where an offset of $v_1(F)$ is applied to the row node labels and an offset of $v_2(F)$ is applied to the column node labels, i.e.

- $V_1(\alpha_F(G)) = \{v_1(F) + 1, ..., v_1(F) + v_1(G)\}$,
- $V_2(\alpha_F(G)) = \{v_2(F) + 1, ..., v_2(F) + v_2(G)\}$,
- $E(\alpha_F(G)) = \{(v_1(F) + x, v_2(F) + y) : (x, y) \in E(G)\}$.

In other words, α_F defines a new numbering for the nodes and edges of G such that the nodes of G are labeled using numbers larger than those used of nodes of F. We define the disjoint union $F \oplus G := F \cup \alpha_F(G)$.

Let $f_1 \in L_2(K_{r_1,c_1})$ and $f_2 \in L_2(K_{r_2,c_2})$. We define $f_1 \otimes f_2 \in L_2(K_{r_1+r_2,c_1+c_2})$ by

$$\begin{aligned}(f_1 \otimes f_2)(H(K_{r_1+r_2,c_1+c_2})) &= (f_1 \otimes f_2)((\xi_i)_{1\le i\le r_1+r_2}, (\eta_j)_{1\le j\le c_1+c_2}, (\zeta_{ij})_{1\le i\le r_1+r_2, 1\le j\le c_1+c_2}) \\ &= f_1((\xi_i)_{1\le i\le r_1}, (\eta_j)_{1\le j\le c_1}, (\zeta_{ij})_{1\le i\le r_1, 1\le j\le c_1}) \\ &\quad \times f_2((\xi_{r_1+i})_{1\le i\le r_2}, (\eta_{c_1+j})_{1\le j\le c_2}, (\zeta_{r_1+i,c_1+j})_{1\le i\le r_2, 1\le j\le c_2}) \\ &= f_1(H(K_{r_1,c_1})) \times f_2(H(\alpha_{K_{r_1,c_1}}(K_{r_2,c_2}))).\end{aligned}$$

In the previous sections, we have not given a precise form for $\Gamma_{r,c}$. We can, without loss of generality, suppose that all the graphs $G \in \Gamma_{r,c}$ which are disconnected with s components are chosen of the form $G_1 \oplus ... \oplus G_s$, where $G_t \in \bar{\Gamma}_{r_t,c_t}$, the complement of Γ_{r_t,c_t} in the set of subgraphs of K_{r_t,c_t} with r_t column row and c_t column nodes, and (r_t, c_t) are the dimensions of G_t, for each $1 \le t \le s$. For a graph $G = G_1 \oplus ... \oplus G_s$, if $f_t \in L_2^*(G_t)$ for each $1 \le t \le s$, then $f_1 \otimes ... \otimes f_s \in L_2^*(G)$. Therefore, if $(e_k^{G_t})_{k\ge1}$ is a orthonormal basis for $L_2^*(G_t)$ for each $1 \le t \le s$, then $(e_{k_1}^{G_s} \otimes ... \otimes e_{k_s}^{G_s})_{(k_1,...,k_s)\in\mathbb{N}^s}$ is a orthonormal basis for $L_2^*(G)$.

Conjecture 4.2.11.

$$N^{d/2}(U_N - p^{\varnothing}) \xrightarrow[N\to\infty]{\mathcal{D}} W, \tag{4.10}$$

for some random variable W. This random variable can be written

$$W = \sum_{\substack{(0,0)<(r,c)\leq(p,q)\\ r+c=d}} \frac{p!q!}{(p-r)!(q-c)!} \sum_{s=1}^{d} \sum_{\substack{(G_1,\dots,G_s)\in\bar{\Gamma}_{r_1,c_1}\times\dots\times\bar{\Gamma}_{r_s,c_s}\\ r_1+\dots+r_s=r, c_1+\dots+c_s=c}} \sum_{\mathbf{t}\in\mathbb{N}^s} \frac{1}{s!} \prod_{k=1}^{s} |\mathrm{Aut}(G_k)|^{-1} \times \left\langle h, e_{t_1}^{G_1}\otimes\dots\otimes e_{t_s}^{G_s}\right\rangle \prod_{\ell=1}^{\infty} H_{\kappa_\ell(\mathbf{t})}(W_{t_\ell}^{G_\ell}), \tag{4.11}$$

where

- *for all G, $(e_t^G)_{t\geq 1}$ is an orthonormal basis of the subspace of symmetric functions in $L_2^*(G)$,*
- *$(W_t^G)_{G,t\geq 1}$ are independent standard normal variables,*
- *H_u is the u-th Hermite polynomial, for $u\geq 0$,*
- *$\kappa_\ell(\mathbf{t})$ is the number of times ℓ appears as an element in the d-tuple $\mathbf{t}$.*

Remark 3. In the Theorem 3 of Janson and Nowicki (1991), the polynomials $\prod_{\ell=1}^{\infty} H_{\kappa_\ell(\mathbf{t})}(W_{t_\ell}^{G_\ell})$ are replaced by Wick products $:W_{t_1}^{G_1}\dots W_{t_s}^{G_s}:$. These quantities are in fact equal. Indeed, the Wick product between s random variables $(X_1,\dots,X_s)$ can be given by the following formula

$$:X_1\dots X_s: = \sum(-1)^{\ell}\prod_{j=1}^{\ell}\left\langle X_{a(j)}, X_{b(j)}\right\rangle \prod_{r\in A} X_r,$$

where the sum is over all subsets $A\subseteq[\![s]\!]$ and partitions of $[\![s]\!]\backslash A$ into ℓ pairs $\{a(j),b(j)\}$. For all integers $a_1,\dots,a_s$, we have $:X_1^{a_1}\dots X_s^{a_s}: = \prod_{t=1}^{s} H_{a_t}(X_s)$ when $X_1,\dots,X_s$ are independent standard normal variables (see for example Major, 1981). This proves that

$$:W_{t_1}^{G_1}\dots W_{t_s}^{G_s}: = \prod_{\ell=1}^{\infty} H_{\kappa_\ell(\mathbf{t})}(W_{t_\ell}^{G_\ell}).$$

Remark 4. The kernel of our U-statistics is a function of the AHK variables $h_\phi((\xi_i)_{i\in\mathbf{i}},(\eta_j)_{j\in\mathbf{j}},(\zeta_{ij})_{i\in\mathbf{i},j\in\mathbf{j}})$ where $|\mathbf{i}|=p$ and $|\mathbf{j}|=q$. The case of U-statistics of i.i.d. observations is obtained when setting $q=0$. As a consequence, Theorem 1.5.13, giving the asymptotic distribution for degenerate U-statistics of i.i.d. observations, can be retrieved from Theorem 4.2.11 by forcing $q=0$. It follows that $c=0$ and $r=s=d$. Necessarily, $r_1=\dots=r_s=1$, $c_1=\dots=c_s=0$ and the graphs in $(G_1,\dots,G_s)$ are each reduced to a single vertex, i.e. they are equal to $K_{1,0}$. This implies that we only need one basis $(e_t)_{t\geq 1}$ of $L_2^*(K_{1,0})$. In this case, the limit becomes

$$W = \binom{p}{d}\sum_{\mathbf{t}\in\mathbb{N}^s}\langle h, e_{t_1}\otimes\dots\otimes e_{t_s}\rangle\prod_{\ell=1}^{\infty} H_{\kappa_\ell(\mathbf{t})}(W_{t_\ell}). \tag{4.12}$$

This expression is equivalent to the expression of the limit in Theorem 1.5.13 by noticing that because $e_{t_1}\otimes\dots\otimes e_{t_s}\in L_2^*(K_{d,0})$, we have $\langle h, e_{t_1}\otimes\dots\otimes e_{t_s}\rangle = \langle h^{(d)}, e_{t_1}\otimes\dots\otimes e_{t_s}\rangle$, where $h^{(d)}$ is the projection of h on $L_2^*(K_{d,0})$.

Example (continued). Let Y be a random matrix such that $Y_{ij}\overset{i.i.d.}{\sim}\mathcal{N}(0,1)$. Let h_2 be the kernel function defined by $h_2(Y_{\{1,2\},\{1,2\}}) = (Y_{11}Y_{22}+Y_{12}Y_{21})/2$ and $U_N^{h_2}$ the U-statistic associated to

this kernel.

Let G_1 and G_2 be the graphs defined by $H(G_1) = (\xi_1, \xi_2, \eta_1, \eta_2, \zeta_{11}, \zeta_{22})$ and $H(G_2) = (\xi_1, \xi_2, \eta_1, \eta_2, \zeta_{12}, \zeta_{21})$. In Section 4.2.6, we have seen that $\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$ if and only if $G_1 \subseteq G$ or $G_2 \subseteq G$. Therefore, G_1 and G_2 are principal support graphs with 2 components and the other principal support graphs of $U_N^{h_2}$ are connected. Thus, if true, Conjecture 4.2.11 would imply

$$N^2 U^{h_2} \xrightarrow[N\to\infty]{\mathcal{D}} W,$$

where W is some polynomial function of independent Gaussian variables of degree 2.

4.2.9. Other asymptotic frameworks

In previous sections, we have assumed that $m_N + n_N = N$ and $m_N/N \to \rho \in]0,1[$. It is in fact possible to extend all our results to any asymptotic behavior. In this section, let us only assume that $m_N \xrightarrow[N\to\infty]{} \infty$ and $n_N \xrightarrow[N\to\infty]{} \infty$ and see how it affects our results.

The principal part of U_N should be the dominant part of the variance. Remember that Proposition 4.2.3 states that

$$\mathbb{V}[U_N] = \sum_{(0,0)<(r,c)\leq(p,q)} \frac{(m_N - r)!}{m_N!} \frac{(n_N - c)!}{n_N!} V^{(r,c)}.$$

We see that $\mathbb{V}[U_N]$ is the sum of the $p \times q$ terms of the form $\frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!}V^{(r,c)}$. Each term behaves like $\frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!}V^{(r,c)} \asymp m_N^{-r} n_N^{-c}$. The dominant part of $\mathbb{V}[U_N]$ is consist of the terms $m_N^{-r} n_N^{-c}$ decreasing the slowest such that $V^{(r,c)} \neq 0$.

There is no equivalent to the previously defined degree of degeneracy, but we can redefine principal degrees. Let the family of couples $((r_\ell, c_\ell))_{1\leq\ell\leq L}$ be such that $m_N^{r_1} n_N^{c_1} \asymp \ldots \asymp m_N^{r_L} n_N^{c_L}$ and $\mathbb{V}[U_N] \asymp \sum_{\ell=1}^{L} \frac{V^{(r_\ell,c_\ell)}}{m_N^{r_\ell} n_N^{c_\ell}}$. We can call these couples the *principal degrees* of U_N, by analogy with the previous case. The quantity $\sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell}$ is called the *principal part* of U_N. We call the *principal support graphs* of U_N the graphs G such that

- $(v_1(G), v_2(G)) \in \{(r_\ell, c_\ell) : 1 \leq \ell \leq L\}$,
- $p^G \neq 0$.

Example 2. Suppose $(m_N, n_N) = (N, \sqrt{N})$ and $V^{(0,1)} = 0$ but $V^{(0,2)} \neq 0$ and $V^{(1,0)} \neq 0$, then the principal degrees are $(1,0)$ and $(0,2)$ because $m_N = n_N^2 = N$ and $\mathbb{V}[U_N] = N^{-1}(V^{(1,0)} + V^{(0,2)})$. In this case, one valid choice of $\gamma(N)$ is $\gamma(N) = N$.

Example 3. Suppose again that $(m_N, n_N) = (N, \sqrt{N})$, but this time $V^{(0,1)} = V^{(0,2)} = V^{(1,0)} = 0$. If $V^{(1,1)} \neq 0$ and $V^{(0,3)} \neq 0$, then the principal degrees are $(1,1)$ and $(0,3)$ because $m_N n_N = n_N^3 = N^{3/2}$. In this case, one valid choice of $\gamma(N)$ is $\gamma(N) = N^{3/2}$.

In this asymptotic framework, there is no reason that $N^{d/2}$ is the right normalization for the weak convergence of U-statistics. If the elements of $((r_\ell, c_\ell))_{1\leq\ell\leq L}$ are the principal degrees of U_N, then there is a function γ such that $m_N^{-r_\ell} n_N^{-c_\ell}\gamma(N) \xrightarrow[N\to\infty]{} \alpha_\ell$, where $\alpha_\ell > 0$ for all $1 \leq \ell \leq L$ and $\gamma(N)\mathbb{V}[U_N] = \sum_{1\leq\ell\leq L} \alpha_\ell V^{(r_\ell,c_\ell)} + o(1)$. Next, we state the equivalent result to Theorem 4.2.7 in the new framework. The proof for this theorem is given in 4.D.1.

Theorem 4.2.12. *There is a random variable W such that $\sqrt{\gamma(N)}\sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell} \xrightarrow{\mathcal{D}} W$ if and only if $\sqrt{\gamma(N)}(U_N - p^\varnothing) \xrightarrow{\mathcal{D}} W$.*

This theorem says that the limit distribution of $U_N - p^\varnothing$ renormalized by $\sqrt{\gamma(N)}$ is the same as that of its principal part $\sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell}$, renormalized by the same quantity. Therefore, similar as in the initial framework, we shall investigate the asymptotic behaviour of U_N by studying its principal part.

In practice, one has to identify the principal part by finding the principal degrees of U_N. The principal degrees depend both on the kernel h and the asymptotic behaviour of (m_N, n_N). After finding the principal degrees, then a function $\gamma(N)$ can be found. With $\gamma(N)$ and the principal degrees, the coefficients α_ℓ can be calculated to yield an expression for the variance. We will illustrate this in examples later.

Now, we derive the equivalent to Theorem 4.2.8, i.e. the convergence result when the principal support graphs of U_N are connected. The proof of this theorem is given in Appendix 4.D.2.

Theorem 4.2.13. *If all principal support graphs of U_N are connected, then*

$$\sqrt{\gamma(N)}(U_N - p^\varnothing) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \sum_{\ell=1}^{L} \alpha_\ell V^{(r_\ell,c_\ell)}.$$

Unsurprisingly, the limit distribution for $\sqrt{\gamma(N)}(U_N - p^\varnothing)$ is still a Gaussian with another expression for the variance. This variance consists of terms associated of the principal degrees of U_N.

4.3. Practical identification of limit distribution of U-statistics

In practice, once the network models and the estimators have been decided, applying the theoretical results is not straightforward. First, we do not generally know *a priori* whether

a U-statistic is degenerate or not. If it is degenerate, the order of degeneracy is also usually unknown. Without any investigation, we do not always know which theorem to apply to obtain the form of the limit distribution. Second, even if some theorem is applied, precisely identifying the limit distribution may be tedious, for example, the asymptotic variance given by the formula in Theorem 4.2.8 is not easy to compute. In this section, we give a few hints and directions to identify the limit distribution of a U-statistic, with emphasis on degenerate cases. We use the asymptotic framework in which $m_N + n_N = N$ and $m_N/N \xrightarrow[N\to\infty]{} \rho \in]0,1[$, but many techniques presented here can be generalized to all frameworks.

4.3.1. Degeneracy

The first step is to determine whether the U-statistic is degenerate or not. This can be done with few analytic calculations, which are fortunately rather simple.

Indeed, if the U-statistic is non-degenerate, then the limit distribution is Gaussian with variance given by Theorem 2.2.7 and Theorem 3.3.1. The two theorems come each with an expression for the asymptotic variance. The two expressions are equivalent. In Theorem 2.2.7, we can transpose the formula to kernels of $p \times q$ submatrices. Therefore, the asymptotic variance is given by

$$V = \frac{p^2}{\rho} c^{1,0} + \frac{q^2}{1-\rho} c^{0,1}, \tag{4.13}$$

where $c^{r,c} = \mathrm{Cov}(h_{\mathbf{i},\mathbf{j}}, h_{\mathbf{i}',\mathbf{j}'})$ where $\mathbf{i}$ and $\mathbf{i}'$ have r elements in common and $\mathbf{j}$ and $\mathbf{j}'$ have c elements in common. In Theorem 3.3.1, the asymptotic variance is given by the variance of the first Hoeffding-type projections, that is

$$V = \frac{p^2}{\rho} v^{1,0} + \frac{q^2}{1-\rho} v^{0,1}, \tag{4.14}$$

where $v^{1,0} = \mathbb{V}[\mathbb{E}[h_{[\![p]\!],[\![q]\!]} \mid \xi_1]]$ and $v^{0,1} = \mathbb{V}[\mathbb{E}[h_{[\![p]\!],[\![q]\!]} \mid \eta_1]]$. Remark that

$$\begin{aligned}
&\mathbb{E}[h_{\{1,\dots,p\},\{1,\dots,q\}} h_{\{1,p+1,\dots,2p-1\},\{q+1,\dots,2q\}}] \\
&\qquad = \mathbb{E}[\mathbb{E}[h_{\{1,\dots,p\},\{1,\dots,q\}} h_{\{1,p+1,\dots,2p-1\},\{q+1,\dots,2q\}} \mid \xi_1]] \\
&\qquad = \mathbb{E}[\mathbb{E}[h_{\{1,\dots,p\},\{1,\dots,q\}} \mid \xi_1] \mathbb{E}[h_{\{1,p+1,\dots,2p-1\},\{q+1,\dots,2q\}} \mid \xi_1]] \\
&\qquad = \mathbb{E}[\mathbb{E}[h_{[\![p]\!],[\![q]\!]} \mid \xi_1]^2].
\end{aligned} \tag{4.15}$$

This proves that $c^{1,0} = v^{1,0}$ and by symmetry, $c^{0,1} = v^{0,1}$.

Any of these two formulas can be used to check the degeneracy, which occurs when this asymptotic variance is equal to 0, i.e. $c^{1,0} = c^{0,1} = 0$ or $v^{1,0} = v^{0,1} = 0$.

4.3.2. Order of degeneracy

Alhough the application of the non-degenerate theorem is rather straightforward after calculating the asymptotic variance, more work is required in the degenerate case. This includes finding the order of degeneracy, which gives the rate of convergence and a list of projections to investigate, and the form of the distribution depending on the principal support graphs.

Although it is possible to find the order of degeneracy and the principal support graphs analytically, it is usually laborious and time-consuming to explore and compute all the required projections, especially when the order of degeneracy is high.

Fortunately, the order of degeneracy can be found by simulation. Below, we suggest two different methods. For the principal support graphs, even if knowing the order of degeneracy shrinks the pool of projections to explore, we have found no general trick to identify them.

Slope method The most intuitive technique is to obtain $\mathbb{V}[U_N]$ for many values of N by simulation. Indeed, $\mathbb{V}[U_N] = V^{(d)}/N^d + o(N^{-d})$, so $\log \mathbb{V}[U_N] = -d \log N + \log(V^{(d)} + o(1))$. Therefore, the slope of the asymptote of the curve representing $\log \mathbb{V}[U_N]$ versus $\log N$ is $-d$. However, this comes with a caveat in practice, as the slope is not always precise for small values of N. One also needs to simulate with different orders of magnitude of N, but simulations are onerous as N grows larger.

The non-degenerate variance estimator method This method exploits an interesting property of the estimator $\widehat{V}_N$ of the non-degenerate asymptotic variance defined in Chapter 3. For non-degenerate U-statistics, we have $N\mathbb{V}[U_N]/\mathbb{E}[\widehat{V}_N] \to 1$. For degenerate U-statistics, the following result holds.

Theorem 4.3.1. *Let U_N be a degenerate U-statistic of order $d-1$, $2 \le d \le p+q-1$. Then, we have*

$$\frac{N\mathbb{V}[U_N]}{\mathbb{E}[\widehat{V}_N]} \xrightarrow[N\to\infty]{} \frac{1}{d}. \tag{4.16}$$

This result, the proof of which is given in Appendix 4.E, is rather surprising. Remember that $\widehat{V}_N$ is the estimator of the asymptotic variance in the non-degenerate case, therefore we have $\widehat{V}_N \xrightarrow[N\to\infty]{\mathbb{P}} 0$. Since $N^{d/2}(U_N - U_\infty) \xrightarrow{\mathcal{D}} W$, where W is a random variable with finite variance, we could expect that $N\mathbb{V}[U_N]/\mathbb{E}[\widehat{V}_N]$ goes to infinity instead. However, it turns out that $\widehat{V}_N$ also captures the variance terms of degenerate U-statistics, which compensate with the N normalization instead of N^d.

Therefore, this method requires to obtain $\mathbb{V}[U_N]$ and $\mathbb{E}[\widehat{V}_N]$ for growing values of N by simulation. This method often finds the order of degeneracy $d-1$ faster than the slope method. Even though the computation of $\widehat{V}_N$ takes $O(N)$ times longer than U_N, because d is obtained as a limit value instead of the slope of an asymptote, it is more precise. The slope method usually requires to simulate networks with larger values of N than this one.

4.4. Example

In this section, we illustrate the example introduced in Section 4.2.6 and continued in following sections. Let Y be a random matrix such that $Y_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Let h_1 and h_2 be the kernel functions defined by $h_1(Y_{\{1\},\{1,2\}}) = Y_{11}Y_{12}$ and $h_2(Y_{\{1,2\},\{1,2\}}) = (Y_{11}Y_{22}+Y_{12}Y_{21})/2$, and $U_N^{h_1}$ and $U_N^{h_2}$ are the U-statistics associated to these kernels.

In Section 4.2.7, we have shown that $U_N^{h_1}$ is degenerate of order 2 and

$$N^{3/2}U_N^{h_1} \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma_1^2),$$

where $\sigma_1^2 = \frac{4}{\rho(1-\rho)^2}$.

In Section 4.2.8, we have shown that $U_N^{h_2}$ is degenerate of order 3 and we have conjectured that

$$N^2U^{h_2} \xrightarrow[N\to\infty]{\mathcal{D}} W,$$

where W is some polynomial function of independent Gaussian variables of degree 2. It is possible to calculate the variance of W, which is $\sigma_2^2 = \rho^{-2}(1-\rho)^{-2}V^{(2,2)}$. However, the calculation of $V^{(2,2)}$ is tedious because one has to explore the principal support graphs of $U_N^{h_2}$. A simpler alternative is to use the Hoeffding-type decomposition of Chapter 3 and use Corollary 3.2.4 stating that

$$\mathbb{V}[U_N^{h_2}] = \sum_{(0,0)<(r,c)\leq(2,2)} \binom{2}{r}^2\binom{2}{c}^2\binom{m_N}{r}^{-1}\binom{n_N}{c}^{-1}\mathbb{V}[p^{r,c}h_2].$$

Since $U_N^{h_2}$ is degenerate of order 3, then $\mathbb{V}[U_N^{h_2}] = \binom{m_N}{2}^{-1}\binom{n_N}{2}^{-1}\mathbb{V}[p^{2,2}h_2] = \frac{4}{N^4\rho^2(1-\rho)^2}\mathbb{V}[h_2(Y_{\{1,2\},\{1,2\}})] + o(N^{-4})$. Since $\mathbb{V}[h_2(Y_{\{1,2\},\{1,2\}})] = \mathbb{E}[(Y_{11}Y_{22}+Y_{12}Y_{21})^2]/4 = \mathbb{E}[Y_{11}^2Y_{22}^2]/2 = 1/2$, we have finally the variance of W,

$$\sigma_2^2 = \frac{2}{\rho^2(1-\rho)^2}.$$

Figures 4.4 and 4.5 present some simulation results. For several values of N, we have simulated $K = 500$ networks and for each network, we have computed $U_N^{h_1}$ and $U_N^{h_2}$. We renormalize them by their respective theoretical asymptotic variance, then we compare their empirical distribution to the standard Gaussian distribution. As expected, the distribution of $N^{3/2}U_N^{h_1}/\sigma_1$

converges to a standard Gaussian distribution. However, the distribution of $N^2 U_N^{h_2}/\sigma_2$ seems to converge to some asymmetric distribution, which is not a Gaussian. It seems to be closer to a centered Chi-squared distribution of the form $W^2 - 1$, where W is a Gaussian standard variable, although this figure is insufficient to illustrate this convergence. Nevertheless, this form was expected because we have conjectured that the limit distribution is a polynomial of independent Gaussians variable of degree 2, which explains the asymmetry.

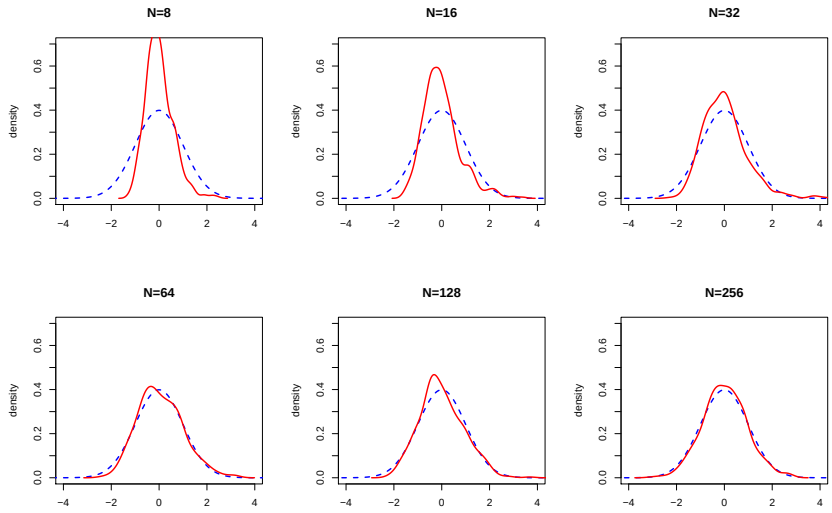

Figure 4.4 – Empirical distribution of $N^{3/2} U_N^{h_1}/\sigma_1$ (solid red), compared to a standard Gaussian distribution (dashed blue). The empirical distribution has been smoothed using the *density* function from the base R stats package.

Remark. This example is basic with a simple i.i.d. model for the network edges, but it captures well the subtleties of the degeneracy phenomenon. Actually, this example is close to a classic example for degenerate U-statistics of i.i.d. observations. Consider that $(X_1, X_2, ...)$ are i.i.d. variables with $\mathbb{E}[X_1] = 0$ and $\mathbb{V}[X_1] = \sigma^2$ and let h be the kernel $h_0(X_1, X_2) = X_1 X_2$. Then, we can show that the corresponding U-statistic $U_N^{h_0} = \binom{N}{2}^{-1} \sum_{1 \le i_1 < i_2 \le N} h_0(X_{i_1}, X_{i_2})$ is degenerate and

$$N U_N^{h_0} \xrightarrow[N \to \infty]{\mathcal{D}} \sigma^2 (W^2 - 1),$$

where W is a standard Gaussian variable.
Now, suppose that the elements $(X_1, X_2, ...)$ are placed in a matrix Y, such that for all N, the submatrix consisting of the first N rows and N columns contains the N^2 first terms of $(X_1, X_2, ...)$. Then, the quantity $U_N^{h_0}$ can be written

$$U_N^{h_0} = \binom{N^2}{2} \sum_{(1,1) \le (i_1, j_1) \ne (i_2, j_2) \le (N, N)} Y_{i_1 j_1} Y_{i_2 j_2}.$$

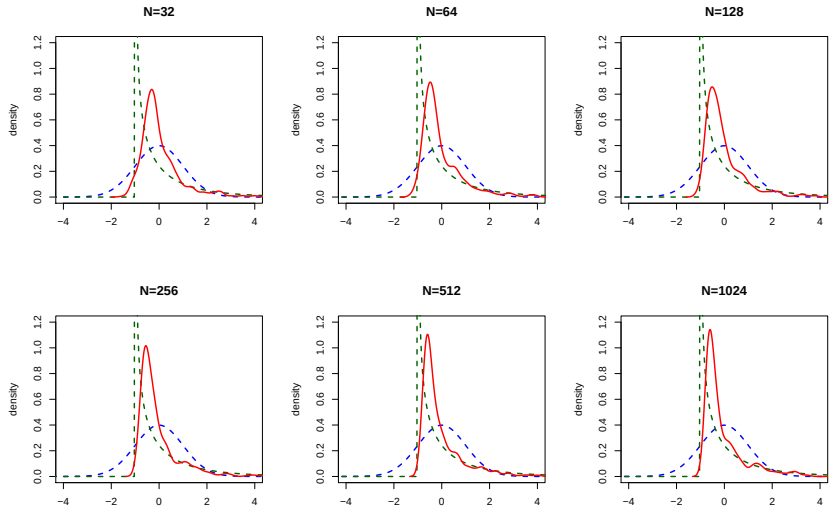

Figure 4.5 – Empirical distribution of $N^2 U_N^{h_2}/\sigma_2$ (solid red), compared to a standard Gaussian distribution (dashed blue) and a centered Chi-squared distribution with one degree of freedom (dashed green). The empirical distribution has been smoothed using the *density* function from the base R stats package.

Remarkably, we observe that

$$U_N^{h_1} = N\binom{N}{2} \sum_{\substack{1\le i_1\le N\\ 1\le j_1\ne j_2\le N}} Y_{i_1 j_1} Y_{i_1 j_2}$$

and

$$U_N^{h_2} = \binom{N}{2}^2 \sum_{\substack{1\le i_1\ne i_2\le N\\ 1\le j_1\ne j_2\le N}} Y_{i_1 j_1} Y_{i_2 j_2}.$$

are sums of similar terms to the ones in $U_N^{h_0}$, but on a restricted sample. Therefore, it may seem surprising that $U_N^{h_1}$ and $U_N^{h_2}$ do not have the same limit distribution.
However, this also shows that, despite being defined as U-statistics of RCE matrices, implicitly, $U_N^{h_1}$ and $U_N^{h_2}$ are incomplete U-statistics based on $U_N^{h_0}$, a U-statistic of i.i.d. observations. The convergence theorem for incomplete U-statistics to a Gaussian variable (Thm. 4.1.1) requires that the size δ_N of the subsample considered in the incomplete U-statistics are negligible compared to the total sample size of the U-statistics, here $\binom{N^2}{2}$. Since for $U_N^{h_2}$, we have $\delta_N^{h_2} = \binom{N}{2}^2 \asymp \binom{N^2}{2}$, then Theorem 4.1.1 does not apply. However, we have $\delta_N^{h_1} = N\binom{N}{2} = o\left(\binom{N^2}{2}\right)$, which explains how $U_N^{h_1}$ can converge to a Gaussian variable, as opposed to $U_N^{h_0}$ and $U_N^{h_2}$.

4.5. Application to the row degree homogeneity test

In this section, we apply our methodology to the network statistical test described in Section 4.1.1. This time, we assume that $Y \sim \mathcal{P} - \mathrm{BEDD}(\Theta)$, where $\Theta = (\lambda, f, g)$. Therefore, the distribution of Y can be defined by

$$\begin{aligned} \xi_i &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall i \geq 1, \\ \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall j \geq 1, \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{P}(\lambda f(\xi_i) g(\eta_j)), && \forall i \geq 1, j \geq 1. \end{aligned}$$

We would like to perform the test $\mathcal{H}_0 : f \equiv 1$ vs. $\mathcal{H}_1 : f \not\equiv 1$ and we use the U-statistic U_N with kernel $h = h1 - h2$ where

$$h_1(Y_{\{i_1,i_2\},\{j_1,j_2\}}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_1j_2} + Y_{i_2j_1}Y_{i_2j_2}),$$

and

$$h_2(Y_{\{i_1,i_2\},\{j_1,j_2\}}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_2j_2} + Y_{i_2j_1}Y_{i_1j_2}).$$

We have already shown that $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}})] = \lambda^2(F_2 - 1)$, which is equal to 0 under $\mathcal{H}_0$. We have also proved that U_N is degenerate because $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1] = \mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \eta_1] = 0$ under $\mathcal{H}_0$.

4.5.1. Definition of the test statistic

Analytic identification of the order of degeneracy Since U_N is degenerate of order at least 1, we check first for graphs $G \in \cup_{r+c=2}\Gamma_{r,c}$ and $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$. In fact, there are only four graphs in $\cup_{r+c=2}\Gamma_{r,c}$. Their corresponding conditional expectations $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid H(G)]$ are calculated in Lemmas 4.F.1 to 4.F.4. Under $\mathcal{H}_0$, they become

- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2] = 0$,
- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \eta_1, \eta_2] = 0$,
- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1] = 0$,
- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \zeta_{11}] = 0$.

Since there are no graph of $\cup_{r+c=2}\Gamma_{r,c}$ such that $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$, that means that U_N is degenerate of order at least 2.

Next, we check for graphs $G \in \cup_{r+c=3}\Gamma_{r,c}$ and $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$. From Lemmas 4.F.5 to 4.F.10, we have under $\mathcal{H}_0$,

- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1] = 0$,
- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] = 0$,

- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] = 0$,
- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2] = 0$,
- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] = 0$,
- $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] = (Y_{11}Y_{12} + \lambda^2 g(\eta_1)g(\eta_2) - \lambda g(V_2)Y_{11} - \lambda g(V_1)Y_{12})/2 \neq 0$.

Therefore, there is one (and only one) graph G satisfying this condition, so we can conclude that the order of degeneracy of U_N is 2.

Identification of the order of degeneracy by simulation Suppose that, instead of the previous calculations, we apply the two techniques suggested in Section 4.3.2 to identify the order of degeneracy $d-1$ of U_N by simulation. Like in Figure 4.1, we have simulated $K = 500$ networks for several values of N

We have already plotted the logarithm of the empirical variance of U_N as a function of $\log N$ in Figure 4.1. The blue dots in Figure 4.6 show the slope of the line between each pair of consecutive dots of Figure 4.1. Because $\mathbb{V}[U_N] = V^{(d)}/N^d + o(N^{-d})$, where $V^{(d)}$ is a constant, the slope should converge to d.

The red dots in Figure 4.6 show the ratio between the empirical mean of $\widehat{V}_N$, the estimator of the non-degenerate asymptotic variance proposed in Chapter 3, and the empirical variance of U_N multiplied by N. According to Theorem 4.3.1, this ratio should converge to d.

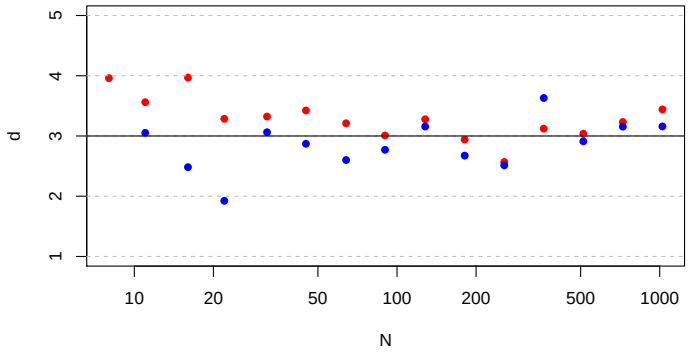

Figure 4.6 – Identification of the order of degeneracy using simulations. For each value of $N \in \{2^{k/2} : 6 \leq k \leq 20\} = \{N_k : 6 \leq k \leq 20\}$, we have simulated 500 networks of size $\lfloor N/2 \rfloor \times \lfloor N/2 \rfloor$, following a Poisson-BEDD model with $\lambda = 1$, $f \equiv 1$ and a power-law form for g, i.e. $g(\eta) = (\alpha_g + 1)\eta^{\alpha_g}$, and α_g chosen such that $G_2 = 2$. A blue dot at a value N_k gives the value of the slope $\left(\log \widehat{\mathbb{V}}[U_{N_k}] - \log \widehat{\mathbb{V}}[U_{N_{k-1}}]\right) / (\log N_k - \log N_{k-1})$, where $\widehat{\mathbb{V}}[U_{N_k}]$ is the empirical variance of U_{N_k} (see Figure 4.1 for the plot of $\log \widehat{\mathbb{V}}[U_N]$ as a function of $\log N$). A red dot at a value N_k show $\widehat{\mathbb{E}}[\widehat{V}_{N_k}] / \left(N_k \widehat{\mathbb{V}}[U_{N_k}]\right)$, where $\widehat{\mathbb{E}}[\widehat{V}_{N_k}]$ is the empirical mean of $\widehat{V}_{N_k}$.

We see that the two sequences of dots converge to 3, confirming the order of degeneracy of 2 found analytically. Both methods are generic and can be implemented with a simple code, while the analytic calculations are tedious. Because the dots of the two methods display some variance, it is necessary to explore several values of N to confirm the order of the degeneracy, but this may still be done quicker than analytic calculations.

Application of the convergence theorem In the analytic identification of the order of degeneracy, we have found that there is only one graph $G \in \cup_{r+c=3}\Gamma_{r,c}$ such that $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid H(G)] \neq 0$. This graph is the graph G such that $H(G) = (\xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12})$, which means $G = K_{1,2}$. Thus, the principal support graphs of U_N are the graphs $(K_{\mathbf{i},\mathbf{j}})_{\mathbf{i}\in\mathcal{P}_1([\![m_N]\!]),\mathbf{j}\in\mathcal{P}_2([\![n_N]\!])}$, which are all connected. Therefore, we can apply Theorem 4.2.8, implying that

$$N^{\frac{3}{2}}U_N \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2),$$

where $\sigma^2 = V^{(1,2)} = \frac{16}{\rho(1-\rho)^2}|\mathrm{Aut}(K_{1,2})|^{-1}\mathbb{E}[(p^{K_{1,2}})^2] = \frac{2\lambda^2}{\rho(1-\rho)^2}$, applying Lemma 4.F.11 with $F_2 = F_3 = F_4 = 1$ under $\mathcal{H}_0$.

The asymptotic variance σ^2 can be estimated by $U_N^{h_2}$, the U-statistic associated to h_2. Let us define the estimator

$$\widehat{\sigma}_N^2 := \frac{2U_N^{h_2}}{\rho(1-\rho)^2},$$

and the test statistic

$$Z_N := \frac{N^{\frac{3}{2}}U_N}{\widehat{\sigma}_N}.$$

Since $U_N^{h_2} \xrightarrow[N\to\infty]{\mathbb{P}} \lambda^2$, it follows that $\widehat{\sigma}_N^2$ is a consistent estimator for σ^2. Therefore, Slutsky's theorem yields

$$Z_N \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

As a result, since $\mathbb{E}[Z_N] > 0$ under $\mathcal{H}_1$, we use a unilateral acceptance interval at significance level $\alpha \in]0,1[$, defined by

$$I(\alpha) :=]\infty, q_{1-\alpha}],$$

where for $x \in]0,1[$, q_x denotes the quantile of order x of a standard Gaussian variable. This interval satisfies

$$\mathbb{P}(Z_N \in I(\alpha)) \xrightarrow[N\to\infty]{} 1-\alpha.$$

Asymptotic behavior of the test statistic under $\mathcal{H}_1$ Because U_N is degenerate under $\mathcal{H}_0$, its asymptotic behavior differs remarkably under $\mathcal{H}_1$. Indeed, under $\mathcal{H}_1$, U_N is not degenerate,

so the non-degenerate convergence theorem (Thm. 3.3.1) applies and

$$\sqrt{\frac{N}{V}}(U_N - \lambda^2(F_2 - 1)) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1),$$

where $V = \frac{4}{\rho}\mathbb{V}\left[\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1]\right] + \frac{4}{1-\rho}\mathbb{V}\left[\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \eta_1]\right]$. Following equations (4.1) and (4.2), we obtain

$$V = \frac{\lambda^4}{\rho}(F_4 - 4F_3 + 8F_2 - F_2^2 - 4) + \frac{4\lambda^4}{1-\rho}(F_2 - 1)^2(G_2 - 1). \tag{4.17}$$

Therefore, replacing U_N with Z_N, we have

$$\frac{\widehat{\sigma}_N}{N\sqrt{V}} Z_N - \sqrt{\frac{N}{V}}\lambda^2(F_2 - 1) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1). \tag{4.18}$$

4.5.2. Simulation study

In this simulation study, we considered that the functions f and g of the BEDD are power-law functions, i.e. $f(\xi) = (\alpha_f + 1)\xi^{\alpha_f}$ and $g(\eta) = (\alpha_g + 1)\eta^{\alpha_g}$, where α_f and α_g are real non-negative numbers. α_f and α_g are directly related to the values $F_k = \int f(\xi)^k d\xi = (\alpha_f + 1)^k/(k\alpha_f + 1)$ and $G_k = \int g(\eta)^k d\eta = (\alpha_g + 1)^k/(k\alpha_g + 1)$. Therefore, this family of BEDD model can be parameterized by (λ, F_2, G_2) instead of (λ, f, g).

Asymptotic normality of the test statistic under $\mathcal{H}_0$ For $N \in \{2^k : 3 \le k \le 10\}$, we have simulated $K = 500$ networks of size $\lfloor N/2 \rfloor \times \lfloor N/2 \rfloor$ ($\rho = 0.5$), $F_2 = 1$ and $G_2 = 2$. The test statistic Z_N is computed for each network. For each N, we observe the empirical distribution of Z_N, which should become closer to a standard Gaussian distribution as N grows.

Figure 4.7 shows the Q-Q plots of the values of Z_N. We observe that for $N = 8$, which corresponds to bipartite networks of size 4×4, the distribution has lighter tails than the standard Gaussian distribution, but this is progressively corrected as N increases.

Power of the test We set a range of values $\{1 + 0.2k : 0 \le k \le 10\}$ for F_2. For each value of F_2 in this set, for $N \in \{2^k : 3 \le k \le 10\}$, we have simulated $K = 500$ networks of size $\lfloor N/2 \rfloor \times \lfloor N/2 \rfloor$ ($\rho = 0.5$), with this value of F_2 and with $G_2 = 2$. The test statistic Z_N is computed for each network and we reject the hypothesis $\mathcal{H}_0$ if $Z_N \notin I(\alpha)$.

The empirical power of the test, i.e. the rate of rejection, is compared to the asymptotic power $\psi_N(F_2)$ for each value of F_2 and N. Since λ and G_2 are fixed in this simulation study, the distribution of U_N only depends on the value chosen for F_2, so we may define it as a function

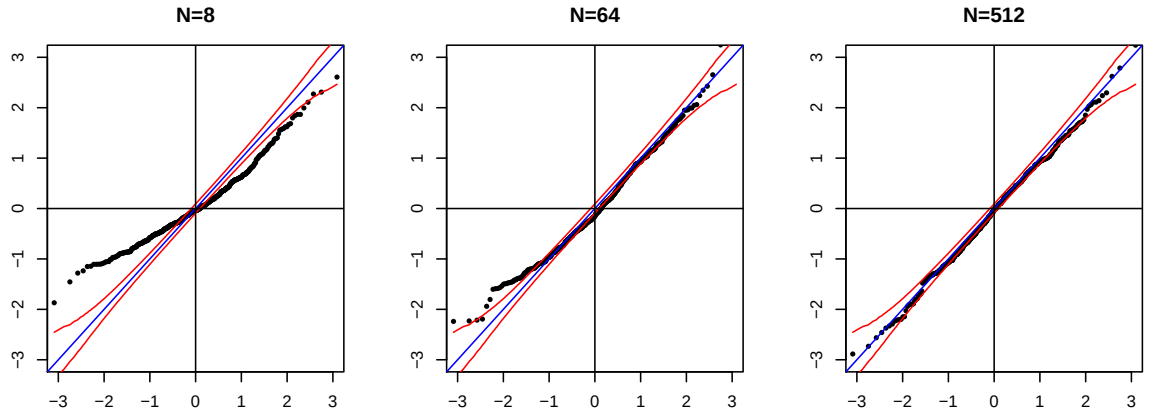

Figure 4.7 – Q-Q plots for the test statistic Z_N under $\mathcal{H}_0$ ($F_2 = 1$).

$U_N(F_2)$ and similarly, we may also write $Z_N(F_2)$ and $V(F_2)$. Note that in the expression of $V(F_2)$ given by (4.17), the terms F_3 and F_4 only depend on F_2.

- For $F_2 = 1$, we are under $\mathcal{H}_0$ and the asymptotic power of the test is $\psi_N(1) = \alpha$.
- For $F_2 > 1$, we are under $\mathcal{H}_1$ and, following (4.18), let the limit variable $\tilde{Z}_N(F_2)$ be such that for all $F_2 > 1$ and $N \geq 1$,

$$\frac{\sigma}{N\sqrt{V(F_2)}}\tilde{Z}_N(F_2) - \sqrt{\frac{N}{V(F_2)}}\lambda^2(F_2 - 1) \overset{\mathcal{D}}{=} \mathcal{N}(0,1).$$

  The asymptotic power is given by the function $\psi_N(F_2) = 1 - \mathbb{P}(\tilde{Z}_N(F_2) \in I(\alpha))$, which translates to

$$\psi_N(F_2) = 1 - \Phi\left(\frac{\sigma q_{1-\alpha}}{N\sqrt{V(F_2)}} - \sqrt{\frac{N}{V(F_2)}}\lambda^2(F_2 - 1)\right), \tag{4.19}$$

  where Φ is the cumulative distribution function of a standard Gaussian variable.

We observe that there are two terms in the argument of Φ, which is an increasing function with $\Phi(0) = 1/2$ and $\lim_{x\to-\infty}\Phi(x) = 0$. We recall that $F_k = (\alpha_f + 1)^k/(k\alpha_f + 1)$. Therefore, we have $F_k \asymp F_2^{k-1}$ and it follows from equation (4.17) that $V(F_2) \asymp F_2^3$. Thus, the first term vanishes when either $F_2 \to \infty$ or $N \to \infty$. For a set value of N, the second term vanishes like $F_2^{-1/2}$. We conclude that $\psi_N(F_2) \xrightarrow[F_2\to\infty]{} 1/2 < 1$, i.e. in equation (4.19), the fact that F_2 moves away from 1 is compensated by the fact that $V(F_2)$ also increases. This also implies that when F_2 becomes larger, it is needed to have at least $N \asymp F_2$ to maintain the second term, thus maintaining the power of this test. In constrast, for a set value of F_2, we have $\psi_N(F_2) \xrightarrow[N\to\infty]{} 1$, which is expected for a statistical test.

The results of the simulations are given in Figure 4.8 for $\alpha = 0.05$. We also see that the test is rather correctly rejected for 5% of the simulated networks at $F_2 = 1$. Overall, the test becomes more powerful when N grows and the rates become nearly perfect for $N \geq 32$, where the test is

rejected for almost all the networks with $F_2 \geq 1.2$. We notice that the empirical power is higher than the theoretical power, especially for small values of $N \leq 32$. This may be explained by the error of approximation of $\mathbb{V}[U_N(F_2)]$ by its asymptotic variance $V(F_2)/N$. For a set value of N, the theoretical power of the test starts by growing rapidly with F_2 but stabilizes and seems to decrease after a certain point, which is in accordance with our predictions.

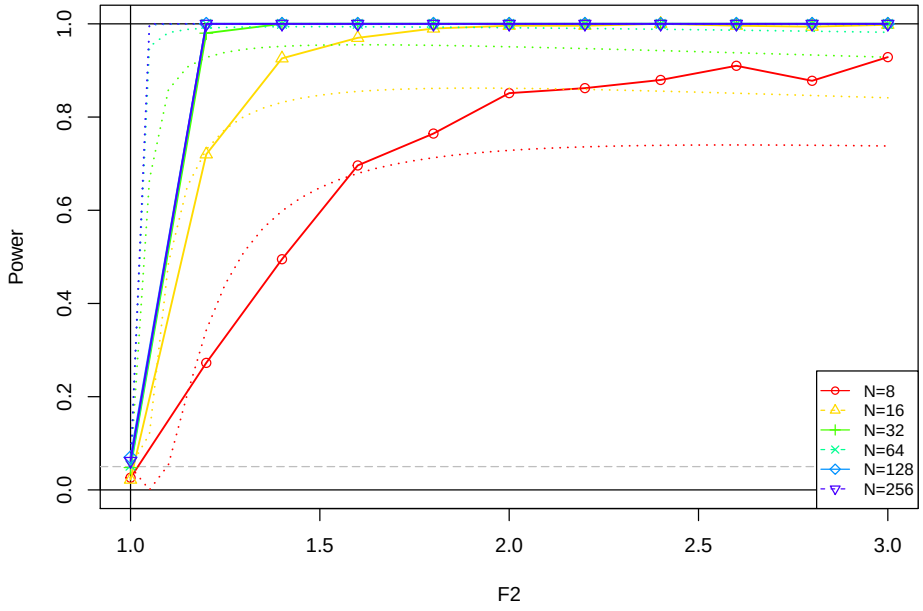

Figure 4.8 – Power of the test $\mathcal{H}_0 : f \equiv 1$ vs. $\mathcal{H}_1 : f \not\equiv 1$. The solid lines represent the empirical power for each value of N. The dashed curves represent the asymptotic power functions $\psi_N(F_2)$ for these same values of N. The grey dashed line is $\alpha = 0.05$, which should be the power of the test at $F_2 = 1$.

Appendix 4.A Proofs for Section 4.2.3

Proof of Proposition 4.2.1. We show that for all F and F' such that $F' \subset F$, we have that $\mathbb{E}[p^F(X) \mid H(F')] = 0$ by induction on F. First, notice that $p^{\varnothing}(X) = \mathbb{E}[X] \in L_2^*(\varnothing)$ being the space of constant variables. Next, fix F and suppose that the induction hypothesis is true for all $\overline{F} \subset F$, i.e. for all $\overline{F}$ and F' such that $F' \subseteq \overline{F} \subset F$, we have that $\mathbb{E}[p^{\overline{F}}(X) \mid H(F')] = 0$. Now

we can calculate for all $F' \subset F$,

$$\begin{aligned}
\mathbb{E}[p^F(X) \mid H(F')] &= \mathbb{E}[\mathbb{E}[X \mid H(F)] \mid H(F')] - \sum_{\overline{F} \subset F} \mathbb{E}[p^{\overline{F}}(X) \mid H(F')] \\
&= \mathbb{E}[X \mid H(F')] - p^{F'}(X) - \sum_{\substack{\overline{F} \subset F \\ \overline{F} \neq F'}} \mathbb{E}[p^{\overline{F}}(X) \mid H(F')] \\
&= \sum_{\overline{F} \subset F'} \mathbb{E}[p^{\overline{F}}(X) \mid H(F')] - \sum_{\substack{\overline{F} \subset F \\ \overline{F} \neq F'}} \mathbb{E}[p^{\overline{F}}(X) \mid H(F')] \\
&= - \sum_{\substack{\overline{F} \subset F \\ \overline{F} \not\subset F'}} \mathbb{E}[p^{\overline{F}}(X) \mid H(F')] \\
&= - \sum_{\substack{\overline{F} \subset F \\ \overline{F} \not\subset F'}} \mathbb{E}[p^{\overline{F}}(X) \mid H(F' \cap \overline{F})].
\end{aligned}$$

By the induction hypothesis, all the terms of this sum are equal to 0, which concludes the proof by induction. □

Appendix 4.B Proofs for Section 4.2.5

Proof of Proposition 4.2.3.

$$\begin{aligned}
\mathbb{V}[U_{m,n}] &= \sum_{(0,0)<(r,c)\leq(p,q)} \mathbb{V}[P^{r,c}_{m,n}] \\
&= \sum_{(0,0)<(r,c)\leq(p,q)} \binom{m}{p}^{-2} \binom{n}{q}^{-2} \sum_{\substack{\mathbf{i},\mathbf{i}' \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j},\mathbf{j}' \in \mathcal{P}_q([\![n]\!])}} \sum_{\substack{G \subseteq K_{\mathbf{i},\mathbf{j}}, G' \subseteq K_{\mathbf{i}',\mathbf{j}'} \\ (v_1(G),v_2(G))=(r,c) \\ (v_1(G'),v_2(G'))=(r,c)}} \mathrm{Cov}(p^G, p^{G'}) \\
&= \sum_{(0,0)<(r,c)\leq(p,q)} \binom{m}{p}^{-1} \binom{n}{q}^{-1} \binom{m-r}{p-r} \binom{n-c}{q-c} r! \binom{p}{r} c! \binom{q}{c} \sum_{G \in \Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1} \mathbb{V}[p^G] \\
&= \sum_{(0,0)<(r,c)\leq(p,q)} \binom{m}{r}^{-1} \binom{n}{c}^{-1} r! \binom{p}{r}^2 c! \binom{q}{c}^2 \sum_{G \in \Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1} \mathbb{E}[(p^G)^2] \\
&= \sum_{(0,0)<(r,c)\leq(p,q)} \frac{(m-r)!}{m!} \frac{(n-c)!}{n!} V^{(r,c)}
\end{aligned}$$

□

Proof of Lemma 4.2.5. Let $G \in \Gamma_{r,c}$.

$$\begin{aligned}
\mathbb{V}[\tilde{P}^G_{m,n}] &= \binom{m}{p}^{-2} \binom{n}{q}^{-2} \sum_{\substack{\mathbf{i},\mathbf{i}' \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j},\mathbf{j}' \in \mathcal{P}_q([\![n]\!])}} \mathrm{Cov}(\tilde{p}^G_{\mathbf{i},\mathbf{j}}, \tilde{p}^G_{\mathbf{i}',\mathbf{j}'}) \\
&= \binom{m}{p}^{-2} \binom{n}{q}^{-2} \sum_{\substack{\mathbf{i},\mathbf{i}' \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j},\mathbf{j}' \in \mathcal{P}_q([\![n]\!])}} \sum_{\Phi,\Phi' \in \mathbb{S}_p \times \mathbb{S}_q} \mathrm{Cov}(p^{\Phi G_{\mathbf{i},\mathbf{j}}}, p^{\Phi' G_{\mathbf{i}',\mathbf{j}'}})
\end{aligned}$$

where for all $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p([\![m]\!]) \times \mathcal{P}_q([\![n]\!])$, $G_{\mathbf{i},\mathbf{j}}$ is any graph of $K_{\mathbf{i},\mathbf{j}}$ which is isomorphic to G.

Now see that if $\Phi G_{\mathbf{i},\mathbf{j}} \neq \Phi' G_{\mathbf{i}',\mathbf{j}'}$, then $\mathrm{Cov}(p^{\Phi G_{\mathbf{i},\mathbf{j}}}, p^{\Phi' G_{\mathbf{i}',\mathbf{j}'}}) = 0$. Otherwise $\Phi G_{\mathbf{i},\mathbf{j}} = \Phi' G_{\mathbf{i}',\mathbf{j}'}$, then $\mathrm{Cov}(p^{\Phi G_{\mathbf{i},\mathbf{j}}}, p^{\Phi' G_{\mathbf{i}',\mathbf{j}'}}) = \mathbb{V}[p^G] = \mathbb{E}[(p^G)^2]$. So, it follows that

$$\mathbb{V}[\tilde{P}^G_{m,n}] = \binom{m}{p}^{-2}\binom{n}{q}^{-2} \sum_{\substack{\mathbf{i},\mathbf{i}' \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j},\mathbf{j}' \in \mathcal{P}_q([\![n]\!])}} \sum_{\Phi,\Phi' \in \mathbb{S}_p \times \mathbb{S}_q} \mathbb{1}(\Phi G_{\mathbf{i},\mathbf{j}} = \Phi' G_{\mathbf{i}',\mathbf{j}'}) \mathbb{E}[(p^G)^2].$$

Finally, applying Lemma 4.2.6, we have

$$\begin{aligned}\mathbb{V}[\tilde{P}^G_{m,n}] &= \binom{m}{p}^{-2}\binom{n}{q}^{-2} \frac{m!(m-r)!}{(m-p)!^2} \frac{n!(n-c)!}{(n-q)!^2} |\mathrm{Aut}(G)| \mathbb{E}[(p^G)^2] \\ &= \frac{(m-r)!}{m!} \frac{(n-c)!}{n!} p!^2 q!^2 |\mathrm{Aut}(G)| \mathbb{E}[(p^G)^2].\end{aligned}$$

□

Proof of Lemma 4.2.6. First, fix $\mathbf{i}_1, \mathbf{j}_1, \Phi_1$. Write $G^1 := \Phi_1 G^1_{\mathbf{i}_1,\mathbf{j}_1}$. We count the number of picks for $\mathbf{i}_2, \mathbf{j}_2, \Phi_2$ such that $\Phi_2 G^2_{\mathbf{i}_2,\mathbf{j}_2} = G^1$.

$\mathbf{i}_2$ and $\mathbf{j}_2$ must contain the r row nodes and the c column nodes of G^1 and Φ_2 must place these nodes in the same order than in G^1, or belong to its automorphism group. This happens for $\binom{m-r}{p-r}\binom{n-c}{q-c}$ picks for $(\mathbf{i}_2, \mathbf{j}_2)$ and for each, there are $(p-r)!(q-c)!|\mathrm{Aut}(G)|$ valid picks for Φ_2.

This happens for all $\binom{m}{p}\binom{n}{q}$ picks of $(\mathbf{i}_1, \mathbf{j}_1)$ and $p!q!$ picks of Φ_1. Therefore,

$$\sum_{\substack{\mathbf{i}_1,\mathbf{i}_2 \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j}_1,\mathbf{j}_2 \in \mathcal{P}_q([\![n]\!])}} \sum_{\Phi_1,\Phi_2 \in \mathbb{S}_p \times \mathbb{S}_q} \mathbb{1}(\Phi_1 G^1_{\mathbf{i}_1,\mathbf{j}_1} = \Phi_2 G^2_{\mathbf{i}_2,\mathbf{j}_2}) = \binom{m}{p}\binom{n}{q}\binom{m-r}{p-r}\binom{n-c}{q-c} p!q!(p-r)!(q-c)!|\mathrm{Aut}(G)|,$$

which develops to the form given by this lemma. □

Appendix 4.C Proofs for Section 4.2.6

Proof of Theorem 4.2.7. Since, $d-1$ is the order of degeneracy, we have $P^{r,c}_N = 0$ for all (r,c) such that $r + c < d$. Therefore, we have $U_N - p^{\varnothing} - \sum_{\substack{(0,0)\leq(r,c)\leq(p,q) \\ r+c=d}} P^{r,c}_N = \sum_{\substack{(0,0)\leq(r,c)\leq(p,q) \\ r+c>d}} P^{r,c}_N$. So

$$\begin{aligned}\mathbb{V}\left[N^{d/2}\left(U_N - p^{\varnothing} - \sum_{\substack{(0,0)\leq(r,c)\leq(p,q) \\ r+c=d}} P^{r,c}_N\right)\right] &= N^d \sum_{\substack{(0,0)\leq(r,c)\leq(p,q) \\ r+c>d}} \mathbb{V}[P^{r,c}_N] \\ &= N^d \sum_{\substack{(0,0)\leq(r,c)\leq(p,q) \\ r+c>d}} \frac{(m_N-r)!}{m_N!} \frac{(n_N-c)!}{n_N!} V^{(r,c)}\end{aligned}$$

But for all (r,c), we have $\frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} = O(N^{-r-c})$, therefore

$$\mathbb{V}\left[N^{d/2}\left(U_N - p^{\varnothing} - \sum_{\substack{(0,0)\le(r,c)\le(p,q)\\ r+c=d}} P_N^{r,c}\right)\right] = N^d \times O\left(\sum_{\substack{(0,0)\le(r,c)\le(p,q)\\ r+c>d}} N^{-r-c}\right)$$
$$= N^d \times o(N^{-d})$$
$$= o(1).$$

Finally, this implies that $N^{d/2}(U_N - p^{\varnothing}) = N^{d/2}\sum_{\substack{(0,0)\le(r,c)\le(p,q)\\ r+c=d}} P_N^{r,c} + o_P(1)$, which proves the theorem. □

Appendix 4.D Proofs for Section 4.2.9

4.D.1 Proof of Theorem 4.2.12

In order to prove Theorem 4.2.12, define $\mathcal{S} = \{(r_\ell, c_\ell) : 1 \le \ell \le L\}$ the set of principal degrees of h. We may define $\mathcal{S}_0$ the set of couples $(0,0) < (r_0,c_0) \le (p,q)$ such that $\gamma(N)^{-1} = o(m_N^{-r_0} n_N^{-c_0})$, for any $(r,c) \in \mathcal{S}$. We may also define $\mathcal{S}_+$, the set of couples $(0,0) < (r_+,c_+) \le (p,q)$ such that $m_N^{-r_+} n_N^{-c_+} = o(\gamma(N)^{-1})$, for any $(r,c) \in \mathcal{S}$. We need the following lemma.

Lemma 4.D.1. *For all $(r,c) \in \mathcal{S}_0$, for all graphs G such that $(v_1(G), v_2(G)) = (r,c)$, we have $p^G = 0$.*

Proof. We have

$$\mathbb{V}[U_N] = \sum_{(0,0)<(r,c)\le(p,q)} \frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} V^{(r,c)}$$
$$= \sum_{(r,c)\in\mathcal{S}_0} \frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} V^{(r,c)} + \sum_{(r,c)\in\mathcal{S}} \frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} V^{(r,c)}$$
$$+ \sum_{(r,c)\in\mathcal{S}_+} \frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} V^{(r,c)}.$$

By definition, $(r,c) \in \mathcal{S}_+$, $m_N^{-r} n_N^{-c} = o(\gamma(N)^{-1})$ and $\mathbb{V}[U_N] = \gamma(N)^{-1}\sum_{1\le\ell\le L}\alpha_\ell V^{(r_\ell,c_\ell)} + o(\gamma(N)^{-1})$. Therefore,

$$\sum_{(r,c)\in\mathcal{S}_0} \frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} V^{(r,c)} = \sum_{l=1}^{L}\left(\gamma(N)^{-1}\alpha_\ell - \frac{(m_N-r_\ell)!}{m_N!}\frac{(n_N-c_\ell)!}{n_N!}\right) V^{(r_\ell,c_\ell)} + o(\gamma(N)^{-1}).$$

Again, by definition, we have for all $1 \leq \ell \leq L$, $\gamma(N)\frac{(m_N-r_\ell)!}{m_N!}\frac{(n_N-c_\ell)!}{n_N!} \xrightarrow[N\to\infty]{} \alpha_\ell$. Therefore, the previous equation yields

$$\gamma(N) \sum_{(r,c)\in\mathcal{S}_0} \frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} V^{(r,c)} = o(1).$$

But for all $(r,c) \in \mathcal{S}_0$, $\gamma(N)\frac{(m_N-r)!}{m_N!}\frac{(n_N-c)!}{n_N!} \xrightarrow[N\to\infty]{} \infty$. Since $V^{(r,c)} \geq 0$ for all $(0,0) \leq (r,c) \leq (p,q)$, this means that for all $(r,c) \in \mathcal{S}_0$, we have $V^{(r,c)} = 0$. Thus, $V^{(r,c)} = \frac{p!}{(p-r)!}\frac{q!}{(q-r)!}\sum_{G\in\Gamma_{r,c}}|\mathrm{Aut}(G)|^{-1}\mathbb{V}[p^G]$, this means $\mathbb{V}[p^G] = 0$ for all $G \in \Gamma_{r,c}$.

Finally, let G be any graph such that $(v_1(G), v_2(G)) = (r,c)$. Then there exists a graph $G^* \in \Gamma_{r,c}$ such that $\mathbb{V}[p^G] = \mathbb{V}[p^{G^*}]$. We have already shown that $\mathbb{V}[p^{G^*}] = 0$ for all $(r,c) \in \mathcal{S}_0$, so adding the fact that $\mathbb{E}[p^G] = 0$ for all graphs $G \neq \varnothing$, it means that $p^G = 0$, for all graphs G such that $(v_1(G), v_2(G)) = (r,c) \in \mathcal{S}_0$.

□

Proof of Theorem 4.2.12.

$$\sqrt{\gamma(N)}\left[U_N - p^\varnothing - \sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell}\right] = \sqrt{\gamma(N)}\left[\sum_{(r,c)\in\mathcal{S}_0} P_N^{r,c} + \sum_{(r,c)\in\mathcal{S}_+} P_N^{r,c}\right].$$

By Lemma 4.D.1, $P_N^{r,c} = 0$ for all $(r,c) \in \mathcal{S}_0$.

$$\begin{aligned}\mathbb{V}[\sqrt{\gamma(N)}\sum_{(r,c)\in\mathcal{S}_+} P_N^{r,c}] &= \gamma(N)\sum_{(r,c)\in\mathcal{S}_+}\frac{(m-r)!}{m!}\frac{(n-c)!}{n!}V^{(r,c)}\\ &= o(1).\end{aligned}$$

That means $\sqrt{\gamma(N)}(U_N - p^\varnothing) = \sqrt{\gamma(N)}\sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell} + o_P(1)$, which concludes the proof. □

4.D.2 Proof of Theorem 4.2.13

Proof. Theorem 4.2.12 states that $\sqrt{\gamma(N)}(U_N - p^\varnothing)$ has the same limit as $\sqrt{\gamma(N)}\sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell}$.

For all $(0,0) < (r,c) \leq (p,q)$,

$$P_N^{r,c} = \sum_{G\in\Gamma_{r,c}} \frac{1}{(p-r)!(q-c)!|\mathrm{Aut}(G)|}\tilde{P}_N^G.$$

So

$$\sqrt{\gamma(N)}\sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell} = \sum_{\ell=1}^{L}\sqrt{\gamma(N)}m_N^{-r_\ell/2}n_N^{-c_\ell/2}\sum_{G\in\Gamma_{r_\ell,c_\ell}}\frac{m_N^{r_\ell/2}n_N^{c_\ell/2}\tilde{P}_N^G}{(p-r_\ell)!(q-c_\ell)!|\mathrm{Aut}(G)|}.$$

By definition, $\gamma(N)m_N^{-r_\ell}n_N^{-c_\ell} \xrightarrow[N\to\infty]{} \alpha_\ell$. Therefore, by Lemma 4.2.10, $\sqrt{\gamma(N)}\sum_{\ell=1}^{L} P_N^{r_\ell,c_\ell}$ converges in distribution to $Z = \sum_{\ell=1}^{L}\sqrt{\alpha_\ell}\sum_{G\in\Gamma_{r_\ell,c_\ell}} W_G$, where all W_G are independent gaussian variables with mean 0 and variance $\frac{(p!)^2(q!)^2}{((p-r_\ell)!)^2((q-c_\ell)!)^2|\mathrm{Aut}(G)|}\mathbb{V}[p^G]$.

Finally, it follows that Z is a gaussian variable with mean 0 and variance $\sum_{\ell=1}^{L}\sqrt{\alpha_\ell}V^{(r_\ell,c_\ell)}$ where

$$V^{(r_\ell,c_\ell)} = \sum_{G\in\Gamma_{r_\ell,c_\ell}} \frac{(p!)^2(q!)^2}{((p-r_\ell)!)^2((q-c_\ell)!)^2|\mathrm{Aut}(G)|}\mathbb{V}[p^G]$$

□

Appendix 4.E Proof of Theorem 4.3.1

In the whole section, U_N is a U-statistic on Y, a row-column exchangeable matrix, of kernel h of size $p\times q$. We remind a few definitions and notations from Chapter 3.

For $(0,0) < (r,c) \le (p,q)$, we denote $v_h^{r,c} = \mathrm{Cov}\left(h_{\mathbf{i},\mathbf{j}}, h_{\mathbf{i}',\mathbf{j}'}\right)$ where $\mathbf{i}$ and $\mathbf{i}'$ have r elements in common and $\mathbf{j}$ and $\mathbf{j}'$ have c elements in common.

For N such that $m_N \ge p$ and $n_N \ge q$, we denote

$$\mathcal{S}_N^{p,q} := \{(\mathbf{i},\mathbf{j}) : \mathbf{i}\in\mathcal{P}_p(\llbracket m_N\rrbracket), \mathbf{j}\in\mathcal{P}_q(\llbracket n_N\rrbracket)\},$$

and for some sets of indices $\underline{\mathbf{i}}\in\mathcal{P}_{\underline{p}}(\llbracket m_N\rrbracket)$ and $\underline{\mathbf{j}}\in\mathcal{P}_{\underline{q}}(\llbracket n_N\rrbracket)$ such that $\underline{p}\le p$ and $\underline{q}\le q$,

$$\mathcal{S}_{N,(\underline{\mathbf{i}},\underline{\mathbf{j}})}^{p,q} := \left\{(\mathbf{i},\mathbf{j})\in\mathcal{S}_N^{p,q} : \underline{\mathbf{i}}\subset\mathbf{i}, \underline{\mathbf{j}}\subset\mathbf{j}\right\}.$$

The estimators of the conditional expectations are

$$\widehat{\mu}_N^{(i)} = \binom{m_N-1}{p-1}^{-1}\binom{n_N}{q}^{-1}\sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_{N,(\{i\},\varnothing)}^{p,q}} h_{\mathbf{i},\mathbf{j}}$$

and

$$\widehat{\nu}_N^{(j)} = \binom{m_N}{p}^{-1}\binom{n_N-1}{q-1}^{-1}\sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_{N,(\varnothing,\{j\})}^{p,q}} h_{\mathbf{i},\mathbf{j}}.$$

The expression of the variance estimator is

$$\widehat{V}_N = \frac{p^2}{\rho}\widehat{v}_N^{1,0} + \frac{q^2}{1-\rho}\widehat{v}_N^{0,1},$$

where

$$\widehat{v}_N^{1,0} = \binom{m_N}{2}^{-1}\sum_{1\le i_1<i_2\le m_N}\frac{(\widehat{\mu}_N^{(i_1)} - \widehat{\mu}_N^{(i_2)})^2}{2}$$

and

$$\widehat{v}_N^{0,1} = \binom{n_N}{2}^{-1} \sum_{1 \le j_1 < j_2 \le n_N} \frac{(\widehat{\nu}_N^{(j_1)} - \widehat{\nu}_N^{(j_2)})^2}{2}.$$

The proof of Theorem 4.3.1 is not technical, but rather tedious. Therefore, we are proving three lemmas first.

Lemma 4.E.1. *If U_N is degenerate of order $d-1$, then*

$$\mathbb{V}[U_N] = \frac{1}{N^d} \sum_{\substack{(0,0)<(r,c)\le(p,q) \\ r+c=d}} \frac{r!c!}{\rho^r(1-\rho)^c} \binom{p}{r}^2 \binom{q}{c}^2 v^{r,c} + o(N^{-d}).$$

Proof.

$$\begin{aligned}\mathbb{V}[U_N] &= \binom{m_N}{p}^{-2} \binom{n_N}{q}^{-2} \sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{S}_N^{p,q} \\ (\mathbf{i}',\mathbf{j}')\in\mathcal{S}_N^{p,q}}} \operatorname{Cov}\left(h_{\mathbf{i},\mathbf{j}}, h_{\mathbf{i}',\mathbf{j}'}\right) \\ &= \binom{m_N}{p}^{-1} \binom{n_N}{q}^{-1} \sum_{(0,0)<(r,c)\le(p,q)} \binom{p}{r}\binom{m_N-p}{p-r}\binom{q}{c}\binom{n_N-q}{q-c} v^{r,c}.\end{aligned}$$

If U_N is degenerate of order $d-1$, then $v^{r,c} = 0$ for $r + c \le d-1$, so

$$\begin{aligned}\mathbb{V}[U_N] &= \sum_{\substack{(0,0)<(r,c)\le(p,q) \\ r+c=d}} \frac{r!c!}{m_N^r n_N^c} \binom{p}{r}^2 \binom{q}{c}^2 v^{r,c} + o(N^{-r-c}) \\ &= \frac{1}{N^d} \sum_{\substack{(0,0)<(r,c)\le(p,q) \\ r+c=d}} \frac{r!c!}{\rho^r(1-\rho)^c} \binom{p}{r}^2 \binom{q}{c}^2 v^{r,c} + o(N^{-d}).\end{aligned}$$

□

Lemma 4.E.2. *If U_N is degenerate of order $d-1$, then*

$$\mathbb{V}[\widehat{\mu}_N^{(1)}] = \frac{1}{N^{d-1}} \sum_{\substack{(1,0)\le(r,c)\le(p,q) \\ r+c=d}} \frac{r}{p^2} \frac{r!c!}{\rho^{r-1}(1-\rho)^c} \binom{p}{r}^2 \binom{q}{c}^2 v^{r,c} + o(N^{-d+1}).$$

and

$$\mathbb{V}[\widehat{\nu}_N^{(1)}] = \frac{1}{N^{d-1}} \sum_{\substack{(0,1)\le(r,c)\le(p,q) \\ r+c=d}} \frac{c}{q^2} \frac{r!c!}{\rho^r(1-\rho)^{c-1}} \binom{p}{r}^2 \binom{q}{c}^2 v^{r,c} + o(N^{-d+1}).$$

Proof. We only prove this lemma for $\mathbb{V}[\widehat{\mu}_N^{(1)}]$. The result for $\mathbb{V}[\widehat{\nu}_N^{(1)}]$ can be obtained analogously.

$$\begin{aligned}\mathbb{V}[\widehat{\mu}_N^{(1)}] &= \binom{m_N-1}{p-1}^{-2} \binom{n_N}{q}^{-2} \sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{S}_{N,(\{1\},\varnothing)}^{p,q} \\ (\mathbf{i}',\mathbf{j}')\in\mathcal{S}_{N,(\{1\},\varnothing)}^{p,q}}} \operatorname{Cov}\left(h_{\mathbf{i},\mathbf{j}}, h_{\mathbf{i}',\mathbf{j}'}\right) \\ &= \binom{m_N-1}{p-1}^{-1} \binom{n_N}{q}^{-1} \sum_{(1,0)\le(r,c)\le(p,q)} \binom{p-1}{r-1}\binom{m_N-p}{p-r}\binom{q}{c}\binom{n_N-q}{q-c} v^{r,c}.\end{aligned}$$

If U_N is degenerate of order $d-1$, then $v^{r,c}=0$ for $r+c\le d-1$, so

$$\begin{aligned}
\mathbb{V}[\widehat{\mu}_N^{(1)}] &= \binom{m_N-1}{p-1}^{-1}\binom{n_N}{q}^{-1}\sum_{\substack{(1,0)\le(r,c)\le(p,q)\\ r+c=d}}\binom{p-1}{r-1}\binom{m_N-p}{p-r}\binom{q}{c}\binom{n_N-q}{q-c}v^{r,c}+o(N^{-d+1})\\
&= \frac{(p-1)!}{m_N^{p-1}}\frac{q!}{n_N^{q}}\sum_{\substack{(1,0)\le(r,c)\le(p,q)\\ r+c=d}}\binom{p-1}{r-1}\frac{m_N^{p-r}}{(p-r)!}\binom{q}{c}\frac{n_N^{q-c}}{(q-c)!}v^{r,c}+o(N^{-d+1})\\
&= \sum_{\substack{(1,0)\le(r,c)\le(p,q)\\ r+c=d}}\frac{r}{p^2}\frac{r!c!}{m_N^{r-1}n_N^{c}}\binom{p}{r}^2\binom{q}{c}^2v^{r,c}+o(N^{-d+1})\\
&= \frac{1}{N^{d-1}}\sum_{\substack{(1,0)\le(r,c)\le(p,q)\\ r+c=d}}\frac{r}{p^2}\frac{r!c!}{\rho^{r-1}(1-\rho)^{c}}\binom{p}{r}^2\binom{q}{c}^2v^{r,c}+o(N^{-d+1}).
\end{aligned}$$

□

Lemma 4.E.3. *If U_N is degenerate of order $d-1$, then*

$$\operatorname{Cov}(\widehat{\mu}_N^{(1)},\widehat{\mu}_N^{(2)})=o(N^{-d+1}).$$

Proof.

$$\begin{aligned}
\operatorname{Cov}(\widehat{\mu}_N^{(1)},\widehat{\mu}_N^{(2)}) &= \binom{m_N-1}{p-1}^{-2}\binom{n_N}{q}^{-2}\sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{S}^{p,q}_{N,(\{1\},\varnothing)}\\ (\mathbf{i}',\mathbf{j}')\in\mathcal{S}^{p,q}_{N,(\{2\},\varnothing)}}}\operatorname{Cov}\left(h_{\mathbf{i},\mathbf{j}},h_{\mathbf{i}',\mathbf{j}'}\right)\\
&= \binom{m_N-1}{p-1}^{-2}\binom{n_N}{q}^{-1}\sum_{(0,0)<(r,c)\le(p,q)}\left[\binom{m_N-2}{p-1}\binom{p}{r}\binom{m_N-p-1}{p-r-1}\right.\\
&\quad\left.+\binom{m_N-2}{p-2}\binom{p-1}{r-1}\binom{m_N-p}{p-r}\right]\binom{q}{c}\binom{n_N-q}{q-c}v^{r,c}
\end{aligned}$$

If U_N is degenerate of order $d-1$, then $v^{r,c}=0$ for $r+c\le d-1$, so

$$\begin{aligned}
\mathrm{Cov}(\widehat{\mu}_N^{(1)},\widehat{\mu}_N^{(2)}) &= \binom{m_N-1}{p-1}^{-2}\binom{n_N}{q}^{-1}\sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}}\left[\binom{m_N-2}{p-1}\binom{p}{r}\binom{m_N-p-1}{p-r-1}\right.\\
&+\left.\binom{m_N-2}{p-2}\binom{p-1}{r-1}\binom{m_N-p}{p-r}\right]\binom{q}{c}\binom{n_N-q}{q-c}v^{r,c}+o(N^{-d})\\
&=\frac{(p-1)!^2}{m_N^{2p-2}}\frac{q!}{n_N^q}\sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}}\left[\frac{m_N^{p-1}m_N^{p-r-1}}{(p-1)!(p-r-1)!}\binom{p}{r}\right.\\
&+\left.\frac{m_N^{p-2}m_N^{p-r}}{(p-2)!(p-r)!}\binom{p-1}{r-1}\right]\binom{q}{c}\frac{n_N^{q-c}}{(q-c)!}v^{r,c}+o(N^{-d})\\
&=\sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}}\left[\frac{r!(p-r)}{m_N^r p}\binom{p}{r}^2+\frac{(p-1)(p-1)!}{m_N^r(p-r)!}\frac{r}{p}\binom{p}{r}\right]\frac{c!}{n_N^c}\binom{q}{c}^2v^{r,c}+o(N^{-d})\\
&=\sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}}\frac{r!c!}{m_N^r n_N^c}\binom{p}{r}^2\binom{q}{c}^2\left[\frac{p-r}{p}+\frac{r(p-1)}{p^2}\right]v^{r,c}+o(N^{-d})\\
&=\frac{1}{N^d}\sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}}\frac{r!c!}{\rho^r(1-\rho)^c}\binom{p}{r}^2\binom{q}{c}^2\left[1-\frac{r}{p^2}\right]v^{r,c}+o(N^{-d})\\
&=o(N^{-d+1})
\end{aligned}$$

□

Proof of Theorem 4.3.1. First, we can rewrite the expression of $\mathbb{V}[U_N]$ in Lemma 4.E.1 as follows

$$\mathbb{V}[U_N]=\frac{1}{N^d}\sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}}\tilde{v}^{r,c}+o(N^{-d}), \tag{4.20}$$

where for $(0,0)<(r,c)\le(p,q), r+c=d$,

$$\tilde{v}^{r,c}=\frac{r!c!}{\rho^r(1-\rho)^c}\binom{p}{r}^2\binom{q}{c}^2v^{r,c}.$$

Second, notice that

$$\begin{aligned}
\mathbb{E}[\hat{v}_N^{1,0}] &= \binom{m_N}{2}^{-1}\sum_{1\le i_1<i_2\le m_N}\mathbb{E}\left[\frac{(\widehat{\mu}_N^{(i_1)}-\widehat{\mu}_N^{(i_2)})^2}{2}\right]\\
&=\binom{m_N}{2}^{-1}\sum_{1\le i_1<i_2\le m_N}\mathbb{E}\left[\frac{(\widehat{\mu}_N^{(1)})^2-2\widehat{\mu}_N^{(1)}\widehat{\mu}_N^{(2)}+(\widehat{\mu}_N^{(2)})^2}{2}\right]\\
&=\mathbb{V}[\widehat{\mu}_N^{(1)}]-\mathrm{Cov}(\widehat{\mu}_N^{(1)},\widehat{\mu}_N^{(2)}).
\end{aligned}$$

Using Lemmas 4.E.2 and 4.E.3, we find

$$\begin{aligned}\mathbb{E}[\hat{v}_N^{1,0}] &= \frac{1}{N^{d-1}} \sum_{\substack{(1,0)\le(r,c)\le(p,q)\\ r+c=d}} \frac{r}{p^2} \frac{r!c!}{\rho^{r-1}(1-\rho)^c} \binom{p}{r}^2 \binom{q}{c}^2 v^{r,c} + o(N^{-d+1}) \\ &= \frac{1}{N^{d-1}} \frac{\rho}{p^2} \sum_{\substack{(1,0)\le(r,c)\le(p,q)\\ r+c=d}} r\tilde{v}^{r,c} + o(N^{-d+1}) \\ &= \frac{1}{N^{d-1}} \frac{\rho}{p^2} \sum_{\substack{(0,0)\le(r,c)\le(p,q)\\ r+c=d}} r\tilde{v}^{r,c} + o(N^{-d+1}). \end{aligned} \tag{4.21}$$

Analogously,

$$\mathbb{E}[\hat{v}_N^{0,1}] = \frac{1}{N^{d-1}} \frac{1-\rho}{q^2} \sum_{\substack{(0,0)\le(r,c)\le(p,q)\\ r+c=d}} c\tilde{v}^{r,c} + o(N^{-d+1}). \tag{4.22}$$

Therefore, combining (4.21), (4.22) and finally (4.20), we have

$$\begin{aligned}\mathbb{E}[\hat{V}_N] &= \frac{p^2}{\rho}\mathbb{E}[\hat{v}_N^{1,0}] + \frac{q^2}{(1-\rho)}\mathbb{E}[\hat{v}_N^{0,1}] \\ &= \frac{1}{N^{d-1}} \sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}} (r+c)\tilde{v}^{r,c} + o(N^{-d+1}) \\ &= \frac{d}{N^{d-1}} \sum_{\substack{(0,0)<(r,c)\le(p,q)\\ r+c=d}} \tilde{v}^{r,c} + o(N^{-d+1}) \\ &= dN\mathbb{V}[U_N^h] + o(N^{-d+1}). \end{aligned}$$

Thus, since $d \ge 1$, we have proved that

$$\frac{N\mathbb{V}[U_N^h]}{\mathbb{E}[\hat{V}_N]} \xrightarrow[N\to\infty]{} \frac{1}{d},$$

which is the targeted result. □

Appendix 4.F Proofs for Section 4.5

In this section, we calculate the conditional expectations and the variances of Section 4.5. Let the distribution of Y be defined by

$$\begin{aligned} &\xi_i \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \le i \le m, \\ &\eta_j \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \le j \le n, \\ &Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{P}(\lambda f(\xi_i) g(\eta_j)), && \forall 1 \le i \le m, 1 \le j \le n. \end{aligned}$$

Let U_N be the U-statistic with kernel $h = h1 - h2$ where

$$h_1(Y_{\{i_1,i_2\},\{j_1,j_2\}}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_1j_2} + Y_{i_2j_1}Y_{i_2j_2}),$$

and

$$h_2(Y_{\{i_1,i_2\},\{j_1,j_2\}}) = \frac{1}{2}(Y_{i_1j_1}Y_{i_2j_2} + Y_{i_2j_1}Y_{i_1j_2}).$$

Lemma 4.F.1. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2] = \frac{\lambda^2}{2}(f(\xi_1) - f(\xi_2))^2$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \xi_2] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \xi_2] \\
&= \frac{1}{2}\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1)g(\eta_2) + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \xi_2] \\
&= \frac{\lambda^2}{2}(f(\xi_1)^2 + f(\xi_2)^2),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \xi_2] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \xi_2] \\
&= \frac{1}{2}\mathbb{E}[2\lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \xi_1, \xi_2] \\
&= \lambda^2 f(\xi_1)f(\xi_2).
\end{aligned}$$

This proves the result. □

Lemma 4.F.2. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \eta_1, \eta_2] = \lambda^2(F_2 - 1)g(\eta_1)g(\eta_2)$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \eta_1, \eta_2] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \eta_1, \eta_2] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \eta_1, \eta_2] \\
&= \frac{1}{2}\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1)g(\eta_2) + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \eta_1, \eta_2] \\
&= \lambda^2 F_2 g(\eta_1)g(\eta_2),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \eta_1, \eta_2] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \eta_1, \eta_2] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \eta_1, \eta_2] \\
&= \frac{1}{2}\mathbb{E}[2\lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \eta_1, \eta_2] \\
&= \lambda^2 g(\eta_1)g(\eta_2).
\end{aligned}$$

This proves the result. □

Lemma 4.F.3. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1] = \frac{\lambda^2}{2}(f(\xi_1)^2 - 2f(\xi_1) + F_2)g(\eta_1)$.

Proof. We have

$$\begin{aligned}\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \eta_1] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \eta_1] \\ &= \frac{1}{2}\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1)g(\eta_2) + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1] \\ &= \frac{\lambda^2}{2}(f(\xi_1)^2 + F_2)g(\eta_1),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \eta_1] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \eta_1] \\ &= \frac{1}{2}\mathbb{E}[2\lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1] \\ &= \lambda^2 f(\xi_1)g(\eta_1).\end{aligned}$$

This proves the result. □

Lemma 4.F.4. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \zeta_{11}] = \frac{\lambda}{2}(f(\xi_1) - 1)Y_{11} + \frac{\lambda^2}{2}(F_2 - f(\xi_1))g(\eta_1)$.

Proof. We have

$$\begin{aligned}\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \zeta_{11}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \eta_1, \zeta_{11}] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}] \mid \xi_1, \eta_1, \zeta_{11}] \\ &= \frac{1}{2}\mathbb{E}[\lambda f(\xi_1)g(\eta_2)Y_{11} + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1, \zeta_{11}] \\ &= \frac{\lambda}{2}f(\xi_1)Y_{11} + \frac{\lambda^2}{2}F_2 g(\eta_1),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \zeta_{11}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \eta_1, \zeta_{11}] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}] \mid \xi_1, \eta_1, \zeta_{11}] \\ &= \frac{1}{2}\mathbb{E}[\lambda Y_{11}f(\xi_2)g(\eta_2) + \lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1, \zeta_{11}] \\ &= \frac{\lambda}{2}Y_{11} + \frac{\lambda^2}{2}f(\xi_1)g(\eta_1).\end{aligned}$$

This proves the result. □

Lemma 4.F.5. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1] = \frac{\lambda^2}{2}(f(\xi_1) - f(\xi_2))^2 g(\eta_1)$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \xi_2, \eta_1] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \xi_2, \eta_1] \\
&= \frac{1}{2}\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1)g(\eta_2) + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \xi_2, \eta_1] \\
&= \frac{\lambda^2}{2}(f(\xi_1)^2 + f(\xi_2)^2)g(\eta_1),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \xi_2, \eta_1] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \xi_2, \eta_1] \\
&= \frac{1}{2}\mathbb{E}[2\lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \xi_1, \xi_2, \eta_1] \\
&= \lambda^2 f(\xi_1)f(\xi_2)g(\eta_1).
\end{aligned}$$

This proves the result. □

Lemma 4.F.6. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] = \frac{\lambda}{2}(f(\xi_1) - f(\xi_2))Y_{11} + \frac{\lambda^2}{2}(f(\xi_2) - f(\xi_1))f(\xi_2)g(\eta_1)$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}] \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\lambda f(\xi_1)g(\eta_2)Y_{11} + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] \\
&= \frac{\lambda}{2}f(\xi_1)Y_{11} + \frac{\lambda^2}{2}f(\xi_2)^2 g(\eta_1),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}] \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\lambda f(\xi_2)g(\eta_2)Y_{11} + \lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}] \\
&= \frac{\lambda}{2}f(\xi_2)Y_{11} + \frac{\lambda^2}{2}f(\xi_1)f(\xi_2)g(\eta_1).
\end{aligned}$$

This proves the result. □

Lemma 4.F.7. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] = \frac{\lambda}{2}(f(\xi_1) - f(\xi_2))(Y_{11} - Y_{21})$.

Proof. We have

$$\begin{aligned}\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}, Y_{21}] \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] \\ &= \frac{1}{2}\mathbb{E}[\lambda f(\xi_1)g(\eta_2)Y_{11} + \lambda^2 f(\xi_2)g(\eta_2)Y_{21} \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] \\ &= \frac{\lambda}{2}(f(\xi_1)Y_{11} + f(\xi_2)Y_{21}),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}, Y_{21}] \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] \\ &= \frac{1}{2}\mathbb{E}[\lambda f(\xi_2)g(\eta_2)Y_{11} + \lambda f(\xi_1)g(\eta_2)Y_{21} \mid \xi_1, \xi_2, \eta_1, \zeta_{11}, \zeta_{21}] \\ &= \frac{\lambda}{2}(f(\xi_2)Y_{11} + f(\xi_1)Y_{21}).\end{aligned}$$

This proves the result. □

Lemma 4.F.8. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2] = \frac{\lambda^2}{2}(f(\xi_1)^2 - 2f(\xi_1) + F_2)g(\eta_1)g(\eta_2)$.

Proof. We have

$$\begin{aligned}\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \eta_1, \eta_2] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \eta_1, \eta_2] \\ &= \frac{1}{2}\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1)g(\eta_2) + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1, \eta_2] \\ &= \frac{\lambda^2}{2}(f(\xi_1)^2 + F_2)g(\eta_1)g(\eta_2),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \eta_1, \eta_2] \\ &= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}] \mid \xi_1, \eta_1, \eta_2] \\ &= \frac{1}{2}\mathbb{E}[2\lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1, \eta_2] \\ &= \lambda^2 f(\xi_1)g(\eta_1)g(\eta_2).\end{aligned}$$

This proves the result. □

Lemma 4.F.9. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] = \frac{\lambda}{2}(f(\xi_1) - 1)g(\eta_2)Y_{11} + \frac{\lambda^2}{2}(F_2 - f(\xi_1))g(\eta_1)g(\eta_2)$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}] \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\lambda f(\xi_1)g(\eta_2)Y_{11} + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] \\
&= \frac{\lambda}{2}f(\xi_1)g(\eta_2)Y_{11} + \frac{\lambda^2}{2}F_2 g(\eta_1)g(\eta_2),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}] \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] \\
&= \frac{1}{2}\mathbb{E}[\lambda f(\xi_2)g(\eta_2)Y_{11} + \lambda^2 f(\xi_1)f(\xi_2)g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}] \\
&= \frac{\lambda}{2}g(\eta_2)Y_{11} + \frac{\lambda^2}{2}f(\xi_1)g(\eta_1)g(\eta_2).
\end{aligned}$$

This proves the result. □

Lemma 4.F.10. *We have* $\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] = \frac{1}{2}Y_{11}Y_{12} - \frac{\lambda}{2}(g(\eta_2)Y_{11} + g(\eta_1)Y_{12}) + \frac{\lambda^2}{2}F_2 g(\eta_1)g(\eta_2)$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[h_1(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{12} + Y_{21}Y_{22} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}, Y_{12}] \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] \\
&= \frac{1}{2}\mathbb{E}[Y_{11}Y_{12} + \lambda^2 f(\xi_2)^2 g(\eta_1)g(\eta_2) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] \\
&= \frac{1}{2}Y_{11}Y_{12} + \frac{\lambda^2}{2}F_2 g(\eta_1)g(\eta_2),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[h_2(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] &= \frac{1}{2}\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] \\
&= \frac{1}{2}\mathbb{E}[\mathbb{E}[Y_{11}Y_{22} + Y_{12}Y_{21} \mid \boldsymbol{\xi}, \boldsymbol{\eta}, Y_{11}, Y_{12}] \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] \\
&= \frac{1}{2}\mathbb{E}[\lambda f(\xi_2)g(\eta_2)Y_{11} + \lambda f(\xi_2)g(\eta_1)Y_{12} \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}] \\
&= \frac{\lambda}{2}(g(\eta_2)Y_{11} + g(\eta_1)Y_{12}).
\end{aligned}$$

This proves the result. □

Lemma 4.F.11. *We have* $\mathbb{E}[\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}]^2] = \frac{\lambda^2}{4}F_2 + \frac{\lambda^3}{2}(F_3 - 2F_2 + 1)G_2 + \frac{\lambda^4}{4}(F_4 - 4F_3 + 3F_2^2)G_2^2$.

Proof. We have

$$\begin{aligned}
&\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}]^2 \\
&\qquad = \left(\frac{1}{2}Y_{11}Y_{12} - \frac{\lambda}{2}(g(\eta_2)Y_{11} + g(\eta_1)Y_{12}) + \frac{\lambda^2}{2}F_2 g(\eta_1)g(\eta_2)\right)^2 \\
&\qquad = \frac{1}{4}Y_{11}^2Y_{12}^2 + \frac{\lambda^2}{4}g(\eta_2)^2Y_{11}^2 + \frac{\lambda^2}{4}g(\eta_1)^2Y_{12}^2 + \frac{\lambda^2}{2}g(\eta_1)g(\eta_2)Y_{11}Y_{12} \\
&\qquad\quad + \frac{\lambda^4}{4}F_2^2 g(\eta_1)^2g(\eta_2)^2 - \frac{\lambda}{2}g(\eta_2)Y_{11}^2Y_{12} - \frac{\lambda}{2}g(\eta_1)Y_{11}Y_{12}^2 \\
&\qquad\quad + \frac{\lambda^2}{2}F_2 g(\eta_1)g(\eta_2)Y_{11}Y_{12} - \frac{\lambda^3}{2}F_2 g(\eta_1)g(\eta_2)^2Y_{11} - \frac{\lambda^3}{2}F_2 g(\eta_1)^2g(\eta_2)Y_{12}.
\end{aligned}$$

Taking the expectation of this random variable and using the row-column exchangeability of Y, it becomes

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}]^2] \\
&\qquad = \frac{1}{4}\mathbb{E}[Y_{11}^2Y_{12}^2] + \frac{\lambda^2}{2}\mathbb{E}[g(\eta_2)^2Y_{11}^2] + \frac{\lambda^2}{2}(F_2 + 1)\mathbb{E}[g(\eta_1)g(\eta_2)Y_{11}Y_{12}] \\
&\qquad\quad + \frac{\lambda^4}{4}F_2^2\mathbb{E}[g(\eta_1)^2g(\eta_2)^2] - \lambda\mathbb{E}[g(\eta_2)Y_{11}^2Y_{12}] - \lambda^3F_2\mathbb{E}[g(\eta_1)g(\eta_2)^2Y_{11}].
\end{aligned}$$

We calculate each term of this expression separately, obtaining

$$\begin{aligned}
\frac{1}{4}\mathbb{E}[Y_{11}^2 Y_{12}^2] &= \mathbb{E}[\mathbb{E}[Y_{11}^2 Y_{12}^2 \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \frac{1}{4}\mathbb{E}[(\lambda f(\xi_1) g(\eta_1) + \lambda^2 f(\xi_1)^2 g(\eta_1)^2) \\
&\quad \times (\lambda f(\xi_1) g(\eta_2) + \lambda^2 f(\xi_1)^2 g(\eta_2)^2)] \\
&= \frac{\lambda^2}{4}\mathbb{E}[f(\xi_1)^2 g(\eta_1) g(\eta_2)] + \frac{\lambda^3}{2}\mathbb{E}[f(\xi_1)^3 g(\eta_1)^2 g(\eta_2)] \\
&\quad + \frac{\lambda^4}{4}\mathbb{E}[f(\xi_1)^4 g(\eta_1)^2 g(\eta_2)^2] \\
&= \frac{\lambda^2}{4} F_2 + \frac{\lambda^3}{2} F_3 G_2 + \frac{\lambda^4}{4} F_4 G_2^2, \\
\frac{\lambda^2}{2}\mathbb{E}[g(\eta_2)^2 Y_{11}^2] &= \frac{\lambda^2}{2}\mathbb{E}[\mathbb{E}[g(\eta_2)^2 Y_{11}^2 \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \frac{\lambda^2}{2}\mathbb{E}[g(\eta_2)^2 (\lambda f(\xi_1) g(\eta_1) + \lambda^2 f(\xi_1)^2 g(\eta_1)^2)] \\
&= \frac{\lambda^3}{2} G_2 + \frac{\lambda^4}{2} F_2 G_2^2, \\
\frac{\lambda^2}{2}(F_2 + 1)\mathbb{E}[g(\eta_1) g(\eta_2) Y_{11} Y_{12}] &= \frac{\lambda^2}{2}(F_2 + 1)\mathbb{E}[\mathbb{E}[g(\eta_1) g(\eta_2) Y_{11} Y_{12} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \frac{\lambda^2}{2}(F_2 + 1)\mathbb{E}[\lambda^2 f(\xi_1)^2 g(\eta_1)^2 g(\eta_2)^2] \\
&= \frac{\lambda^4}{2}(F_2 + 1) F_2 G_2^2, \\
\lambda\mathbb{E}[g(\eta_2) Y_{11}^2 Y_{12}] &= \lambda\mathbb{E}[\mathbb{E}[g(\eta_2) Y_{11}^2 Y_{12} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \lambda\mathbb{E}[g(\eta_2)(\lambda f(\xi_1) g(\eta_1) + \lambda^2 f(\xi_1)^2 g(\eta_1)^2)\lambda f(\xi_1) g(\eta_2)] \\
&= \lambda^3 F_2 G_2 + \lambda^4 F_3 G_2^2, \\
\lambda^3 F_2 \mathbb{E}[g(\eta_1) g(\eta_2)^2 Y_{11}] &= \lambda^3 F_2 \mathbb{E}[\mathbb{E}[g(\eta_1) g(\eta_2)^2 Y_{11} \mid \boldsymbol{\xi}, \boldsymbol{\eta}]] \\
&= \lambda^3 F_2 \mathbb{E}[\lambda f(\xi_1) g(\eta_1)^2 g(\eta_2)^2] \\
&= \lambda^4 F_2 G_2^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[h(Y_{\{1,2\},\{1,2\}}) \mid \xi_1, \eta_1, \eta_2, \zeta_{11}, \zeta_{12}]^2] \\
&\quad = \frac{\lambda^2}{4} F_2 + \frac{\lambda^3}{2} F_3 G_2 + \frac{\lambda^4}{4} F_4 G_2^2 + \frac{\lambda^3}{2} G_2 + \frac{\lambda^4}{2} F_2 G_2^2 \\
&\qquad + \frac{\lambda^4}{2}(F_2 + 1) F_2 G_2^2 + \frac{\lambda^4}{4} F_2^2 G_2^2 - \lambda^3 F_2 G_2 - \lambda^4 F_3 G_2^2 - \lambda^4 F_2 G_2^2 \\
&\quad = \frac{\lambda^2}{4} F_2 + \frac{\lambda^3}{2}(F_3 - 2F_2 + 1) G_2 + \frac{\lambda^4}{4}(F_4 - 4F_3 + 3F_2^2) G_2^2,
\end{aligned}$$

which is the expression given by the lemma. □

Chapter 5

Perspectives

Contents

The main contribution of this thesis is the conception of a methodology to perform statistical inference on networks with two main objects, row-column exchangeable matrices and U-statistics. The three previous chapters of this thesis have shown the progressive evolution of this methodology. In the first chapter, a CLT has been derived for U-statistics of quadruplet kernels. The second chapter extends it to kernels of all sizes and proposes a generic estimator for the asymptotic variance. The third chapter gives the CLT in the degenerate case, when the limit is Gaussian. The different examples show the broad range of applications of this methodology.

I have identified three main directions for future works. The most direct direction consists in completing this methodology, especially for degenerate cases. This includes proving the CLT conjectured for degenerate U-statistics when the principal support graphs are not connected and deriving generic estimators for the asymptotic variance of degenerate U-statistics when the limit is Gaussian, similar to the ones in the non-degenerate case. The second direction consists in improving this methodology by adding new theoretical results. This means investigating alternatives approaches and comparing their performances with the existing ones. The use of bootstrap estimators, especially for degenerate U-statistics, is particularly interesting. There

are also other ways to build confidence intervals, for example with concentration inequalities. Finally, Berry-Esseen theorems could add more guarantees to our asymptotic results. The third direction consists in extending my work to other models, which may be more suited to real networks. Notably, two issues with exchangeable models is not only they allow isolated nodes, which are usually not observed or recorded in ecological networks, but they are also dense, while observed networks are commonly sparse.

5.1. Completion of the methodology

The methodology proposed in this thesis is complete for non-degenerate U-statistics. For degenerate U-statistics, the following ideas aim to obtain the remaining building blocks to round it off. They will be explored in priority to be included in a future article, based on Chapter 4.

5.1.1. Non-Gaussian degenerate limit theorem

Chapters 2 and 3 both establish limit theorems for U-statistics, but they do not solve the degenerate cases. Chapter 4 partially fills this gap by transposing the theory of generalized U-statistics of Janson and Nowicki (1991), stating that the limit distribution of U-statistics is determined by properties of its principal support graphs. More precisely, the limit distribution is a polynomial function of independent Gaussian variables of degree b, the maximum number of components in the principal support graphs.

In Janson and Nowicki (1991), proofs for generalized U-statistics are given for the Gaussian limit ($b = 1$), the Chi-squared limit ($b = 2$) and the general case ($b \geq 1$). Chapter 3 has proved the limit theorem for degenerate U-statistics of RCE matrices when $b = 1$, which corresponds to the Gaussian limit. The general case ($b \geq 1$) has only been conjectured. Since the essential part of the theory, the Hoeffding-type decomposition, has already been figured out in Chapter 3, it is expected that this conjecture can be proved with less effort.

5.1.2. Plug-and-play variance estimators for degenerate U-statistics

Chapter 3 suggests a generic estimator of the asymptotic variance of non-degenerate U-statistics. This estimator can be implemented by a simple algorithm and applies to all non-degenerate kernels. Remarkably, it does not require any specific analytic calculation, so in this section, I will be calling it the "plug-and-play" estimator. In the non-degenerate case, the asymptotic variance can be decomposed in two terms, each is the variance of the expectation

of the kernel conditionally to one row or one column. There is a hope that, similar to the non-degenerate case, plug-and-play estimators for the variance can be derived in the degenerate case.

In degenerate case, the terms to estimate are the variance of higher order projections. For example, when the degeneracy order is $d-1$, all the following terms must be estimated, for $(0,0) < (r,c) \leq (p,q)$, $r+c=d$,

$$\mathbb{V}\left[\psi^{r,c}_{(\llbracket r\rrbracket,\llbracket c\rrbracket)}h\right] = \mathbb{V}\left[\mathbb{E}[h_{\llbracket p\rrbracket,\llbracket q\rrbracket} \mid \mathcal{A}_{\llbracket r\rrbracket,\llbracket c\rrbracket}]\right], \tag{5.1}$$

where $\mathcal{A}_{\mathbf{i},\mathbf{j}} = ((\xi_i)_{i\in\mathbf{i}}, (\eta_j)_{j\in\mathbf{j}}, (\zeta_{ij})_{\substack{i\in\mathbf{i}\\ j\in\mathbf{j}}})$. As a reminder, the estimator for non-degenerate cases is the empirical variance of the set of estimators of all the conditional expectations $(\psi^{r,c}_{(\mathbf{i}',\mathbf{j}')}h)_{\mathbf{i}'\in\mathcal{P}_r(\llbracket m_N\rrbracket),\mathbf{j}'\in\mathcal{P}_c(\llbracket n_N\rrbracket)}$, associated to a Hoeffding projection. However, there is one possible obstacle to extending it to degenerate cases. The estimators of the conditional expectations are of the form

$$\widehat{\kappa}_N^{(\mathbf{i}',\mathbf{j}')} := \binom{m_N-r}{p-r}^{-1}\binom{n_N-c}{q-c}^{-1} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}^{p,q}_{N,(\mathbf{i}',\mathbf{j}')}} h_{\mathbf{i},\mathbf{j}}.$$

This is only an average of $O(N^{p+q-d})$ terms, instead of $O(N^{p+q})$ in the non-degenerate case. Therefore, as the order of degeneracy $d-1$ increases, the estimation of the conditional expectations is less precise. Besides, the consistency of the empirical variances of the $(\widehat{\kappa}_N^{(\mathbf{i}',\mathbf{j}')})_{\mathbf{i}'\in\mathcal{P}_r(\llbracket m_N\rrbracket),\mathbf{j}'\in\mathcal{P}_c(\llbracket n_N\rrbracket)}$ has also to be proved, so this requires some calculations.

5.2. Improvements to the methodology

The following ideas might improve the methodology developed in this thesis in two different ways. First, bootstrap methods can substitute current techniques to estimate the variance of our U-statistics and approximate their limit distribution. Second, concentration inequalities and Berry-Esseen theorems give more insights about the distribution of the U-statistics for finite networks. This may help with providing more precise estimates leading to better guarantees in the statistical inference process.

5.2.1. Bootstrap methods

Bootstrap variance estimator The most common method for estimating the variance of U-statistics of i.i.d. observations by bootstrapping. One of the aim of McCullagh (2000) was to show that no bootstrap scheme can generate a consistent estimate of the variance of the mean of a row-column exchangeable matrix. Let Y be a RCE matrix of size $m_N \times n_N$ and

$\bar{Y}_N = \frac{1}{m_N n_N} \sum_{ij} Y_{ij}$. Then,

$$\mathbb{V}[\bar{Y}_N] = \frac{v^{1,0}}{m_N} + \frac{v^{0,1}}{n_N} + \frac{v^{0,0}}{m_N n_N},$$

where $v^{0,0} = \mathrm{Cov}(Y_{11}, Y_{22})$, $v^{1,0} = \mathrm{Cov}(Y_{11}, Y_{12})$ and $v^{0,1} = \mathrm{Cov}(Y_{11}, Y_{21})$. One of the bootstrap scheme suggested by McCullagh (2000) is defined by the following procedure.

- Draw a random sample $(\alpha_i)_{1 \leq i \leq m_N}$ of m_N row indices from $[\![m_N]\!]$ with replacement.
- Draw a random sample $(\beta_j)_{1 \leq j \leq n_N}$ of n_N row indices from $[\![n_N]\!]$ with replacement.
- The bootstrap matrix is defined by $Y_N^b = (Y_{\alpha_i \beta_j})_{1 \leq i \leq m_N, 1 \leq j \leq n_N}$.

This bootstrap scheme is also called the *pigeonhole bootstrap* (Owen, 2007). McCullagh (2000) stated that the variance estimate obtained with this bootstrap scheme has expectation

$$\mathbb{E}[\mathbb{V}[\bar{Y}_N^b]] = \left(\frac{v^{1,0}}{m_N} + \frac{v^{0,1}}{n_N} + \frac{3v^{0,0}}{m_N n_N} \right) (1 + o(1)).$$

McCullagh (2000) assumed that $v^{1,0}$, $v^{0,1}$ and $v^{0,0}$ could vary with N, so in general, this bootstrap scheme is not consistent. However, in our case, $v^{1,0}$, $v^{0,1}$ and $v^{0,0}$ are constant, so $\mathbb{E}\left[\mathbb{V}[\bar{Y}_N^b]\right] / \mathbb{V}[\bar{Y}_N] \xrightarrow[N \to \infty]{} 1$ and the bootstrap estimator is in fact acceptable.

McCullagh (2000) investigated the mean of RCE matrices but he did not consider U-statistics. The recent work of Davezies et al. (2021) extended the pigeonhole bootstrap to means of jointly exchangeable arrays and separately exchangeable arrays, for which they have proved its asymptotic validity. Means of jointly separately exchangeable arrays share many properties with U-statistics of RCE matrices (which are sums of π-echangeable arrays, see Sect. 1.6.1), so there is a reasonable chance that the bootstrap would work for U-statistics of RCE matrices.

Bootstrapping degenerate U-statistics If the validity of a bootstrap scheme was proved for our U-statistics, then not only we could estimate variances, but we could also be able to estimate any function of their limit distribution, e.g. quantiles to directly build confidence intervals. This might be a blessing for the analysis of degenerate U-statistics, because we would not need to know their limit distribution. Instead of trying to identify the form of the complex distribution of a degenerate U-statistic, we could, for instance, hope to directly bootstrap the quantiles of this complex distribution.

However, even if the bootstrap was valid for our U-statistics, some additional work would be required to be able to capture the degenerate limit distribution with bootstrapping methods. In fact, a naïve bootstrap for degenerate U-statistics for i.i.d. observations has been known to fail since Bretagnolle (1983). Without much detail, we give next an intuition about the reason why a naïve bootstrap does not work in degenerate cases.

Let $(X_1, X_2, ...)$ be i.i.d. variables with distribution $P \in \mathcal{P}$. Let U_n be a U-statistic on the variables $(X_1, ..., X_n)$. This U-statistic can be seen as a functional $U_n(P)$ defined on $\mathcal{P}$. For

simplicity, suppose that $\mathbb{E}[U_n(P)] = 0$. Let us define a bootstrap scheme. The bootstrap sample of size n is denoted $(X_1^b, ..., X_n^b)$ where all the X_i^b are i.i.d. with distribution P_n, such that $P_n \xrightarrow[n\to\infty]{} P$. For example, the classic bootstrap scheme is defined by $P_n = \sum_{i=1}^n \delta_{X_i}$ where δ_{X_i} is the Dirac measure at X_i. The principle of the bootstrap is that P is unknown. By naïvely replacing P by P_n, we hope to approximate the distribution of a statistic $T(P)$ by $T(P_n)$. As $P_n \xrightarrow[n\to\infty]{} P$, we hope that the distribution of $T(P_n)$ converges to the distribution of $T(P)$. However, this requires that $T(P)$ is a continuous functional of P, as we illustrate it now.

First, suppose that $U_n(P)$ is non-degenerate. The CLT for U-statistics states that

$$\sqrt{n}U_n(P) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, V_1(P)),$$

where $V_1(P)$ is the asymptotic variance. In some way, the distribution of $\sqrt{n}U_n(P)$ is a continuous function of P, because even if $V_1(P) = 0$, this result would be still valid where "$\mathcal{N}(0,0)$" is the Dirac distribution at 0. Therefore, the bootstrap could work when replacing $\sqrt{n}U_n(P)$ by $\sqrt{n}U_n(P_n)$, in order to approximate $\mathcal{N}(0, V_1(P))$.

Now, note that $nU_n(P)$ diverges unless $V_1(P) = 0$. If $V_1(P) = 0$, then the degeneracy occurs and we have

$$nU_n(P) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{L}_2(V_2(P)),$$

where $\mathcal{L}_2$ is some family of distribution indexed by its variance, and $V_2(P)$ is the asymptotic variance. Therefore, we see that the distribution of $nU_n(P)$ is not a continuous function of P. Because of the discontinuous nature of $nU_n(P)$, a slight change in P might imply a completely different distribution. Therefore, $nU_n(P_n)$ does not approximate $nU_n(P)$ when $U_n(P)$ is degenerate, because $nU_n(P_n)$ is not degenerate in general.

This issue has been tackled by Arcones and Gine (1992) and Dehling and Mikosch (1994), who proposed a workaround. Remember the Hoeffding decomposition established in Section 1.5.2,

$$U_n(P) = \sum_{c=1}^{k} \binom{k}{c} P_n^c(P),$$

where for $1 \le c \le k$, $P_n^c = \binom{n}{c}^{-1} \sum_{\mathbf{i}\in\mathcal{P}_c([\![n]\!])} p^c(X_{\mathbf{i}})$ and p^c is the Hoeffding projection of order c. If $U_n(P)$ is degenerate, then $nP_n^1(P) = 0$ and $nU_n(P) = n\sum_{c=2}^k \binom{k}{c} P_n^c(P)$. Therefore, instead of using $nU_n(P_n)$ as the bootstrap statistic, we may use $n(U_n(P_n) - P_n^1(P_n))$. In this case, the approximated distribution is $n(U_n(P) - P_n^1(P)) = n\sum_{c=2}^k \binom{k}{c} P_n^c(P)$, for which the discontinuity issue disappears.

Adapting such ideas to RCE matrices with the pigeonhole bootstrap is certainly possible, but not trivial. Yet, bootstrapping degenerate U-statistics would provide an alternative approach in my methodology. If we can bootstrap degenerate U-statistics, we only need to identify the order

of degeneracy $d-1$ and use the correct bootstrap statistic $N^{d/2}(U_N - \sum_{r+c\leq d} P_N^{r,c})$ to approximate the limit distribution, instead of identifying this distribution analytically.

5.2.2. Non-asymptotic results

In the proposed methodology, the asymptotic confidence intervals used are derived from limit theorems of the type $\sqrt{N^d/V}(\widehat{\theta}_N - \theta) \xrightarrow[N\to\infty]{\mathcal{D}} W$ where V is the asymptotic variance and W is a random variable of variance 1. With an estimator $\widehat{V}_N$ for V, an asymptotic confidence interval is

$$I_N(\alpha) = \left[\widehat{\theta}_N - q_{1-\alpha/2}\sqrt{\frac{\widehat{V}_N}{N^d}}, \widehat{\theta}_N - q_{\alpha/2}\sqrt{\frac{\widehat{V}_N}{N^d}}\right],$$

where q_x denotes the quantile of order x of W. We have

$$\mathbb{P}(\theta \in I_N(\alpha)) \xrightarrow[N\to\infty]{} 1-\alpha.$$

However, for a finite value of N, we have no guarantee on accuracy of $I_N(\alpha)$, e.g. we do not know $\mathbb{P}(\theta \in I_N(\alpha))$.

Concentration inequalities give bounds on the deviation of a random variable from its mean. For an estimator $\widehat{\theta}_N$ of θ, they are results of the type

$$\mathbb{P}(|\widehat{\theta}_N - \theta| \geq t) \leq \phi(t, N),$$

where ϕ is some function with values in $[0,1]$. This inequality is valid for all values of N. Suppose that for some $N \geq 1$ and $\alpha \in]0,1[$, there is a value $t_{\alpha,N}$ such that $\phi(t_{\alpha,N}, N) \geq \alpha$. In this case, the interval

$$J_N(\alpha) = \left[\widehat{\theta}_N - t_{\alpha,N}, \widehat{\theta}_N + t_{\alpha,N}\right]$$

is such that

$$\mathbb{P}(\theta \in J_N(\alpha)) \geq 1-\alpha.$$

This type of intervals can be used instead of the asymptotic confidence intervals used in my methodology.

However, there are several potential caveats to this approach. First, although the guarantees of $J_N(\alpha)$ are clearer than the guarantees of $I_N(\alpha)$, it does not necessarily mean that $\mathbb{P}(\theta \in J_N(\alpha))$ is closer to $1-\alpha$ than $\mathbb{P}(\theta \in I_N(\alpha))$ is. In fact, we can only conclude when $J_N(\alpha) \subset I_N(\alpha)$, because in this case, we have $1-\alpha \leq \mathbb{P}(\theta \in J_N(\alpha)) \leq \mathbb{P}(\theta \in I_N(\alpha))$. Second, finding the value $t_{\alpha,N}$ to build $J_N(\alpha)$ might not be obvious. Most of the time, it can only be estimated with an estimator $\widehat{t}_{\alpha,N}$. Replacing $t_{\alpha,N}$ with $\widehat{t}_{\alpha,N}$ to obtain the interval $\widehat{J}_N(\alpha)$ usually introduces an asymptotic approximation. Therefore, $\mathbb{P}(\theta \in \widehat{J}_N(\alpha))$ is only asymptotically bounded. For this

reason, in practice, intervals built from concentration inequalities may be decent alternatives to the asymptotic confidence intervals used in my methodology, but contrarily to what one might hope, they do not always have non-asymptotic guarantees.

To illustrate this, one simple useful concentration inequality is Chebyshev's inequality. If $\widehat{\theta}_N$ is unbiased, then we have

$$\mathbb{P}(|\widehat{\theta}_N - \theta| \geq t) \leq \frac{\mathbb{V}[\widehat{\theta}_N]}{t^2}.$$

In this case, $\mathbb{V}[\widehat{\theta}_N]$ must be estimated. I have investigated the use of this inequality to build confidence intervals, but I have found that such confidence intervals are usually less conservative than the ones currently used in my methodology.

Because $\widehat{\theta}_N$ is made of U-statistics, one hopes to be able to derive Bernstein-type and Hoeffding-type inequalities. These inequality exploit the sum structure of random variables such as U-statistics. For U-statistics of i.i.d. observations with bounded kernels h of size k, Hoeffding (1963) has proved these two inequalities

$$\mathbb{P}(|U_N - U_\infty| \geq t) \leq \exp\left(-\frac{2\lfloor N/k \rfloor t^2}{(b-a)^2}\right),$$

$$\mathbb{P}(|U_N - U_\infty| \geq t) \leq \exp\left(-\frac{\lfloor N/k \rfloor t^2}{2\sigma^2 + \frac{2}{3}(b-a)t}\right),$$

where $\sigma = \mathbb{V}[h(X_1, ..., X_k)]$ and $a \leq h(X_1, ..., X_k) \leq b$. In future works, it might be possible to generalize these inequalities to U-statistics of RCE matrices. However, both inequalities require the kernel to be bounded, which is a strong restriction, unfit to some applications.

5.2.3. Berry-Esseen theorems

When a sum of random variables converges weakly to a standard Gaussian variable, a Berry-Esseen theorem gives a bound to the approximation error between the distribution of the sum and the Gaussian limit. For a given limit theorem, several Berry-Esseen bounds might exist in the literature because most of the time, the form of Berry-Esseen bounds heavily depends on the moments of the random variables. This is especially true for U-statistics of i.i.d. observations, for which the successive works of Grams and Serfling (1973), Bickel (1974), Chan and Wierman (1977) and Callaert and Janssen (1978) explored the assumptions on the moments of the random variables to find the sharpest bound. As an example, I am giving the typical form of a Berry-Esseen result for sums of i.i.d. variables. Let $X_1, ..., X_n$ be i.i.d. random variables with moments $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = \sigma^2$ and $\mathbb{E}[|X_1|^3] = \beta^3$. Let $S_n = (\sigma\sqrt{n})^{-1} \sum_{i=1}^n X_i$. Then, we have

$$\sup_x |\mathbb{P}(S_n \leq x) - \Phi(x)| \leq \frac{C\beta^3}{\sigma^3\sqrt{n}}, \tag{5.2}$$

where Φ is the cumulative distribution function of a standard Gaussian variable and C is a constant.

Similar results have been established for U-statistics of jointly exchangeable arrays by van Zwet (1984) and Friedrich (1989). Recently, many additional Berry-Esseen bounds have been found with proofs revolving around the application of Stein's method (Zhang, 2022; Austern and Orbanz, 2022). Stein's method also allows to identify Berry-Esseen bounds for non-Gaussian limits (Chatterjee and Shao, 2011; Shao and Zhang, 2019; Han and Kato, 2022), although I have not found any for degenerate U-statistics. I have not conducted a thorough literature review on Berry-Esseen bounds, so applicable results might already exist for our U-statistics. However, even though a Berry-Esseen theorem may indicate how quickly a U-statistic converges to its limiting distribution, it is hard to exploit the bound in our framework. The first potential difficulty is the calculation of the constants in the bound, which are not often obvious. The other issue is that the approximation error cannot be directly used to correct the confidence intervals built in my methodology. In constrast, one possibility is to use the bound on the approximation error to build confidence intervals for p-values or for the upper and lower bounds of the asymptotic confidence intervals used in the methodology.

5.3. Beyond RCE models

The random network models used in my work are only required to be exchangeable and dissociated. The dissociation assumption is not very restrictive, but the exchangeability assumption is a strong one. Exchangeability is mathematically very convenient to derive the theoretical results of this thesis. In the introductory chapter, I have explained how exchangeability is a reasonable assumption for ecological networks. However, it is also legitimate to wonder if exchangeable models represent well real-life networks. Without condemning exchangeability, the following two ideas discuss two potential caveats of exchangeable models while providing guidance for future works to adapt my methodology to other models, which are better at representing ecological networks.

5.3.1. Missing species

Ecological interaction networks result from the aggregation of individual interactions. This means that every species in the dataset has been seen interacting at least once. In this respect, all RCE models share a caveat when it comes to representing real networks because they allow nodes to be isolated, i.e. rows and columns of the adjacency matrices to be empty. Let us discuss this for binary networks (the same reasoning can be applied for valued networks).

For binary networks, we recall the definition of the graphon model, which encompasses all dissociated RCE models. An adjacency matrix Y of size $m_N \times n_N$ is generated by the model $G(N, w)$ means

$$\begin{aligned}
&\xi_i \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq i \leq m_N,\\
&\eta_j \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq j \leq n_N,\\
&Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{B}(w(\xi_i, \eta_j)), && \forall 1 \leq i \leq m_N, 1 \leq j \leq n_N,
\end{aligned}$$

where $w : [0,1]^2 \to [0,1]$. For all $1 \leq i \leq m$, the probability of the i-th row to be empty is

$$\begin{aligned}
\mathbb{P}(\sum_{j=1}^{n_N} Y_{ij} = 0) &= \iint \mathbb{P}(Y_{i1} = 0, ..., Y_{in} = 0 \mid \xi_i, \eta_1, ..., \eta_n) d\xi_i d(\eta_1, ..., \eta_n)\\
&= \iint \prod_{j=1}^{n_N} \mathbb{P}(Y_{ij} = 0 \mid \xi_i, \eta_j) d\xi_i d(\eta_1, ..., \eta_n)\\
&= \int \left(\int \big(1 - w(\xi_i, \eta)\big) d\eta \right)^{n_N} d\xi_i.
\end{aligned}$$

Since $\int \big(1 - w(\xi_i, \eta)\big) d\eta > 0$ unless $w(\xi_i, \eta) = 1$ for all $\eta \in [0,1]$, we have $\mathbb{P}(\sum_{j=1}^{n_N} Y_{ij} = 0) > 0$ unless $w \equiv 1$.

An observed network cannot have been generated by a random model allowing isolated nodes. It is hard to quantify how much the presence of isolated nodes affects the results of the analysis. I have thought about two approaches to investigate the issue. One could either consider that the observed adjacency matrix only consists in the non-empty rows and columns of the network generated by the model, or one could change the models to make them generate only adjacency matrices with no empty rows or columns.

This second solution is rather difficult to implement. Certainly, in principle, it is easy to simulate networks following a graphon model conditioned to having no isolated nodes. It suffices to generate candidate networks from a graphon model until one satisfies this condition. However, the inclusion of this constraint to the graphon model makes the probability distribution complex and unlikely to be tractable. The row-column exchangeability property is also lost, which means that the convergence results on U-statistics of this thesis no longer apply. Later, we will introduce the graphex models, a class of models solving somes issues.

The first solution is a simpler approach for future works. The observed networks are considered as subnetworks of the networks generated by the models, but where isolated nodes are removed. Inversely, we simply need to add these empty rows and columns back to the observed adjacency matrices before performing statistical inference with the proposed methodology. However, the tricky question is to determine how many empty rows and columns should be added. One promising idea is to look into the literature of species richness estimation.

Richness estimation is one motivation of ecological surveys. The list of species recorded in the data may only consist in a subset of the actual species present in the area. Many techniques have been developed to infer the number of missing species in presence-absence or abundance data. Non-parametric methods based on resampling methods (jackknife, bootstrap) have been developed (Burnham and Overton, 1978; Heltshe and Forrester, 1983; Smith and van Belle, 1984). However, their variance is an issue and they seem to heavily underestimate or overestimate the number of missing species. In contrast, Chao (1984) gives an estimator of a lower bound for the number of missing species in abundance dat. The so-called Chao's estimator has been found to be rather statistically robust (Bunge and Fitzpatrick, 1993; Walther and Moore, 2005).

Once the numbers of missing rows and missing columns (here denoted m_0 and n_0) are estimated, is it tempting to use them to correct the networks. One may complete the adjacency matrix by adding $\widehat{m}_0$ empty rows and $\widehat{n}_0$ empty columns and perform the usual statistical inference on the completed matrix. However, since $\widehat{m}_0$ and $\widehat{n}_0$ are estimates, they may not correspond to the true numbers of missing rows and columns (m_0 and n_0). Therefore, this requires that the potential error of estimation is propagated in my methodology. For example, if confidence intervals are available for m_0 and n_0, then one might explore how changes in these values affect the estimation of the quantity of interest, when these different values within the confidence intervals are used to complete the matrix instead. Alternatively, all these couples of possible values for (m_0, n_0) can be considered as different models in a Model Averaging framework (Raftery et al., 1997; Wasserman, 2000), which accounts for the estimation uncertainty.

5.3.2. Sparse networks

On one hand, we have just seen that ecological interaction networks do not usually feature isolated nodes. On the other hand, these networks have been found to be sparse, i.e. their connectance decreases when they have more nodes (Busiello et al., 2017). This seems to be in contradiction with the exchangeability assumption, which ensures that the connectance remains constant when the network size increases. Many generalizations of RCE models to sparse networks exist, to which there is a hope that my methodology can be adapted.

Before introducing some models of sparse networks, I would like to emphasize that non-sparsity does not fatally invalidate our approach. In reality, one driver of sparsity might be species abundance, which greatly affects the structure of network. Rare species are less likely to be sampled, therefore, they appear in fewer interactions of a dataset. In observed networks, these rare species appear to be specialists (Fründ et al., 2016), i.e. they have only few interactions. Abundant species are always sampled first, but as the sampling effort grows, there comes a point after which every new species added is a rare species. This might explain the apparent

sparsity in ecological networks. Although the observed networks are sparser, it is possible that the complete network, showing all the interactions actually occurring, including missing links, is a dense network.

Sparse graphons A graphon w can be written $w = \lambda\bar{w}$ where $\lambda = \iint w$ is the density of the network and $\iint \bar{w} = 1$. When the network dimensions m_N and n_N increase, it is natural to replace the constant density by a sequence $\lambda = (\lambda_N)_{N\geq 1}$, where $\lambda_N \to 0$ when $N \to \infty$. The *sparse graphon* model $G(N, \bar{w}, \lambda)$ can be defined as

$$\begin{aligned}
&\xi_i \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq i \leq m_N,\\
&\eta_j \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq j \leq n_N,\\
&Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{B}(\lambda_N \bar{w}(\xi_i, \eta_j)), && \forall 1 \leq i \leq m_N, 1 \leq j \leq n_N.
\end{aligned}$$

This model has been proposed and studied by Bickel et al. (2011), Wolfe and Olhede (2013) and Klopp et al. (2017) for unipartite networks. Significant works by Bhattacharyya and Bickel (2015), Levin and Levina (2019), Green and Shalizi (2022) and Lunde and Sarkar (2023) investigated motif densities, which are usually U-statistics, but for sparse networks, they are renormalized by λ_N and estimated by resampling methods instead. Currently, no general theory of U-statistics exist for this family of models. It is expected that λ_N affects the behavior of U-statistics. Therefore, either U-statistics should be renormalized by λ_N to cancel it out, or one should study the effect of λ_N on the leading terms of the Hoeffding-type decomposition, but this provides a basis for future works.

Graphex model Another recent series of models addressing the issue of sparsity has been developed around discrete exchangeable measures on $\mathbb{R}_+^2$ (Herlau et al., 2016; Caron and Fox, 2017; Todeschini et al., 2020). The associated class models has been theoretically studied under the name of *graphex models*, *Kallenberg exchangeable graphs* or *graphon processes* by Veitch and Roy (2015) and Borgs et al. (2016). To distinguish them from these new models, we qualify our exchangeable models as "dense". Instead of using dense exchangeable adjacency matrices, these models construct graphs from an exchangeable point process on $\mathbb{R}_+^2$. Instead of the Aldous-Hoover representation, exchangeable point processes on $\mathbb{R}_+^2$ admit a representation as a function of uniform random variables and unit rate Poisson point processes (Kallenberg, 2005). Instead of the graphon model encompassing all the dense exchangeable models, the graphex model encompasses all the models arising from these exchangeable measures.

In the formalism of Veitch and Roy (2015), a graphex consists in a triple (I, S, W), where $I \in \mathbb{R}_+$ is a non-negative number, $S : \mathbb{R}_+ \to \mathbb{R}_+$ is an integrable function and $W : \mathbb{R}_+^2 \to [0,1]$ is a symmetric function. Sacrificing generality for the sake of clarity, we describe, with the help of

Figure 5.1, a particular case of the graphex model where $I = 0$ and $S \equiv 0$. Let Π be a unit rate Poisson process on $\mathbb{R}_+^2$. Each point sampled Z_i is latent and has coordinates (θ_i, ϑ_i). The value θ_i is equivalent to a label for the point and ϑ_i is a latent parameter. Each pair of points (Z_i, Z_j) is included in the graph as an edge with probability $W(\vartheta_i, \vartheta_j)$. To obtain a finite graph, a value $\nu \in \mathbb{R}_+$ is set and only the edges with labels $\theta_i, \theta_j < \nu$ are kept, alongside with the vertices that participate in at least one edge.

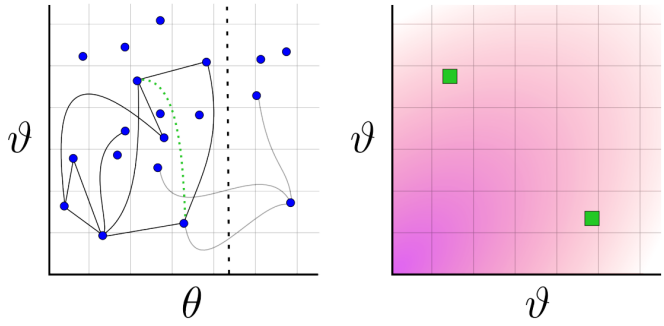

Figure 5.1 – An illustration of the graph building process of the graphex $(0, 0, W)$. The points are the sampled from the unit rate Poisson process on $\mathbb{R}_+^2$, with coordinates (θ_i, ϑ_i) (left). The green dashed edge is included with probability given by the function W taken on one of the green squares (right, it is a symmetric function). After all the pairs of points have been explored the same way, the black edges have been included to the graph. The finite graph is obtained by setting a value ν, here $\nu = 4.2$. The grey edges have been excluded because they involve at least a point with $\theta_i > \nu$. Finally, the finite graph consists in the points participating in at least one (black) edge. Figure taken from Veitch and Roy (2015).

From that perspective, we see that the graphex model explicitly excludes isolated vertices. Also, Veitch and Roy (2015) proved that the graphex model generates sparse networks in general, if W is integrable and does not have a compact support. In this case, the graphex model would solve the questions of both sparsity and isolated vertices. It is interesting to note that otherwise, if the support of the symmetric function W is compact, then the graphs generated by the graphex models are dense. In that case, the graphex model actually becomes a graphon model conditioned to exclude isolated vertices, which was actually what we were looking for in a previous paragraph. Therefore, the flexibility of the graphex model combined with its desirable property to exclude isolated vertices indicates that it may be appropriate to represent real ecological networks.

The graphex model is a sparse network model, with some probabilistic symmetries, although the exchangeability assumption is weaker than in the graphon model. Some work is required before using these models to investigate ecological networks. First, only few models of this class

have been built for binary bipartite graphs (Caron, 2012; Veitch et al., 2017), and not in the general framework of the graphex model as Veitch and Roy (2015) defined it. Furthermore, I have found no reference to weighted or valued graphex models. Second, the probability distribution of the graphex model is not as simple as the one of the graphon model. The adjacency matrix of a graphex network is not exchangeable in general. Still, the probabilistic symmetries of the underlying process can be used to investigate the distribution of network statistics. U-statistics appear to have been only used to count motifs (Caron et al., 2017). The asymptotic behavior of U-statistics can be investigated through the point process structure (Reitzner and Schulte, 2013) of the network. One can also think about other quantities of interest and maybe other types of statistics, for which we hope to be able to identify the limit distribution using the exchangeability property of the graphex model.

In this chapter, a number of ideas have been suggested for further work. They have been organised in such a way that the former are direct extensions of my work, complementing or improving my methodology, while the latter are more exploratory. The first ideas can be developed immediately, while the others are better suited to a future research plan over several years.

Chapitre 6

Résumé

The following chapter is a summary of the thesis in French.

Sommaire

Cette thèse présente les travaux que j'ai menés pendant trois ans en tant que doctorant au MIA Paris-Saclay, sous la direction de Stéphane Robin, Sophie Donnet et François Massol. Ce chapitre résume les différentes contributions scientifiques après avoir présenté le contexte dans lequel elles s'inscrivent.

6.1. Contexte

Cette thèse porte essentiellement sur le comportement asymptotique des U-statistiques sur des matrices échangeables ligne-colonne (*row column exchangeable*, RCE). Ces travaux sont principalement théoriques, mais ils sont fortements motivés par des applications à l'analyse de réseaux écologiques. En effet, en écologie, la plupart des études ne se limitent pas à l'analyse

d'un seul réseau, mais d'une collection de réseaux échantillonnés sous différentes conditions, à des endroits ou des moments différents. En analysant les réseaux de la collection de manière conjointe, on espère obtenir des informations sur la variabilité des réseaux dans l'espace, dans le temps et sur l'influence des perturbation extérieures. Cependant, il existe peu de méthodologie générale pour étudier cela. Dans la littérature écologique, les méthodes utilisées sont spécifiques aux données disponibles et à la problématique biologique étudiée. L'objectif de cette thèse est de proposer une méthode plus générale pour étudier les réseaux écologiques.

Il existe de nombreux types de réseaux en écologie. Les réseaux les plus étudiés sont les réseaux d'interaction entre espèces. Ces réseaux représentent le plus souvent un seul type d'interaction particulier. Nous avons choisi ici de nous concentrer sur les réseaux bipartites, ayant deux types de nœuds tels que seuls deux nœuds de types différents peuvent être reliés, comme les réseaux de pollinisation (Fig. 6.1). Le point de départ de ces travaux a donc été de développer une méthode de comparaison pour ces réseaux. En comparaison aux réseaux rencontrés dans d'autres domaines scientifiques, les réseaux écologiques ont des propriétés spécifiques.

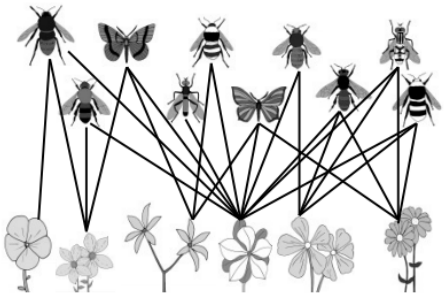

FIGURE 6.1 – Un réseau de pollinisation binaire relie des plantes aux insectes qui les visitent. Figure tirée de Fontaine (2009).

Tout d'abord, ces réseaux résultent de l'agrégation de données d'interaction individuelles. Selon la manière dont ces données sont agrégées, différents réseaux peuvent être obtenus (Fig. 6.2). Par exemple, les interactions peuvent être regroupées à des niveaux taxonomiques différents (par espèce, genre, famille, etc.), sur des périodes différentes (par jour, mois, année, etc.), sur des zones géographiques différentes (par site d'observation, région, etc.).

La méthode d'échantillonnage des interactions affecte également la topologie des réseaux. Par exemple, dans les réseaux de pollinisation, un insecte peut être échantillonné comme interagissant avec seulement une espèce de plante, par exemple s'il est capturé à l'aide d'un filet après une visite. Le même insecte peut également être observé comme ayant visité plusieurs fleurs de cette même plante, ce qui donne lieu soit à plusieurs interactions, au lieu d'une seule, soit à une interaction quantifiée. Enfin, cet insecte aurait pu être échantillonné comme ayant interagi avec différentes plantes, si les interactions sont analysées par le biais du pollen trouvé sur l'insecte.

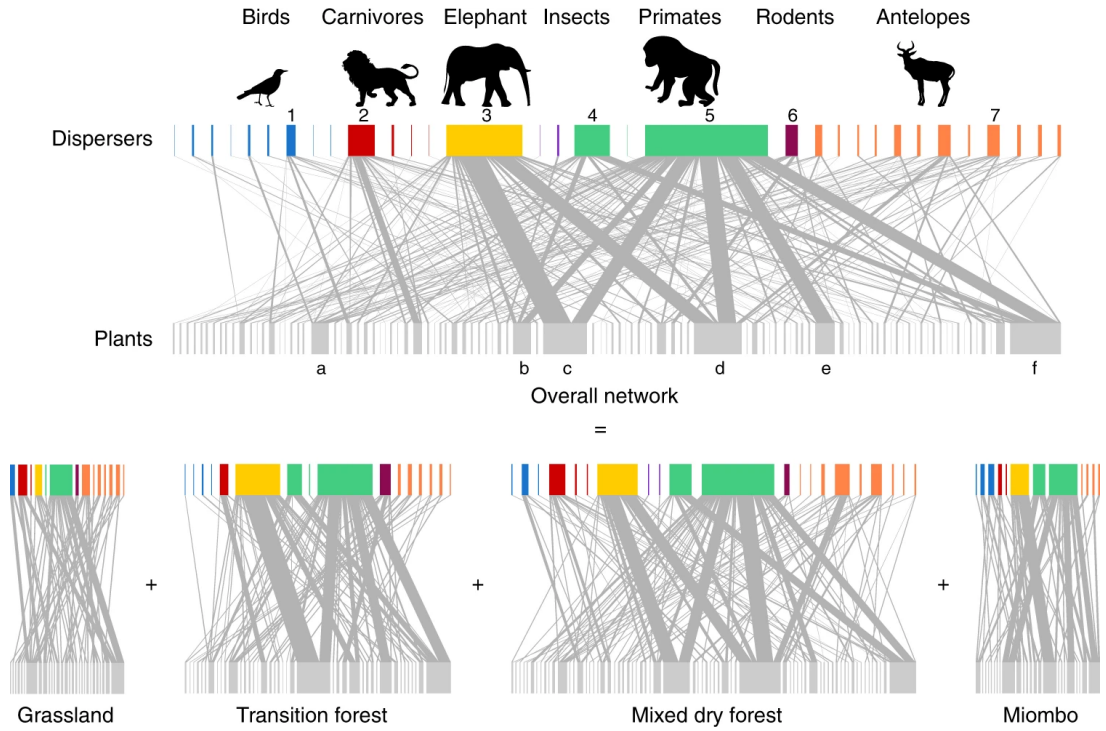

FIGURE 6.2 – La question de l'agrégation des données est récurrente dans les réseaux écologiques. Haut : un réseau de dispersion de graines. Bas : les données d'interaction ont été divisées par le type de végétation, ce qui amène à des structures différentes du réseau agrégé. Figure tirée de Timóteo et al. (2018).

Bien qu'il s'agisse du même individu, les interactions peuvent donc être échantillonnées de différentes manières, ce qui conduit à des données différentes. La liste des nœuds au réseau, la liste des arêtes, ainsi que la valeur des arêtes peuvent être affectées.

Surtout, l'effort d'échantillonnage joue un rôle crucial dans les réseaux d'interaction écologiques. Le réseau reconstruit n'est souvent qu'une fraction du réseau complet, et il est difficile de savoir quel effort est nécessaire pour avoir la certitude d'avoir suffisamment échantillonné l'ensemble du réseau. Il existe donc une notion d'incertitude inhérente aux données.

Enfin, en raison de la question de l'échantillonnage ou, plus simplement, parce qu'ils sont observés dans des lieux ou à des moments différents, les réseaux d'une collection n'impliquent pas nécessairement les mêmes espèces. Par conséquent, les réseaux peuvent avoir des ensembles de nœuds différents, potentiellement de tailles différentes.

L'approche que j'ai utilisée pour concevoir une méthodologie a donc été motivée par trois lignes directrices. Premièrement, la méthode doit non seulement permettre de caractériser un réseau individuellement, mais surtout d'analyser plusieurs réseaux conjointement. Par exemple, une telle méthode doit pouvoir être utilisée pour comparer des réseaux. Deuxièmement, elle

doit prendre en compte les spécificités des réseaux d'interaction écologiques, notamment les caractéristiques décrites ci-dessus. Enfin, elle doit pouvoir répondre à des questions écologiques. En effet, pour ce dernier point, il est facile de dire si deux réseaux observés sont identiques, mais il est plus difficile de savoir dans quelle mesure ils diffèrent, si ces différences sont significatives et surtout, comment les interpréter d'un point de vue écologique.

La méthodologie élaborée dans cette thèse repose sur deux principaux piliers : les modèles de graphes aléatoires échangeables et les U-statistiques. Le premier, les modèles de graphes aléatoires, permet de considérer que chaque réseau observé est la réalisation d'un certain modèle probabiliste. Cela permet de capturer la source de variabilité des réseaux qui n'est due qu'au hasard. L'hypothèse d'échangeabilité est une hypothèse portant sur les espèces présentes dans un réseau. Les modèles échangeables permettent d'obtenir des propriétés intéressantes pour les U-statistiques, qui constituent le deuxième pilier de cette méthodologie. Ces dernières forment une classe de statistiques généralisant le concept de moyenne empirique et seront principalement utilisées comme estimateurs. Pour développer ma méthodologie, j'ai établi des propriétés théoriques pour les objets mathématiques concernés, c'est-à-dire les modèles de graphes aléatoires échangeables et les U-statistiques sur des matrices générées par ces modèles. Néanmoins, je n'ai jamais perdu de vue la visée applicative initiale.

6.2. Cadre général

6.2.1. Modèles de réseaux bipartites échangeables

Un réseau est dit aléatoire si le graphe qui le représente est une variable aléatoire. Cela signifie que les graphes des réseaux observés résultent d'une expérience aléatoire. Un modèle de réseau (ou de graphe, on utilisera désormais les deux termes de manière équivalente) aléatoire définit la loi de cette variable aléatoire. Il est parfois plus simple de définir un modèle en définissant la loi d'une matrice, qui peut être considérée comme la matrice d'adjacence d'un graphe.

Un modèle de réseau bipartite est dit échangeable si la loi de la matrice est invariante par permutations séparées des lignes et des colonnes. On dit que la matrice d'adjacence est échangeable ligne-colonne (row column-exchangeable, RCE). On appellera ces modèles des *modèles RCE*.

La forme générale des modèles RCE est donnée par la classe des graphons bipartites colorés. Un graphon bipartite coloré est une application $\mathcal{W} : [0,1]^2 \to \Pi(E)$, où $\Pi(E)$ est l'ensemble des lois de probabilité sur E. Le modèle associé peut s'écrire de la forme suivante, pour générer la

matrice d'adjacence Y d'un réseau ayant m nœuds en ligne et n nœuds en colonne :

$$\begin{aligned} \xi_i &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq i \leq m, \\ \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq j \leq n, \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{W}(\xi_i, \eta_j), && \forall 1 \leq i \leq m, 1 \leq j \leq n. \end{aligned}$$

Dans cette thèse, une sous-classe des graphons colorés va être considérée. Les graphons colorés étudiés seront de la forme $\mathcal{W}(\cdot,\cdot) = \mathcal{L}(w(\cdot,\cdot))$, où $w : [0,1]^2 \to \mathbb{R}$ et $(\mathcal{L}(\mu))_{\mu\in\mathbb{R}}$ est une famille de lois de probabilité avec un paramètre unique. Les modèles associés sont donc

$$\begin{aligned} \xi_i &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq i \leq m, \\ \eta_j &\overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \leq j \leq n, \\ Y_{ij} \mid \xi_i, \eta_j &\sim \mathcal{L}(w(\xi_i, \eta_j)), && \forall 1 \leq i \leq m, 1 \leq j \leq n. \end{aligned}$$

Par la suite, la fonction w sera simplement appelée un graphon et on se référera aux *modèles graphons* pour désigner cette sous-classe.

Le modèle à blocs latents

Le modèle à blocs latents (latent block model, LBM, Govaert and Nadif, 2003) est un modèle supposant que les deux ensembles de nœuds des graphes bipartites peuvent être divisés séparément en plusieurs groupes. La loi de l'interaction entre deux nœuds est déterminée par les groupes auxquels appartiennent ces nœuds. En général, cette loi est paramétrée par un paramètre unique dépendant de ces groupes. La forme hiérarchique du LBM est donnée par la formule

$$\begin{aligned} Z_i &\overset{iid}{\sim} \mathcal{M}(1;\boldsymbol{\alpha}), && \forall 1 \leq i \leq m, \\ W_j &\overset{iid}{\sim} \mathcal{M}(1;\boldsymbol{\beta}), && \forall 1 \leq j \leq n, \\ Y_{ij} \mid Z_i = k, W_j = \ell &\sim \mathcal{L}(\pi_{k\ell}), && \forall 1 \leq i \leq m, 1 \leq j \leq n, \end{aligned}$$

où α et β sont des vecteurs de probabilités de tailles K_1 et K_2, les nombres respectifs de groupes pour les nœuds en ligne et les nœuds en colonne, $(\mathcal{L}(\theta))_{\theta\in\Theta}$ est une famille de lois de probabilité et $\boldsymbol{\pi} \in \mathcal{M}_{K_1,K_2}(\Theta)$ est une matrice de $K_1 \times K_2$ paramètres (Fig. 6.3).

Le LBM peut s'écrire comme un graphon constant par blocs, avec

$$w(\xi_i, \eta_j) = \sum_{k,\ell} \pi_{k\ell} \mathbb{1}\{s(\xi_i) = k\} \mathbb{1}\{t(\eta_j) = \ell\},$$

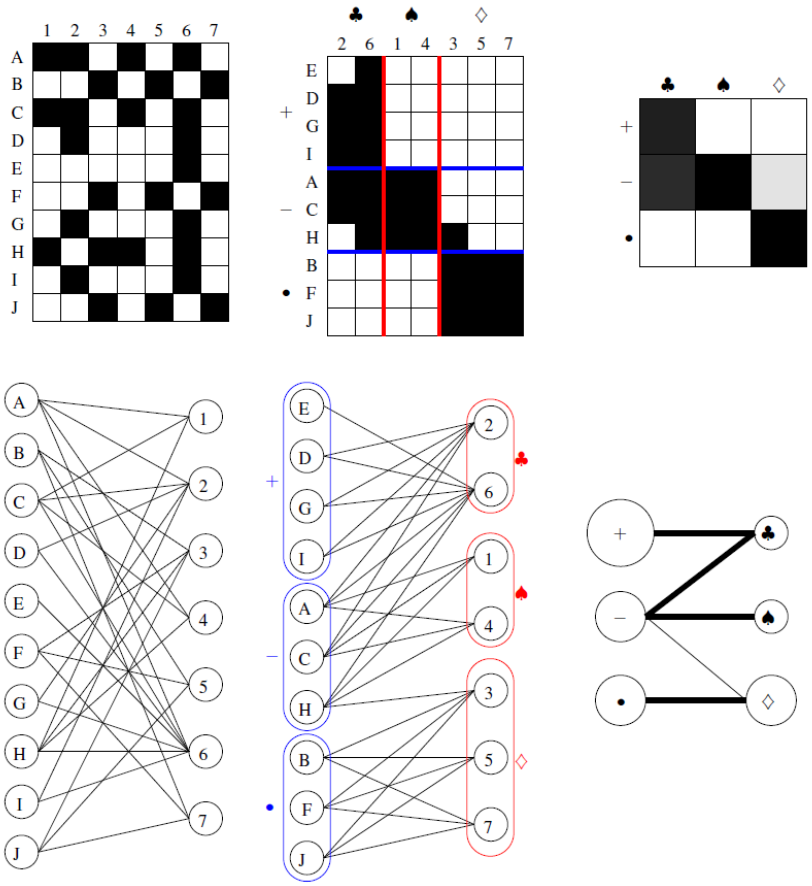

FIGURE 6.3 – Le modèle à blocs latents (LBM). Gauche : un réseau bipartite et sa matrice d'adjacence. Milieu : le réseau et la matrice réordonnés selon l'appartenance des nœuds aux groupes. Droite : un réseau "résumé", représentant la matrice $\boldsymbol{\pi}$ comme un réseau. Figure tirée de Brault (2014).

où les fonctions s et t sont constantes par morceaux :

$$s(\xi_i) = 1 + \sum_{k=1}^{K} \mathbb{1}\left\{\xi_i > \sum_{k'=1}^{k} \alpha_{k'}\right\}, \qquad \text{and} \qquad t(\eta_j) = 1 + \sum_{\ell=1}^{L} \mathbb{1}\left\{\eta_j > \sum_{\ell'=1}^{\ell} \beta_{\ell'}\right\}.$$

Parfois, pour spécifier la loi $\mathcal{L}$ utilisée dans le modèle, on peut écrire $\mathcal{L}$-LBM, par exemple Bernoulli-LBM ou Poisson-LBM.

Modèle à distributions de degrés attendus bipartite

Le modèle à distributions de degrés attendus bipartite (bipartite expected degree distribution model, BEDD, Ouadah et al., 2022) utilise deux distributions de degrés, une pour chaque

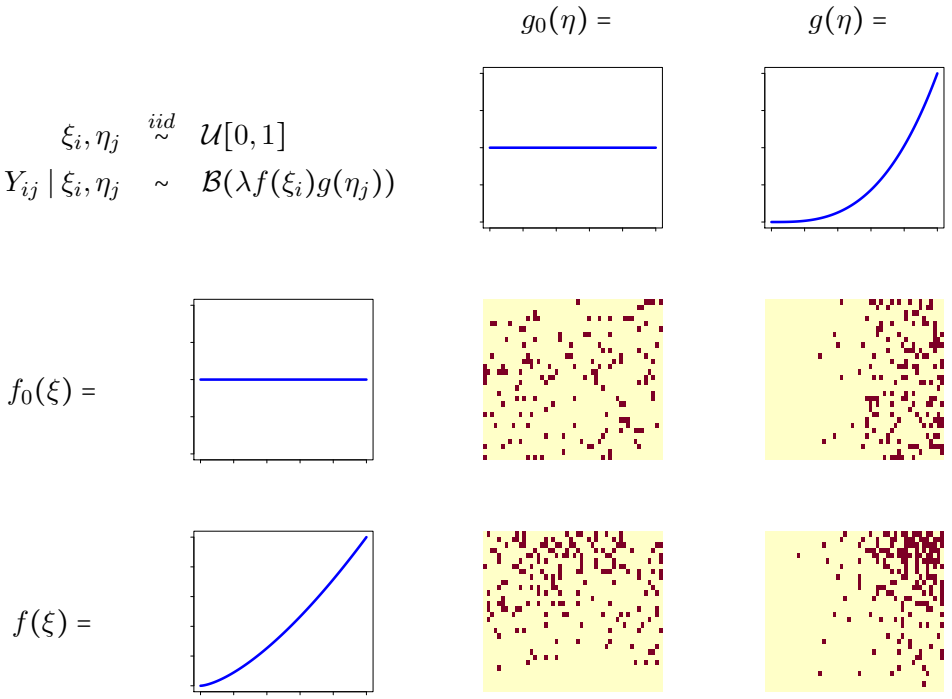

FIGURE 6.4 – Le modèle à distributions de degrés attendus bipartite (BEDD). La figure représente plusieurs matrices d'adjacence (réordonnées) générées par des BEDD binaires avec des distributions de degrés constantes (f_0, g_0) ou quelconques (f, g).

ensemble de nœuds (en ligne ou en colonne). Il peut être écrit de la manière suivante

$$\begin{aligned} &\xi_1, ..., \xi_m \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \le i \le m, \\ &\eta_1, ..., \eta_n \overset{i.i.d.}{\sim} \mathcal{U}[0,1], && \forall 1 \le j \le n, \\ &Y_{ij} \mid \xi_i, \eta_j \sim \mathcal{L}(\lambda f(\xi_i) g(\eta_j)), && \forall 1 \le i \le m, 1 \le j \le n, \end{aligned}$$

où $(\mathcal{L}(\mu))_{\mu \in \mathbb{R}}$ est une famille de lois de probabilité paramétrisée par leur espérance μ, $\lambda \in \mathbb{R}$ est la densité du graphe, f et g sont des fonctions de $[0,1] \to \mathbb{R}$ telles que $\int f = \int g = 1$.

Avec cette formulation, les distributions de degrés sont caractérisées par les fonctions f et g (Fig. 6.4). Plus précisément, les fonctions f et g peuvent être vues comme les inverses généralisées des fonctions de répartition des distributions de degrés renormalisés, puisque $\mathbb{E}[n^{-1} \sum_{j=1}^{n} Y_{ij} \mid \xi_i] = \lambda f(\xi_i)$ et $\mathbb{E}[m^{-1} \sum_{i=1}^{m} Y_{ij} \mid \eta_j] = \lambda g(\eta_j)$. Le BEDD est un modèle graphon, où le graphon a une forme produit $w(\xi, \eta) = \lambda f(\xi) g(\eta)$. Parfois, pour spécifier la loi $\mathcal{L}$ utilisée dans le modèle, on peut écrire $\mathcal{L}$-BEDD, par exemple Bernoulli-BEDD or Poisson-BEDD.

6.2.2. Cadre asymptotique

Un cadre asymptotique définit ce qui se passe lorsque la taille des données disponbles augmente. Deux cadres peuvent être définis en analyse de réseaux. Dans le premier, les nouvelles données apportent de nouveaux réseaux qui sont considérés comme des expériences supplémentaires d'un modèle unique. Dans l'autre, les nouvelles données font grandir le réseau en ajoutant des nœuds et des arêtes à un réseau unique mais incomplet, supposé être un sous-échantillon du réseau complet, éventuellement infini, tiré du modèle d'intérêt.

Les données d'interaction des écosystèmes sont collectées par les écologues à des endroits différents, à des moments différents et dans des conditions différentes. Par conséquent, contrairement à de nombreux autres types de données de réseaux, les réseaux écologiques sont, par essence, le résultat d'une agrégation de données d'interaction échantillonnées individuellement. L'échantillonnage des interactions écologiques étant coûteux et la plupart du temps incomplet, on souhaite utiliser autant de données que possible pour construire un réseau. Malgré cela, les réseaux obtenus par agrégation sont susceptibles d'être incomplets. L'effort d'échantillonnage nécessaire pour obtenir un réseau complet n'est pas facile à estimer. Pour ces raisons, il est moins courant d'utiliser des mesures répétées de réseaux qui peuvent être considérées comme des réplicats tirés du même modèle. Même si des réplicats sont disponibles, elles sont souvent agrégées pour obtenir un réseau plus complet, avec plus de nœuds et plus d'arêtes. Il est donc raisonnable de travailler avec un cadre asymptotique dans lequel la quantité croissante est le nombre de nœuds d'un réseau unique, plutôt qu'une quantité croissante de réseaux réplicats.

Cette méthode est particulièrement bien adaptée à l'hypothèse d'échangeabilité, puisque chaque réseau peut être représenté par un modèle unique, indépendamment de sa taille. En effet, chaque matrice d'adjacence observée peut-être considérée comme une sous-matrice d'une matrice RCE infinie. Les réseaux peuvent être étudiés dans un espace de modèles. Cela peut être considéré comme une sorte de plongement (*embedding*). Faire un plongement d'un réseau signifie le représenter dans un espace différent, souvent $\mathbb{R}^d$ où $d > 1$. Dans le cadre de cette thèse, chaque réseau est représenté dans l'espace de tous les modèles de réseaux RCE, éventuellement restreint à une certaine famille de modèles. Par conséquent, les réseaux peuvent être étudiés dans l'espace des modèles par l'intermédiaire de leur représentant à l'aide d'outils probabilistes et statistiques. Cette approche est intéressante pour les données écologiques car, pour comparer deux écosystèmes, on aime généralement comparer les deux réseaux associés à ces écosystèmes. Une analyse *model-based* comparerait les deux modèles ajustés à ces réseaux, c'est-à-dire les représentants des deux réseaux dans l'espace des modèles. Si l'espace est restreint à une certaine famille de modèles paramétriques, il est possible de le faire en comparant leurs paramètres. Étant donné que les graphons caractérisent les modèles de graphes échangeables, ceux-là peuvent

également être utilisés pour construire une métrique de comparaison.

Dans cette thèse, on considère que les dimensions $m_N \times n_N$ de la matrice d'adjacence observée croissent, avec $m_N \to \infty$ et $n_N \to \infty$. Dans la plupart des résultats, on suppose que les nombres de lignes et de colonnes restent du même ordre de grandeur, avec $N = m_N + n_N$ et $m_N/N \to \rho \in]0,1[$, mais il est en fait facile de généraliser les résultats à d'autres comportements de m_N et n_N.

L'un des avantages de travailler avec des modèles échangeables et ce cadre asymptotique est la facilité avec laquelle il est possible d'étudier des réseaux de tailles différentes. Si l'estimation statistique des paramètres est une manière d'étudier les modèles, de nombreux autres outils peuvent être utilisés. Comme les modèles sont caractérisés par leurs lois de probabilité, de nombreuses approches utilisant des outils de la statistique paramétrique et non paramétrique, de la théorie des probabilités et même de la théorie de l'information peuvent être combinées pour analyser les réseaux. L'approche développée dans cette thèse consiste à identifier des quantités d'intérêt pouvant être estimées avec les U-statistiques.

6.2.3. U-statistiques sur des réseaux bipartites

Les U-statistiques sont définies comme la moyenne d'une fonction, appelée noyau, d'un sous-échantillon des données. Quand les données se présentent comme une matrice d'adjacence bipartite de taille $m \times n$, le noyau $h : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ est une fonction d'une sous-matrice de taille $p \times q$, avec $1 \leq p \leq m$ et $1 \leq q \leq n$. On peut supposer que le noyau vérifie la propriété de symétrie suivante : pour tout $(\sigma_1, \sigma_2) \in \mathbb{S}_p \times \mathbb{S}_q$,

$$h(Y_{(i_{\sigma_1(1)},\dots,i_{\sigma_1(p)};j_{\sigma_2(1)},\dots,j_{\sigma_2(q)})}) = h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}),$$

où $Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}$ est la sous-matrice de Y de taille $p \times q$ composée des lignes et des colonnes de Y indexées par $i_1, \dots, i_p$ et $j_1, \dots, j_q$ respectivement.

Avec cette hypothèse de symétrie, l'ordre des éléments dans $\mathbf{i} = \{i_1, \dots, i_p\}$ et $\mathbf{j} = \{j_1, \dots, j_q\}$ n'importe pas et on peut utiliser la notation $h_{\mathbf{i},\mathbf{j}} := h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)})$. Alors, la U-statistique associé est

$$U_{m,n} = \binom{m}{p}^{-1} \binom{n}{q}^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_p([\![m]\!]) \\ \mathbf{j} \in \mathcal{P}_q([\![n]\!])}} h_{\mathbf{i},\mathbf{j}}. \tag{6.1}$$

Cette hypothèse sur la symétrie de h peut être faite sans perte de généralité. En effet, si $h^0 : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ n'est pas une fonction symétrique, alors $h : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ définie par

$$h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)}) = (p!q!)^{-1} \sum_{(\sigma_1,\sigma_2) \in \mathbb{S}_p \times \mathbb{S}_q} h^0(Y_{(i_{\sigma_1(1)},\dots,i_{\sigma_1(p)};j_{\sigma_2(1)},\dots,j_{\sigma_2(q)})}) \tag{6.2}$$

possède la propriété de symétrie voulue tout en ayant la même moyenne que h^0.

Les propriétés de $U_{m,n}$ viennent de la structure de dépendance des éléments moyennés. Soit $X_{(i_1,\dots,i_p;j_1,\dots,j_q)} \coloneqq h(Y_{(i_1,\dots,i_p;j_1,\dots,j_q)})$ de telle manière que $U_{m,n}(Y)$ soit la moyenne des premiers éléments du tableau multidimensionnel infini X. Dans le cadre de cette thèse, Y est une matrice RCE. Par conséquent, X est un tableau π-échangeable telle que pour toutes permutations σ_1 et σ_2 de $\mathbb{S}_\infty$, on ait

$$X \overset{\mathcal{D}}{=} \left(X_{(\sigma_1(i_1),\dots,\sigma_1(i_p);\sigma_2(j_1),\dots,\sigma_2(j_q))}\right)_{\substack{1\le i_1\neq\dots\neq i_p\le\infty \\ 1\le j_1\neq\dots\neq j_q\le\infty}}.$$

Cela signifie aussi que

$$\mathbb{E}[U_{m,n}] = \mathbb{E}[X_{(1,\dots,p;1,\dots,q)}] = \mathbb{E}[h_{[\![p]\!],[\![q]\!]}].$$

Par conséquent, $U_{m,n}$ est un estimateur non-biaisé de $X_{(1,\dots,p;1,\dots,q)} = h_{[\![p]\!],[\![q]\!]}$. La méthodologie développée fait usage des U-statistiques pour réaliser des tâches d'inférence statistique sur les modèles de réseaux RCE. Ainsi, elle ne nécessite pas de modèle paramétrique. Elle s'applique à toutes les quantités pouvant être estimées par une fonction d'un sous-réseau.

6.2.4. Méthodologie proposée

Soit θ une quantité d'intérêt à estimer dans un réseau observé. Généralement, θ est une fonction des paramètres d'un modèle de réseau aléatoire. L'objectif est de trouver un estimateur $\widehat{\theta}_N$ ainsi qu'un intervalle de confiance pour θ. L'idée est d'utiliser des estimateurs qui sont soit des U-statistiques, soit des fonctions d'U-statistiques, c'est-à-dire de la forme $\widehat{\theta}_N = U_N \coloneqq U_{m_N,n_N}$ ou $\widehat{\theta}_N \coloneqq g(U_N^{h_1},\dots,U_N^{h_D})$, où g est une fonction dérivable, $(h_1,\dots,h_D)$ un vecteur de noyaux, potentiellement de tailles différentes, et $(U_N^{h_1},\dots,U_N^{h_D})$ les U-statistiques associées à ces noyaux.

Si le modèle de réseau est RCE, on espère pouvoir utiliser des résultats théoriques pour identifier la loi limite de $\widehat{\theta}_N$ quand $N \to \infty$. Ces résultats théoriques sont souvent de la forme

$$\sqrt{\frac{\gamma(N)}{v}}\left(\widehat{\theta}_N - \theta\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{Q},$$

où $\gamma(N)$ est une certaine fonction connue de N telle que $\gamma(N) \to \infty$, v est la variance asymptotique de $\sqrt{\gamma(N)}\widehat{\theta}_N$ et $\mathcal{Q}$ est une loi de probabilité de variance 1. $\gamma(N)$ et v sont reliés à la variance de $\widehat{\theta}_N$ par la relation $\mathbb{V}[\theta_N] = v/\gamma(N) + o(\gamma(N)^{-1})$. Cependant, v doit être en général estimé à partir des données. Si $\widehat{v}_N$ est un estimateur consistant de v, alors le théorème de Slutsky implique

$$\sqrt{\frac{\gamma(N)}{\widehat{v}_N}}\left(\widehat{\theta}_N - \theta\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{Q}. \tag{6.3}$$

Pour $x \in]0,1[$, on note q_x le quantile d'ordre x de la loi $\mathcal{Q}$, c'est-à-dire si W est une variable aléatoire de loi $\mathcal{Q}$, alors $\mathbb{P}(W \le q_x) = x$. Des intervalles de confiance asymptotiques pour θ

peuvent être construits à partir de (6.3). Par exemple,

$$I_N(\alpha) = \left[\widehat{\theta}_N - q_{1-\alpha/2}\sqrt{\frac{\widehat{v}_N}{\gamma(N)}}, \widehat{\theta}_N - q_{\alpha/2}\sqrt{\frac{\widehat{v}_N}{\gamma(N)}}\right]$$

est un intervalle de confiance au niveau α pour θ, c'est-à-dire $\mathbb{P}(\theta \in I_N(\alpha)) \xrightarrow[N\to\infty]{} 1-\alpha$.

Avec de tels intervalles de confiance, on est capable d'obtenir des garanties statistiques pour $\widehat{\theta}_N$. Par conséquent, on est capable de réaliser la plupart des tâches d'inférence statistique sur les modèles de réseaux, dont l'estimation de paramètres, le test d'hypothèses et la comparaison de réseaux.

6.3. Résultats

La liste des ingrédients requis pour appliquer cette méthode à des réseaux observés est composée de :

1. un modèle de réseau RCE, à partir duquel la quantité d'intérêt peut être identifiée,
2. un estimateur de cette quantité d'intérêt, pouvant être écrit comme une fonction d'U-statistiques,
3. un résultat de convergence en loi pour cette fonction d'U-statistiques,
4. un estimateur de la variance asymptotique de cette fonction d'U-statistiques,
5. une manière de calculer efficacement les estimateurs de la quantité d'intérêt et de la variance asymptotique.

La partie théorique de cette thèse s'intéresse principalement à établir des théorèmes limites pour les fonctions d'U-statistiques et des estimateurs de leur variance asymptotique. À travers différents exemples illustrant ces résultats théoriques, je vais montrer comment choisir un modèle de réseau et des estimateurs pertinents, selon le problème étudié. Pour le dernier point, le coût de calcul de ces estimateurs dépend habituellement du problème. Malgré un coût élevé dans le cas général, dans plusieurs exemples, je donne quelques techniques pour remédier à ce problème. Le code source, disponible sur un dépôt en ligne, implémente quelques uns de ces exemples en utilisant des méthodes de calcul efficaces.

6.3.1. Caractérisation des modèles BEDD

Le premier résultat de cette thèse est légèrement différent du reste. Il ne concerne pas directement les U-statistiques, mais c'est l'une des motivations principales derrière leur usage dans ma méthodologie. Ce résultat établit la caractérisation du modèle BEDD par la loi jointe d'un *quadruplet*, c'est-à-dire une sous-matrice de taille 2×2.

Dans le modèle BEDD, les fonctions f et g sont identifiables à leurs classes d'équivalences près, générées par toutes les transformations sur $[0,1]$ conservant la mesure. En outre, il est possible de poser des contraintes sur les fonctions f et g dans la définition du BEDD de telle manière à ce qu'il n'y ait qu'une seule fonction éligible au sein de leurs classes d'équivalences respectives, par exemple, on peut se référer à Yang et al. (2014) pour un argument similaire sur les graphons.

Le théorème suivant stipule que f et g sont charactérisées de manière unique, à leur classes d'équivalences près, par des certaines familles $(F_k)_{k\geq 1}$ et $(G_k)_{k\geq 1}$.

Théorème 2.3.2 (Chp. 2, Le Minh, 2023)**.** *Soit $\Theta = (\lambda, f, g)$ des paramètres du BEDD et $Y \sim BEDD(\Theta)$. La loi de Y est uniquement déterminée par λ, $(F_k)_{k\geq 1}$ et $(G_k)_{k\geq 1}$, où $F_k := \int f^k$ et $G_k := \int g^k$ pour tout $k \geq 1$.*

De plus, pour prouver la caractérisation du modèle BEDD par un quadruplet, il est suffisant de montrer que deux ensembles distincts de paramètres (non-équivalents) pour le BEDD mènent à deux lois nécessairement différentes pour un quadruplet. Avec le théorème précédent, on n'a besoin que de le prouver pour des λ, $(F_k)_{k\geq 1}$ et $(G_k)_{k\geq 1}$ distincts. On a prouvé ce résultat pour une classe de modèles BEDD pour lesquels la loi $\mathcal{L}$ vérifie une hypothèse.

Théorème 2.3.4 (Chp. 2, Le Minh, 2023)**.** *Supposons que pour la famille de lois $\mathcal{L}(\mu)$, il existe une suite de fonctions $(\Psi_k)_{k\geq 1}$ telle que si une variable aléatoire $X \sim \mathcal{L}(\mu)$, alors pour tout $k \geq 1$,*

$$\mathbb{E}[\Psi_k(X)] = \mu^k.$$

Alors, dans ce cas, pour tout $k \in \mathbb{N}$, F_k et G_k sont uniquement déterminés par la loi jointe d'un quadruplet.

Ce résultat a été la motivation derrière mon usage des U-statistiques : puisqu'un quadruplet est suffisant pour contenir toute l'information du modèle, alors il y a un espoir que toute quantité d'intérêt puisse être estimée par une U-statistique moyennant une fonction de quadruplet.

6.3.2. Décompositions de type Hoeffding

Décomposition sur les variables AHK La décomposition de Hoeffding est un outil pratique pour étudier les U-statistiques de variables i.i.d. Cependant, généraliser la décomposition aux U-statistiques sur une matrice RCE ne peut se faire directement. Dans le cas i.i.d., la décomposition de Hoeffding consiste à projeter le noyau h sur des sous-espaces fonctionnels orthogonaux générés par des fonctions de sous-ensembles des observations $(X_1, X_2, ...)$. Pour une matrice RCE Y, une décomposition fondée sur les sous-espaces générés par les fonctions d'éléments de Y pourrait ne

pas vérifier les conditions d'orthogonalité, parce que ces derniers ne sont pas i.i.d. L'idée clé pour trouver une décomposition de Hoeffding dans ce cas est d'utiliser la représentation d'Aldous-Hoover-Kallenberg (AHK) des matrices RCE. Cette représentation permet d'écrire Y, et ainsi la U-statistique, comme une fonction des variables aléatoires i.i.d. $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ et $(\zeta_{ij})_{i\geq 1, j\geq 1}$. Soient $(\xi_i)_{i\geq 1}$, $(\eta_j)_{j\geq 1}$ et $(\zeta_{ij})_{i\geq 1,j\geq 1}$ des familles de variables AHK associées à Y, c'est-à-dire qu'il existe une fonction ϕ telle que pour tout $i \geq 1$, $j \geq 1$, on ait

$$Y_{ij} \overset{a.s.}{=} \phi(\xi_i, \eta_j, \zeta_{ij}). \tag{6.4}$$

Au lieu de projeter sur les sous-espaces générés par les fonctions d'observations, on peut projeter sur les sous-espaces générés par les fonctions de variables AHK. Étant donné que les variables AHK sont i.i.d., l'orthogonalité des sous-espaces peut être obtenue en choisissant des ensembles judicieux de variables AHK sur lesquelles projeter.

Ensuite, la décomposition de la U-statistique se déduit directement de la décomposition du noyau. Dans cette thèse, j'ai finalement trouvé deux systèmes de projection distincts pouvant être qualifiés de décomposition de Hoeffding pour les U-statistiques de matrices RCE.

Première décomposition La première décomposition a été inspirée par le fait que la décomposition de Hoeffding dans le cas i.i.d. est une décomposition sur des sous-espaces générés par des observations, c'est-à-dire sur des ensembles de la forme $(X_{\mathbf{i}})$. Puisqu'on ne peut pas directement utiliser les éléments de Y sans enfreindre l'orthogonalité, on utilise les ensembles de variables AHK $A_{\mathbf{i},\mathbf{j}}$ définis comme suit. Pour tout $\mathbf{i} \in \mathcal{P}(\llbracket m \rrbracket)$ et $\mathbf{j} \in \mathcal{P}(\llbracket n \rrbracket)$, la variable aléatoire $h(Y_{\mathbf{i},\mathbf{j}})$ est mesurable par l'ensemble de variables AHK

$$A_{\mathbf{i},\mathbf{j}} := ((\xi_i)_{i\in\mathbf{i}}, (\eta_j)_{j\in\mathbf{j}}, (\zeta_{ij})_{\substack{i\in\mathbf{i}\\ j\in\mathbf{j}}}).$$

Les projections sont finalement définies de la manière suivante. Soient $\mathbf{i}' \in \mathcal{P}_{r'}(\mathbb{N})$ et $\mathbf{j}' \in \mathcal{P}_{c'}(\mathbb{N})$.

$$p^{\mathbf{i}',\mathbf{j}'}(h_{\mathbf{i},\mathbf{j}}) = \mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid A_{\mathbf{i}',\mathbf{j}'}] - \sum_{(0,0)\leq(r',c')<(r,c)} \sum_{\substack{\mathbf{i}''\subset\mathbf{i}'\\ \mathbf{i}''\subset\mathbf{i}'}} p^{\mathbf{i}'',\mathbf{j}''}(h_{\mathbf{i},\mathbf{j}}).$$

Alors, la décomposition est donnée par

$$h_{\mathbf{i},\mathbf{j}} = \sum_{\substack{\mathbf{i}'\subseteq\mathbf{i}\\ \mathbf{i}'\subseteq\mathbf{i}}} p^{\mathbf{i}',\mathbf{j}'}(h_{\mathbf{i},\mathbf{j}}).$$

Seconde décomposition La seconde décomposition est apparue en adaptant la théorie des U-statistiques généralisées de Janson and Nowicki (1991) aux matrices RCE. Soit G =

$(V_1(G), V_2(G), E(G))$ un graphe bipartite. On peut définir l'ensemble $H(G)$ des variables AHK associées à G comme

$$H(G) = ((\xi_i)_{i\in V_1(G)}, (\eta_j)_{j\in V_2(G)}, (\zeta_{ij})_{(i,j)\in E(G)})$$

et $\mathcal{H}(G) = \sigma(H(G))$, la tribu engendrée par les variables de $H(G)$. Les projections sont alors définies de manière similaire que dans le système précédent

$$p^G(h_{\mathbf{i},\mathbf{j}}) = \mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid H(G)] - \sum_{F\subset G} p^F(h_{\mathbf{i},\mathbf{j}}),$$

et la décomposition de h s'ensuit

$$h_{\mathbf{i},\mathbf{j}} = \sum_{G\subseteq K_{\mathbf{i},\mathbf{j}}} p^G(h_{\mathbf{i},\mathbf{j}}).$$

Lien entre les deux décompositions Les deux décompositions sont bien des décompositions sur des sous-espaces orthogonaux, mais elles ont des applications différentes. La première décomposition est une décomposition de Hoeffding minimale, dans le sens où la U-statistique est décomposée en le nombre de termes le plus petit possible, tout en capturant tous les ordres de grandeurs de la variance de $U_{m,n}$. En effet, on montrera que

$$\mathbb{V}[U_{m,n}] = \sum_{(0,0)<(r,c)\leq(p,q)} \binom{p}{r}^2 \binom{q}{c}^2 \binom{m}{r}^{-1} \binom{n}{c}^{-1} \mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})], \tag{6.5}$$

donc pour un certain (r, c) tel que $(0, 0) < (r, c) \leq (p, q)$, la contribution de $\mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})]$ dans $\mathbb{V}[U_{m,n}]$ est $O(m^{-r}n^{-c})$.

La seconde décomposition utilise des projections dans plus de sous-espaces que dans la première décomposition. En effet, les ensembles utilisés dans la première décomposition peuvent aussi s'écrire avec le formalisme des graphes bipartites de la seconde, avec $A_{\mathbf{i},\mathbf{j}} = H(K_{\mathbf{i},\mathbf{j}})$, où $K_{\mathbf{i},\mathbf{j}} = (V_1(K_{\mathbf{i},\mathbf{j}}), V_2(K_{\mathbf{i},\mathbf{j}}), E(K_{\mathbf{i},\mathbf{j}})) := (\mathbf{i}, \mathbf{j}, \mathbf{i} \times \mathbf{j})$ est le graphe bipartite complètement connecté avec les nœuds en ligne $\mathbf{i}$ et les nœuds en colonne $\mathbf{j}$. En revanche, la seconde décomposition considère aussi les espaces générés par $H(G)$ pour tout $G \subseteq K_{\mathbf{i},\mathbf{j}}$. Certes, elle est bien plus complexe que la première décomposition, mais elle peut identifier toutes les lois limites des U-statistiques dégénérées, ce qui n'est pas possible avec la première décomposition.

6.3.3. Théorèmes limites

Les résultats théoriques clefs de mon travail portent sur l'identification de la loi limite des U-statistiques $U_N := U_{m_N,n_N}$ sur des matrices RCE de taille $m_N \times n_N$ quand $N \to \infty$, où $N = m_N + n_N$ et $m_N/N \to \rho \in]0, 1[$. Tous ces résultats peuvent être étendus au cas multivarié, c'est-à-dire pour obtenir la convergence jointe d'un vecteur d'U-statistiques via le théorème de

Cramér-Wold. De plus, si ces U-statistiques ont le même taux de convergence, par exemple s'ils sont tous non-dégénérés ou dégénérés du même ordre, alors ces résultats peuvent être aussi étendus aux fonctions de ces U-statistiques, en utilisant la *delta méthode*.

Noyaux de quadruplets Le premier résultat de convergence s'applique aux U-statistiques associées à des noyaux fonctions de quadruplets, c'est-à-dire de sous-matrices de taille 2×2.

Théorème 2.2.5 (Chp. 2, Le Minh, 2023)**.** *Soit Y une matrice RCE. Soit h un noyau, fonction de quadruplet telle que $\mathbb{E}[h^2_{\{1,2\},\{1,2\}}] < \infty$. Soient $\mathcal{F}_N = \sigma\big((U^h_{k,l}, k \geq m_N, l \geq n_N)\big)$ et $\mathcal{F}_\infty := \bigcap_{N=1}^{\infty} \mathcal{F}_N$. On note $U_\infty = \mathbb{E}[h_{\{1,2\};\{1,2\}} \mid \mathcal{F}_\infty]$ et*

$$V = \frac{4}{\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{1,3\},\{3,4\}} \mid \mathcal{F}_\infty\big) + \frac{4}{1-\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{3,4\},\{1,3\}} \mid \mathcal{F}_\infty\big).$$

Si $\mathbb{P}(V > 0) > 0$, alors

$$\sqrt{N}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} W,$$

où W est une variable aléatoire de fonction caractéristique $\phi(t) = \mathbb{E}[\exp(-\frac{1}{2}t^2 V)]$.

Théorème 2.2.7 (Chp. 2, Le Minh, 2023)**.** *En plus des hypothèses du Theorème 2.2.5, si Y est dissociée, alors U^h_∞ et V^h sont constants et*

$$\sqrt{N}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V),$$

Plus précisément,

1. $U_\infty = \mathbb{E}[h_{\{1,2\},\{1,2\}}]$,
2. $V = \frac{4}{\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{1,3\},\{3,4\}}\big) + \frac{4}{1-\rho}\mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{3,4\},\{1,3\}}\big)$.

Le premier théorème s'applique à toutes les matrices RCE. Il a été prouvé à l'aide du théorème de convergence des martingales inverses de Eagleson and Weber (1978) (Thm. 1.5.8). L'hypothèse $\mathbb{P}(V > 0) > 0$ assure que la U-statistique n'est pas dégénérée. Ce résultat montre que dans le cas général, où Y n'est pas forcément dissociée, la loi d'une U-statistique converge vers une loi de mélange de variables gaussiennes. Cette loi étant complexe, ce résultat n'est pas facilement exploitable. Le second théorème est une conséquence du premier dans le cas où Y est dissociée, c'est-à-dire que toutes ses sous-matrices ne partageant ni de ligne, ni de colonne en commun sont indépendantes. Il a été déduit grâce à un argument de type Hewitt-Savage. Pour les applications statistiques, c'est bien ce dernier théorème qui est exploité.

Noyaux non-dégénérés de taille $p\times q$ Même si les théorèmes précédents ont été prouvés pour des noyaux de taille 2×2, les preuves peuvent être en fait généralisées aux noyaux de taille $p \times q$ quelconque, au prix de notations et d'une combinatoire plus complexes. Les décompositions de

Hoeffding donnent des démonstrations plus simples pour ces résultats, même pour des noyaux de taille $p \times q$. Le théorème suivant est la généralisation du théorème précédent dans le cas dissocié, démontré avec une décomposition de Hoeffding.

Théorème 3.3.1 (Chp. 3)**.** *Soit Y une matrice dissociée. Soit h un noyau de taille $p \times q$ tel que $\mathbb{E}[h^2_{[\![p]\!],[\![q]\!]}] < \infty$. Soient $U_\infty = \mathbb{E}[h_{[\![p]\!],[\![q]\!]}]$ et*

$$V = \frac{p^2}{\rho}\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})] + \frac{q^2}{1-\rho}\mathbb{V}[p^{\varnothing,\{1\}}(h_{[\![p]\!],[\![q]\!]})].$$

Si $V > 0$, alors

$$\sqrt{N}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V).$$

Remarquons que même si elle a des expressions différentes en apparence dans les deux théorèmes, la variance asymptotique V représente bien la même quantité. Si $p = q = 2$, alors

$$\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![2]\!],[\![2]\!]})] = \mathbb{V}[\mathbb{E}[h_{\{1,2\},\{1,2\}} \mid \xi_1]] = \mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{1,3\},\{3,4\}}\big) \tag{6.6}$$

et

$$\mathbb{V}[p^{\varnothing,\{1\}}(h_{[\![2]\!],[\![2]\!]})] = \mathbb{V}[\mathbb{E}[h_{\{1,2\},\{1,2\}} \mid \eta_1]] = \mathrm{Cov}\big(h_{\{1,2\},\{1,2\}}, h_{\{3,4\},\{1,3\}}\big). \tag{6.7}$$

Noyaux dégénérés de taille $p \times q$ Finalement, dans le cas dégénéré, le problème est beaucoup plus complexe, même si Y est dissociée. Afin de comprendre la dégénérescence, on peut utiliser l'expression développée de la variance de $\mathbb{V}[U_N]$ donnée par l'équation (6.5). Le cas dégénéré se produit lorsque V, qui aurait été le terme dominant de $\mathbb{V}[U_N]$, vaut 0. Dans ce cas, le terme dominant correspond aux termes $\mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})]$ non-nuls avec la plus petite valeur de $d = r + c$, ayant une contribution en $O(N^{-d})$. La bonne normalisation pour le théorème limite n'est alors plus $\sqrt{N}$, mais plutôt $N^{d/2}$.

Quant à la loi limite, elle n'est pas forcément gaussienne. Elle peut être identifiée avec la deuxième décomposition de Hoeffding. Encore une fois, la variance peut être décomposée selon ce système de projection comme dans (6.5). Soit $\Gamma_{r,c}$ l'ensemble des graphes bipartites G avec $V_1(G) = [\![r]\!]$ et $V_2(G) = [\![c]\!]$ tels que tout graphe bipartite avec r nœuds de ligne et c nœuds de colonnes est isomorphe à un et un seul élément de $\Gamma_{r,c}$. Alors

$$\mathbb{V}[U_{m,n}] = \sum_{(0,0)<(r,c)\le(p,q)} \binom{p}{r}^2 \binom{q}{c}^2 \binom{m}{r}^{-1} \binom{n}{c}^{-1} r!c! \sum_{G\in\Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1} \mathbb{V}[p^G].$$

La partie dominante est composée de termes de variance positifs correspondants aux projections caractérisées par les graphes $G \in \bigcup_{(r,c):r+c=d} \Gamma_{r,c}$, pour un certain d. d est appelé le degré principal de h et ces graphes G sont appelés les graphes de support principaux. La loi limite de U_N dépend de la forme de ses graphes de support principaux. Le théorème suivant identifie le cas gaussien.

Théorème 4.2.8 (Chp. 4)**.** *Si tous les graphes de support principaux de U_N sont connectés, alors*

$$N^{d/2}(U_N - U_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

où

$$\sigma^2 = \sum_{\substack{(0,0)<(r,c)\leq(p,q)\\ r+c=d}} \frac{p!^2 q!^2}{(p-r)!^2(q-c)!^2} \sum_{G\in\Gamma_{r,c}} |\mathrm{Aut}(G)|^{-1}\mathbb{V}[p^G].$$

La démonstration dans les cas non-gaussiens n'est pas donnée dans cette thèse. Cependant, on peut supposer que, comme dans Janson and Nowicki (1991), le type de loi limite dépend du plus grand nombre, noté b, de composantes connectées trouvé dans les graphes de support principaux. Dans Janson and Nowicki (1991), si $b = 1$, alors la loi limite est gaussienne, si $b = 2$, alors c'est une somme de variables suivant des lois du χ^2. Pour un b quelconque, ce serait une fonction polynomiale de variables gaussiennes indépendantes de degré b. Même si ce résultat reste non-démontré, plusieurs exemples corroborant cette conjecture seront donnés.

6.3.4. Estimateurs de la variance

Dans cette thèse, j'ai utilisé deux approches pour estimer la variance asymptotique des U-statistiques. Remarquons que, tout comme les théorèmes limites pour les U-statistiques peuvent être généralisés aux fonctions d'U-statistiques par le biais de la delta méthode, les estimateurs de la variance asymptotique peuvent l'être aussi.

Premier estimateur de la variance Les variances asymptotiques données par les théorèmes limites dépendent de la loi de Y, c'est-à-dire du modèle de réseau, et du noyau h. Une expression analytique peut être développée pour la variance asymptotique. Alors, on espère que toutes les quantités qui apparaissent dans cette expression analytique peuvent être estimées pour construire un estimateur consistant de la variance asymptotique. Cette approche marche autant pour les U-statistiques dégénérées que non-dégénérées, mais elle a deux défauts. D'abord, le calcul analytique de la variance peut être laborieux, surtout dans les cas dégénérés. Ensuite, il n'y a pas de technique générale pour estimer les quantités qui apparaissent dans l'expression analytique. Dans mes exemples, toutes ces quantités peuvent être estimées à l'aide d'autres U-statistiques. Cependant, cela est peut-être dû à la chance.

Deuxième estimateur de la variance J'ai défini un estimateur générique pour V du Theorème 3.3.1. Remarquons que le premier terme apparaissant dans l'expression de V est

$\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})] = \mathbb{V}[\mathbb{E}[h_{[\![p]\!],[\![q]\!]} \mid \xi_1]]$. Comme les variables $(\xi_i)_{i\geq 1}$ sont i.i.d., les espérances conditionnelles $\mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid \xi_i]$ sont aussi i.i.d. pour tout $i \geq 1$ tant que $i \in \mathbf{i}$. On note

$$\mu^{(i)} = \mathbb{E}[h_{\mathbf{i},\mathbf{j}} \mid \xi_i]$$

pour tout $(\mathbf{i},\mathbf{j}) \in \mathcal{P}_p(\mathbb{N}) \times \mathcal{P}_q(\mathbb{N})$ tel que $i \in \mathbf{i}$. Par échangeabilité, cette quantité ne dépend pas des éléments de $\mathbf{i}$ (autres que i) et $\mathbf{j}$. Ainsi, l'estimateur sans biais de la variance de $\mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})]$ utilisant m_N lignes de Y est

$$\frac{1}{m_N(m_N-1)} \sum_{1\leq i_1<i_2\leq m_N} (\mu^{(i_1)} - \mu^{(i_2)})^2.$$

Cependant, les $(\mu^{(i)})_{i\geq 1}$ ne sont pas connus, donc on doit également les estimer. Cela peut être fait avec les estimateurs suivants

$$\widehat{\mu}_N^{(i)} := \binom{m_N-1}{p-1}^{-1}\binom{n_N}{q}^{-1} \sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{P}_p(\mathbb{N})\times\mathcal{P}_q(\mathbb{N}) \\ i\in\mathbf{i}}} h_{\mathbf{i},\mathbf{j}}. \tag{6.8}$$

En remplaçant les espérances conditionnelles par leurs estimateurs dans l'expression de l'estimateur sans biais de la variance, on obtient

$$\widehat{v}_N^{1,0} := \binom{m_N}{2}^{-1} \sum_{1\leq i_1<i_2\leq m_N} \frac{(\widehat{\mu}_N^{(i_1)} - \widehat{\mu}_N^{(i_2)})^2}{2}.$$

Par symétrie, on définit pour tout $j \geq 1$,

$$\widehat{\nu}_N^{(j)} := \binom{m_N}{p}^{-1}\binom{n_N-1}{q-1}^{-1} \sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{P}_p(\mathbb{N})\times\mathcal{P}_q(\mathbb{N}) \\ j\in\mathbf{j}}} h_{\mathbf{i},\mathbf{j}}. \tag{6.9}$$

et

$$\widehat{v}_N^{0,1} := \binom{n_N}{2}^{-1} \sum_{1\leq j_1<j_2\leq n_N} \frac{(\widehat{\nu}_N^{(j_1)} - \widehat{\nu}_N^{(j_2)})^2}{2}.$$

Alors, le théorème suivant nous permet d'utiliser $\widehat{v}_N^{1,0}$ et $\widehat{v}_N^{0,1}$ pour construire un estimateur consistant de V.

Théorème 3.4.4 (Chp. 3)**.** *On a* $\widehat{v}_N^{1,0} \xrightarrow[N\to\infty]{\mathbb{P}} \mathbb{V}[p^{\{1\},\varnothing}(h_{[\![p]\!],[\![q]\!]})]$ *et* $\widehat{v}_N^{0,1} \xrightarrow[N\to\infty]{\mathbb{P}} \mathbb{V}[p^{\varnothing,\{1\}}(h_{[\![p]\!],[\![q]\!]})]$. *Par conséquent,*

$$\widehat{V}_N := \frac{p^2}{\rho}\widehat{v}_N^{1,0} + \frac{q^2}{1-\rho}\widehat{v}_N^{0,1} \xrightarrow[N\to\infty]{\mathbb{P}} V.$$

$\widehat{V}_N$ nous permet d'estimer V de manière consistante et sans faire de calculs analytiques. Pour cet aspect, il présente quelques similarités avec les méthodes de rééchantillonnage telles que les méthodes de bootstrap ou de jackknife. Dans mes études de simulation, $\widehat{V}_N$ est au moins aussi précis que la variance estimée par la première méthode. Ainsi, cet estimateur fonctionne

bien pour les cas non-dégénérés. Pour les cas dégénérés, il est en réalité possible de construire de manière analogue des estimateurs consistants $\mathbb{V}[p^{[\![r]\!],[\![c]\!]}(h_{[\![p]\!],[\![q]\!]})]$ pour tout $(0,0) \leq (r,c) \leq (p,q)$, mais quand r et c deviennent proches de p et de q respectivement, ils sont moins précis. Cela est dû au fait que le nombre de termes dans l'estimateur du type (6.8) ou (6.9) est la moyenne de $O(m_N^{p-r} n_N^{q-c})$ termes.

6.4. Plan de la thèse

Le chapitre 1 est un chapitre introductif situant avec plus de détails la problématique écologique et les motivations des travaux de cette thèse. Les notions théoriques de base nécessaires pour la construction de ma méthodologie sont également données. Pour ces raisons, ce chapitre présente une revue de littérature en écologie, en analyse de réseaux et en mathématiques. Enfin, les contributions de cette thèse y sont brièvement présentées, comme dans ce résumé.

Le chapitre 2 porte principalement sur les théorèmes limites pour les U-statistiques des matrices RCE non-dissociées et dissociées, avec des noyaux de quadruplets. Ces résultats sont obtenus en utilisant des arguments de martingales inverses. Dans le cas non-dissocié, la U-statistique converge vers un mélange de gaussiennes, qui devient une simple gaussienne dans le cas dissocié. La caractérisation des modèles BEDD est également étudiée dans ce chapitre. Les exemples incluent l'estimation de l'hétérogénéité des degrés des nœuds, la comparaison de réseaux et le comptage des motifs dans les réseaux. Ce chapitre correspond à un article publié dans le journal ESAIM : Probability & Statistics.

Le chapitre 3 définit une première décomposition de Hoeffding des U-statistiques sur les matrices RCE dissociées. Cette décomposition est utilisée pour prouver le théorème limite des U-statistiques, avec des noyaux de taille quelconque. Un estimateur consistant de leur variance asymptotique est également construit. Les exemples incluent le comptage de motifs dans les réseaux, l'estimation de la distance entre des graphons et l'estimation de l'hétérogénéité des degrés des nœuds. Une analyse de données de réseaux politiques illustre la méthode. Ce chapitre correspond à un article soumis dans une revue de statistique théorique.

Le chapitre 4 traite des U-statistiques dégénérées sur des matrices RCE dissociées. Une deuxième décomposition de Hoeffding est présentée. En utilisant cette nouvelle décomposition, un théorème limite est prouvé pour les U-statistiques dégénérées lorsque la limite est gaussienne. Dans le cas général (limite non-gaussienne), une conjecture est formulée et commentée. Des exemples de statistiques dégénérées sont donnés, incluant le test de l'hétérogénéité des degrés des nœuds d'un réseau. Ce chapitre correspond à un travail en cours, destiné à être soumis plus tard.

Le chapitre 5 propose des idées pour des travaux futurs, fondées sur les recherches présentées dans cette thèse. Certaines de ces idées consistent à compléter la méthodologie, en y ajoutant les éléments qui manquent. D'autres idées visent à améliorer la méthodologie, en étudiant l'erreur d'approximation lors de l'utilisation de résultats asymptotiques. Enfin, les dernières idées concernent l'extension du cadre d'application actuel utilisant des matrices RCE à des modèles de réseaux mieux adaptés aux réseaux réels, en particulier des modèles de réseaux *sparse*.

Bibliography

Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., and Weiss, B. (1996). Strong laws for l-and u-statistics. *Transactions of the American Mathematical Society*, 348(7):2845–2866.

Adamczak, R., Chafaï, D., and Wolff, P. (2016). Circular law for random matrices with exchangeable entries. *Random Structures & Algorithms*, 48(3):454–479.

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(Sep):1981–2014.

Aizen, M. A., Morales, C. L., and Morales, J. M. (2008). Invasive mutualists erode native pollination webs. *PLoS Biology*, 6(2):e31.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598.

Ali, W., Rito, T., Reinert, G., Sun, F., and Deane, C. M. (2014). Alignment-free protein interaction network comparison. *Bioinformatics*, 30(17):i430–i437.

Allee, W. C., Park, O., Emerson, A. E., Park, T., Schmidt, K. P., et al. (1949). *Principles of animal ecology.* Number Edn 1. WB Saundere Co. Ltd.

Almeida-Neto, M., Guimaraes, P., Guimaraes Jr, P. R., Loyola, R. D., and Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239.

Ambrosiano, J., Sims, B., Bartlow, A. W., Rosenberger, W., Ressler, M., and Fair, J. M. (2020). Ontology-based graphs of research communities: a tool for understanding threat reduction networks. *Frontiers in Research Metrics and Analytics*, 5:3.

Arcones, M. A. and Gine, E. (1992). On the bootstrap of u and v statistics. *The Annals of Statistics*, pages 655–674.

Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood estimation: some examples. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 233–243.

Arvesen, J. N. (1969). Jackknifing u-statistics. *The Annals of Mathematical Statistics*, 40(6):2076–2100.

Asta, D. and Shalizi, C. R. (2014). Geometric network comparison. *arXiv preprint arXiv:1411.1350*.

Asta, D. M. and Shalizi, C. R. (2015). Geometric network comparisons. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 102–110. AUAI Press.

Astegiano, J., Guimarães Jr, P. R., Cheptou, P.-O., Vidal, M. M., Mandai, C. Y., Ashworth, L., and Massol, F. (2015). Persistence of plants and pollinators in the face of habitat loss: insights from trait-based metacommunity models. In *Advances in ecological research*, volume 53, pages 201–257. Elsevier.

Austern, M. and Orbanz, P. (2022). Limit theorems for distributions invariant under groups of transformations. *The Annals of Statistics*, 50(4):1960–1991.

Baker, N. J., Kaartinen, R., Roslin, T., and Stouffer, D. B. (2015). Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography*, 38(2):130–139.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113.

Barbour, A. D. and Eagleson, G. (1985). Multiple comparisons and sums of dissociated random variables. *Advances in applied probability*, 17(1):147–162.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752.

Barthélemy, M. (2011). Spatial networks. *Physics reports*, 499(1-3):1–101.

Bartomeus, I. (2013). Understanding linkage rules in plant-pollinator networks by using hierarchical models that incorporate pollinator detectability and plant traits. *PloS one*, 8(7):e69200.

Bascompte, J. and Jordano, P. (2013). *Mutualistic networks*, volume 70. Princeton University Press.

Bascompte, J., Jordano, P., Melián, C. J., and Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387.

Bascompte, J. and Melián, C. J. (2005). Simple trophic modules for complex food webs. *Ecology*, 86(11):2868–2873.

Baskerville, E. B., Dobson, A. P., Bedford, T., Allesina, S., Anderson, T. M., and Pascual, M. (2011). Spatial guilds in the Serengeti food web revealed by a bayesian group model. *PLoS Computational Biology*, 7(12):e1002321.

Bassetti, F., Lagomarsino, M. C., and Mandrà, S. (2007). Exchangeable random networks. *Internet Mathematics*, 4(4):357–400.

Bearman, P. S., Moody, J., and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American journal of sociology*, 110(1):44–91.

Bell, G. (2003). The interpretation of biological surveys. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1533):2531–2542.

Bender, E. A. and Canfield, E. R. (1978). The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58.

Bhamidi, S., Bresler, G., Sly, A., et al. (2011). Mixing time of exponential random graphs. *The Annals of Applied Probability*, 21(6):2146–2170.

Bhattacharyya, S. and Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *The Annals of Statistics*, 43(6):2384–2411.

Bickel, P. J. (1974). Edgeworth expansions in nonparametric statistics. *The Annals of Statistics*, pages 1–20.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073.

Bickel, P. J., Chen, A., and Levina, E. (2011). The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The annals of statistics*, 9(6):1196–1217.

Biella, P., Ollerton, J., Barcella, M., and Assini, S. (2017). Network analysis of phenological units to detect important species in plant-pollinator assemblages: can it inform conservation strategies? *Community Ecology*, 18(1):1–10.

Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence*, 22(7):719–725.

Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons, 3rd edition.

Blanchard, J. L. (2015). Climate change: A rewired food web. *Nature*, 527(7577):173–174.

Blüthgen, N., E. Stork, N., and Fiedler, K. (2004). Bottom-up control and co-occurrence in complex communities: honeydew and nectar determine a rainforest ant mosaic. *Oikos*, 106(2):344–358.

Blüthgen, N., Fründ, J., Vázquez, D. P., and Menzel, F. (2008). What do interaction network metrics tell us about specialization and biological traits. *Ecology*, 89(12):3387–3399.

Blüthgen, N., Menzel, F., and Blüthgen, N. (2006). Measuring specialization in species interaction networks. *BMC ecology*, 6(1):1–12.

Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316.

Bollobás, B., Janson, S., and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122.

Borgs, C., Chayes, J. T., Cohn, H., and Holden, N. (2016). Sparse exchangeable graphs and their limits via graphon processes. *arXiv preprint arXiv:1601.07134*.

Bosch, J., Martín González, A. M., Rodrigo, A., and Navarro, D. (2009). Plant–pollinator networks: adding the pollinator's perspective. *Ecology letters*, 12(5):409–419.

Brault, V. (2014). *Estimation et sélection de modèle pour le modèle des blocs latents*. PhD thesis, Paris-Sud University, France.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Annales de l'IHP Probabilités et statistiques*, 19(3):281–296.

Brown, B. and Kildea, D. (1978). Reduced u-statistics and the hodges-lehmann estimator. *The Annals of Statistics*, pages 828–835.

Brualdi, R. A. and Sanderson, J. G. (1999). Nested species subsets, gaps, and discrepancy. *Oecologia*, 119:256–264.

Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: a review. *Journal of the American Statistical Association*, 88(421):364–373.

Burkle, L. A., Marlin, J. C., and Knight, T. M. (2013). Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science*, 339(6127):1611–1615.

Burnham, K. P. and Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, 65(3):625–633.

Busiello, D. M., Suweis, S., Hidalgo, J., and Maritan, A. (2017). Explorability and the origin of network sparsity in living systems. *Scientific reports*, 7(1):12323.

Cai, D., Campbell, T., and Broderick, T. (2016). Edge-exchangeable graphs and sparsity. *Advances in Neural Information Processing Systems*, 29.

Callaert, H. and Janssen, P. (1978). The berry-esseen theorem for u-statistics. *The Annals of Statistics*, pages 417–421.

Callaert, H. and Veraverbeke, N. (1981). The order of the normal approximation for a studentized u-statistic. *The Annals of Statistics*, pages 194–200.

Carlo, T. A., Collazo, J. A., and Groom, M. J. (2003). Avian fruit preferences across a puerto rican forested landscape: pattern consistency and implications for seed removal. *Oecologia*, 134:119–131.

Caron, F. (2012). Bayesian nonparametric models for bipartite graphs. *Advances in Neural Information Processing Systems*, 25.

Caron, F. and Fox, E. B. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 79(5):1295.

Caron, F., Panero, F., and Rousseau, J. (2017). On sparsity, power-law and clustering properties of graphex processes. *arXiv preprint arXiv:1708.03120*.

Carstensen, D. W., Sabatino, M., and Morellato, L. P. C. (2016). Modularity, pollination systems, and interaction turnover in plant-pollinator networks across space. *Ecology*, 97(5):1298–1306.

Chan, Y.-K. and Wierman, J. (1977). On the berry-esseen theorem for u-statistics. *The Annals of Probability*, pages 136–139.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of statistics*, pages 265–270.

Chatterjee, S. (2008). A new method of normal approximation. *The Annals of Probability*, 36(4):1584–1610.

Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461.

Chatterjee, S. and Shao, Q.-M. (2011). Non-normal approximation by stein's method of exchangeable pairs with application to the curie-weiss model. *The Annals of Applied Probability*, 21(2):464–483.

Chiang, H. D., Kato, K., and Sasaki, Y. (2021). Inference for high-dimensional exchangeable arrays. *Journal of the American Statistical Association*, pages 1–11.

Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882.

Clauset, A., Moore, C., and Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101.

Clements, F. E. and Long, F. L. (1923). *Experimental pollination: an outline of the ecology of flowers and insects*, volume 336 of *Carnegie Institution of Washington publication*. Carnegie institution of Washington.

Cohen, R. and Havlin, S. (2003). Scale-free networks are ultrasmall. *Physical review letters*, 90(5):058701.

Côme, E. and Latouche, P. (2015). Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. *Statistical Modelling*, 15(6):564–589.

Connor, E. F. and Simberloff, D. (1979). The assembly of species communities: chance or competition? *Ecology*, 60(6):1132–1140.

Coulson, M., Gaunt, R. E., and Reinert, G. (2016). Poisson approximation of subgraph counts in stochastic block models and a graphon model. *ESAIM: Probability and Statistics*, 20:131–142.

Crane, H. and Dempsey, W. (2018). Edge exchangeable models for interaction networks. *Journal of the American Statistical Association*, 113(523):1311–1326.

Dall'Asta, L., Alvarez-Hamelin, I., Barrat, A., Vázquez, A., and Vespignani, A. (2006). Exploring networks with traceroute-like probes: Theory and simulations. *Theoretical Computer Science*, 355(1):6–24.

Dalsgaard, B., Magård, E., Fjeldså, J., Martín González, A. M., Rahbek, C., Olesen, J. M., Ollerton, J., Alarcón, R., Cardoso Araujo, A., Cotton, P. A., et al. (2011). Specialization in plant-hummingbird networks is associated with species richness, contemporary precipitation and quaternary climate-change velocity. *PLoS One*, 6(10):e25891.

Daudin, J.-J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Statistics and computing*, 18(2):173–183.

Davezies, L., D'Haultfœuille, X., and Guyonvarch, Y. (2021). Empirical process results for exchangeable arrays. *The Annals of Statistics*, 49(2):845–862.

David, P., Thebault, E., Anneville, O., Duyck, P.-F., Chapuis, E., and Loeuille, N. (2017). Impacts of invasive species on food webs: a review of empirical data. *Advances in ecological research*, 56:1–60.

de La Fortelle, A. (2020). Generalized inverses of increasing functions and lebesgue decomposition. *Markov Processes And Related Fields*.

de Manincor, N., Hautekèete, N., Mazoyer, C., Moreau, P., Piquot, Y., Schatz, B., Schmitt, E., Zélazny, M., and Massol, F. (2020a). How biased is our perception of plant-pollinator networks? a comparison of visit-and pollen-based representations of the same networks. *Acta Oecologica*, 105:103551.

de Manincor, N., Hautekèete, N., Piquot, Y., Schatz, B., Vanappelghem, C., and Massol, F. (2020b). Does phenology explain plant–pollinator interactions at different latitudes? an assessment of its explanatory power in plant–hoverfly networks in french calcareous grasslands. *Oikos*, 129(5):753–765.

de Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, 27(5):292–306.

Dehling, H., Denker, M., and Philipp, W. (1984). Invariance principles for von mises and u-statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 67:139–167.

Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for u-statistics. *Journal of Multivariate Analysis*, 51(2):392–413.

Denker, M. and Keller, G. (1983). On u-statistics and v. mise'statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64(4):505–522.

Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII*, 28(1):33–61.

Diamond, J. M. (1975). Assembly of species communities. In Cody, M. L. and Diamond, J. M., editors, *Ecology and evolution of communities*, pages 342–444. Harvard University Press, Cambridge, Massachusetts, USA.

Dicks, L., Corbet, S., and Pywell, R. (2002). Compartmentalization in plant-insect flower visitor webs. *Journal of Animal Ecology*, pages 32–43.

Dobson, A. (2009). Food-web structure and ecosystem services: insights from the serengeti. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524):1665–1682.

Doob, J. L. (1953). *Stochastic processes*, volume 7. Wiley New York.

Dormann, C. F., Fründ, J., Blüthgen, N., and Gruber, B. (2009). Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal*, 2(1).

Dormann, C. F., Fründ, J., and Schaefer, H. M. (2017). Identifying causes of patterns in ecological networks: opportunities and limitations. *Annual Review of Ecology, Evolution, and Systematics*, 48:559–584.

Dubart, M., Alonso, P., Barroso-Bergada, D., Becker, N., Bethune, K., Bohan, D. A., Boury, C., Cambon, M., Canard, E., Chancerel, E., et al. (2021). Coupling ecological network analysis with high-throughput sequencing-based surveys: Lessons from the next-generation biomonitoring project. In *Advances in ecological research*, volume 65, pages 367–430. Elsevier.

Duchemin, Q., De Castro, Y., and Lacour, C. (2020). Concentration inequality for u-statistics of order two for uniformly ergodic markov chains. *arXiv preprint arXiv:2011.11435*.

Duchemin, Q., De Castro, Y., and Lacour, C. (2022). Three rates of convergence or separation via u-statistics in a dependent framework. *Journal of Machine Learning Research*, 23:1–59.

Dunne, J. A., Williams, R. J., and Martinez, N. D. (2002a). Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922.

Dunne, J. A., Williams, R. J., and Martinez, N. D. (2002b). Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology letters*, 5(4):558–567.

Dupont, Y. L., Hansen, D. M., and Olesen, J. M. (2003). Structure of a plant–flower-visitor network in the high-altitude sub-alpine desert of tenerife, canary islands. *Ecography*, 26(3):301–310.

Dupont, Y. L. and Olesen, J. M. (2009). Ecological modules and roles of species in heathland plant-insect flower visitor networks. *Journal of Animal Ecology*, pages 346–353.

Dupont, Y. L., Padrón, B., Olesen, J. M., and Petanidou, T. (2009). Spatio-temporal variation in the structure of pollination networks. *Oikos*, 118(8):1261–1269.

Durrett, R. (2019). *Probability: theory and examples*, volume 49. Cambridge university press.

Eagleson, G. (1979). Orthogonal expansions and u-statistics. *Australian Journal of Statistics*, 21(3):221–237.

Eagleson, G. K. and Weber, N. C. (1978). Limit theorems for weakly exchangeable arrays. *Mathematical Proceedings of the Cambridge Philosophical Society*, 84(1):123–130.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti, M., Guimaraes, P. R., et al. (2013). The dimensionality of ecological networks. *Ecology letters*, 16(5):577–583.

Emmert-Streib, F., Dehmer, M., and Shi, Y. (2016). Fifty years of graph matching, network alignment and network comparison. *Information sciences*, 346:180–197.

Erdős, P. and Rényi, A. (1959). On random graphs I. *Publications Mathematicae Debrecen*, 6(290-297):18.

Faust, K. and Skvoretz, J. (2002). Comparing networks across space and time, size and species. *Sociological Methodology*, 32(1):267–299.

Feller, W. (1971). *An Introduction to Probability theory and its application Vol II.* John Wiley and Sons.

Fonseca, C. R. and Ganade, G. (1996). Asymmetries, compartments and null interactions in an amazonian ant-plant community. *Journal of Animal Ecology*, pages 339–347.

Fontaine, C. (2009). Stabilité et évolution des réseaux d'interactions écologiques.

Fontaine, C., Guimarães Jr, P. R., Kéfi, S., Loeuille, N., Memmott, J., van Der Putten, W. H., Van Veen, F. J., and Thébault, E. (2011). The ecological and evolutionary implications of merging different types of networks. *Ecology letters*, 14(11):1170–1181.

Fortanier, F., Liberatore, A., Maurer, A., Pilgrim, G., and Thomson, L. (2017). The oecd-wto balanced trade in services database. *World Trade Organization and Organisation for Economic Co-operation and Development.*

Fortuna, M. A., Stouffer, D. B., Olesen, J. M., Jordano, P., Mouillot, D., Krasnov, B. R., Poulin, R., and Bascompte, J. (2010). Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of animal ecology*, pages 811–817.

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5):75–174.

Fortunato, S. and Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the national academy of sciences*, 104(1):36–41.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., et al. (2018). Science of science. *Science*, 359(6379):eaao0185.

Frank, O. (2005). Network sampling and model fitting. *Models and methods in social network analysis*, pages 31–56.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Friedrich, K. O. (1989). A berry-esseen bound for functions of independent random variables. *The Annals of Statistics*, pages 170–183.

Fründ, J., McCann, K. S., and Williams, N. M. (2016). Sampling bias is a challenge for quantifying specialization and network structure: lessons from a quantitative niche model. *Oikos*, 125(4):502–513.

Funke, T. and Becker, T. (2019). Stochastic block models: A comparison of variants and inference methods. *PloS one*, 14(4):e0215296.

Gao, C. and Lafferty, J. (2017). Testing network structure using relations between small subgraph probabilities. *arXiv preprint arXiv:1704.06742*.

Geslin, B., Gauzens, B., Thébault, E., and Dajoz, I. (2013). Plant pollinator networks along a gradient of urbanisation. *PloS one*, 8(5):e63421.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683.

Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144.

Gill, R. J., Baldock, K. C., Brown, M. J., Cresswell, J. E., Dicks, L. V., Fountain, M. T., Garratt, M. P., Gough, L. A., Heard, M. S., Holland, J. M., et al. (2016). Protecting an ecosystem service: approaches to understanding and mitigating threats to wild insect pollinators. In *Advances in ecological research*, volume 54, pages 135–206. Elsevier.

Gilpin, M. E. and Diamond, J. M. (1982). Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia*, 52(1):75–84.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). *A survey of statistical network models*. Now Publishers Inc.

Gotelli, N. J. and Entsminger, G. L. (2001). Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia*, 129(2):281–291.

Gotelli, N. J. and Entsminger, G. L. (2003). Swap algorithms in null model analysis. *Ecology*, 84(2):532–535.

Govaert, G. and Nadif, M. (2003). Clustering with block mixture models. *Pattern Recognition*, 36(2):463–473.

Grams, W. F. and Serfling, R. (1973). Convergence rates for u-statistics and related statistics. *The Annals of Statistics*, pages 153–160.

Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B., Tylianakis, J. M., Wood, S. A., et al. (2019). Bringing elton and grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks. *Ecography*, 42(3):401–415.

Green, A. and Shalizi, C. R. (2022). Bootstrapping exchangeable random graphs. *Electronic Journal of Statistics*, 16(1):1058–1095.

Gregory, G. G. (1977). Large sample theory for u-statistics and tests of fit. *The annals of statistics*, pages 110–123.

Grisi-Filho, J. H., Ossada, R., Ferreira, F., and Amaku, M. (2013). Scale-free networks with the same degree distribution: Different structural properties. *Physics Research International*, 2013.

Guimera, R. and Nunes Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *nature*, 433(7028):895–900.

Hadfield, J. D., Krasnov, B. R., Poulin, R., and Nakagawa, S. (2014). A tale of two phylogenies: comparative analyses of ecological interactions. *The American Naturalist*, 183(2):174–187.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic press.

Halmos, P. R. (1946). The theory of unbiased estimation. *The Annals of Mathematical Statistics*, 17(1):34–43.

Han, Q. and Kato, K. (2022). Berry–esseen bounds for chernoff-type nonstandard asymptotics in isotonic regression. *The Annals of Applied Probability*, 32(2):1459–1498.

Handcock, M. S. (2003). Assessing degeneracy in statistical models of social networks. Technical report, Center for Statistics and the Social Sciences, University of Washington.

Handcock, M. S. and Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1):5.

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354.

Heinen, J. H., Florens, F. B. V., Baider, C., Kissling, W. D., Whittaker, R. J., Rahbek, C., and Borregaard, M. (2023). Novel plant–frugivore network on mauritius is unlikely to compensate for the extinction of seed dispersers. *Nature Communications*, 14(1):1019.

Heleno, R. H., Olesen, J. M., Nogales, M., Vargas, P., and Traveset, A. (2013). Seed dispersal networks in the galápagos and the consequences of alien plant invasions. *Proceedings of the Royal Society B: Biological Sciences*, 280(1750):20122112.

Heltshe, J. F. and Forrester, N. E. (1983). Estimating species richness using the jackknife procedure. *Biometrics*, pages 1–11.

Herlau, T., Schmidt, M. N., and Mørup, M. (2016). Completely random measures for modelling block-structured sparse networks. *Advances in Neural Information Processing Systems*, 29.

Herrera, J. P., Moody, J., and Nunn, C. L. (2023). Predicting primate-parasite associations using exponentional random graph models. *Journal of Animal Ecology*.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325.

Hoeffding, W. (1961). The strong law of large numbers for u-statistics. Technical report, North Carolina State University. Dept. of Statistics.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*, 2:275.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258.

Jacobs, A. Z. and Clauset, A. (2014). A unified view of generative models for networks: models, methods, opportunities, and challenges. *arXiv preprint arXiv:1411.4070*.

Janson, S. and Nowicki, K. (1991). The asymptotic distributions of generalized u-statistics with applications to random graphs. *Probability theory and related fields*, 90:341–375.

Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *The American Naturalist*, 129(5):657–677.

Jordano, P. (2016). Sampling networks of ecological interactions. *Functional ecology*, 30(12):1883–1893.

Jordano, P., Bascompte, J., and Olesen, J. M. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology letters*, 6(1):69–81.

Kaiser-Bunbury, C. N., Mougal, J., Whittington, A. E., Valentin, T., Gabriel, R., Olesen, J. M., and Blüthgen, N. (2017). Ecosystem restoration strengthens pollination network resilience and function. *Nature*, 542(7640):223–227.

Kallenberg, O. (1999). Multivariate sampling and the estimation problem for exchangeable arrays. *Journal of Theoretical Probability*, 12(3):859–883.

Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles*, volume 9. Springer.

Kashtan, N., Itzkovitz, S., Milo, R., and Alon, U. (2004). Topological generalizations of network motifs. *Physical Review E*, 70(3):031909.

Klopp, O., Tsybakov, A. B., and Verzelen, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Annals of Statistics*, 45(1):316–354.

Konstantopoulos, T. and Yuan, L. (2019). On the extendibility of finitely exchangeable probability measures. *Transactions of the American Mathematical Society*, 371(10):7067–7092.

Korolyuk, V. S. and Borovskich, Y. V. (1993). *Theory of U-statistics*. Springer Dordrecht.

Koutrouli, M., Karatzas, E., Paez-Espino, D., and Pavlopoulos, G. A. (2020). A guide to conquer the biological network era using graph theory. *Frontiers in bioengineering and biotechnology*, 8:34.

Lafortune, G., Fuller, G., Moreno, J., Schmidt-Traub, G., and Kroll, C. (2018). Sdg index and dashboards detailed methodological paper. *Sustainable Development Solutions Network*, pages 1–56.

Lanuza, J. B., Allen-Perkins, A., and Bartomeus, I. (2023). The non-random assembly of network motifs in plant–pollinator networks. *Journal of Animal Ecology*, 92(3):760–773.

Latouche, P., Birmele, E., and Ambroise, C. (2012). Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1):93–115.

Lauritzen, S., Rinaldo, A., and Sadeghi, K. (2018). Random networks, graphical models and exchangeability. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 80(3):481–508.

Le Minh, T. (2023). U-statistics on bipartite exchangeable networks. *ESAIM: Probability and Statistics*, 27:576–620.

Le Minh, T., Donnet, S., Massol, F., and Robin, S. (2023). Hoeffding-type decomposition for u-statistics on bipartite networks. *arXiv preprint arXiv:2308.14518*.

Lee, A. J. (1990). *U-statistics: Theory and Practice*. Routledge.

Levin, K. and Levina, E. (2019). Bootstrapping networks with latent space structure. *arXiv preprint arXiv:1907.10821*.

Lindenstrauss, E. (1999). Pointwise theorems for amenable groups. *Electronic Research Announcements of the American Mathematical Society*, 5(12):82–90.

Lovász, L. (2012). *Large networks and graph limits*, volume 60. American Mathematical Soc.

Lovász, L. and Szegedy, B. (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957.

Lovász, L. and Szegedy, B. (2010). Limits of compact decorated graphs. *arXiv preprint arXiv:1010.5155*.

Lubold, S., Chandrasekhar, A. G., and McCormick, T. H. (2023). Identifying the latent space geometry of network models through analysis of curvature. *ournal of the Royal Statistical Society Series B: Statistical Methodology*. qkad002.

Luczkovich, J. J., Borgatti, S. P., Johnson, J. C., and Everett, M. G. (2003). Defining and measuring trophic role similarity in food webs using regular equivalence. *Journal of Theoretical Biology*, 220(3):303–321.

Lunde, R. and Sarkar, P. (2023). Subsampling sparse graphons under minimal assumptions. *Biometrika*, 110(1):15–32.

Maesono, Y. (1991). On the normal approximation of u-statistics of degree two. *Journal of statistical planning and inference*, 27(1):37–50.

Mai, J.-F. (2020). The infinite extendibility problem for exchangeable real-valued random vectors. *Probability Surveys*, 17:677–753.

Major, P. (1981). Wick polynomials. In *Multiple Wiener-Itô Integrals: With Applications to Limit Theorems*, pages 6–13. Springer Berlin Heidelberg, Berlin, Heidelberg.

Martin Gonzalez, A. M., Allesina, S., Rodrigo, A., and Bosch, J. (2012). Drivers of compartmentalization in a mediterranean pollination network. *Oikos*, 121(12):2001–2013.

Maruyama, P. K., Vizentin-Bugoni, J., Oliveira, G. M., Oliveira, P. E., and Dalsgaard, B. (2014). Morphological and spatio-temporal mismatches shape a neotropical savanna plant-hummingbird network. *Biotropica*, 46(6):740–747.

Massol, F., Dubart, M., Calcagno, V., Cazelles, K., Jacquet, C., Kefi, S., and Gravel, D. (2017). Island biogeography of food webs. In *Advances in ecological research*, volume 56, pages 183–262. Elsevier.

Matias, C. and Robin, S. (2014). Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM: Proceedings and Surveys*, 47:55–74.

Maugis, P.-A., Olhede, S., Priebe, C., and Wolfe, P. (2020). Testing for equivalence of network distribution using subgraph counts. *Journal of Computational and Graphical Statistics*, 29(3):455–465.

McCormick, T. H. and Zheng, T. (2015). Latent surface models for networks using aggregated relational data. *Journal of the American Statistical Association*, 110(512):1684–1695.

McCullagh, P. (2000). Resampling and exchangeable arrays. *Bernoulli*, pages 285–301.

Memmott, J., Waser, N. M., and Price, M. V. (2004). Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1557):2605–2611.

Michalska-Smith, M. J. and Allesina, S. (2019). Telling ecological networks apart by their structure: A computational challenge. *PLoS Computational Biology*, 15(6):e1007076. Data available at `github.com/mjsmith037/Classifying_Bipartite_Networks`.

Miele, V., Ramos-Jiliberto, R., and Vázquez, D. P. (2020). Core–periphery dynamics in a plant–pollinator network. *Journal of Animal Ecology*, 89(7):1670–1677.

Miklós, I. and Podani, J. (2004). Randomization of presence-absence matrices: comments and new algorithms. *Ecology*, 85(1):86–92.

Milgram, S. (1967). The small world problem. *Psychology Today*, 2(1):60–67.

Montoya, J. M., Pimm, S. L., and Solé, R. V. (2006). Ecological networks and their fragility. *Nature*, 442(7100):259–264.

Nandi, H. and Sen, P. (1963). On the properties of U-statistics when the observations are not independent: Part two unbiased estimation of the parameters of a finite population. *Calcutta Statistical Association Bulletin*, 12(4):124–148.

Naulet, Z., Roy, D. M., Sharma, E., and Veitch, V. (2021). Bootstrap estimators for the tail-index and for the count statistics of graphex processes. *Electronic Journal of Statistics*, 15:282–325.

Neff, F., Brändle, M., Ambarlı, D., Ammer, C., Bauhus, J., Boch, S., Hölzel, N., Klaus, V. H., Kleinebecker, T., Prati, D., et al. (2021). Changes in plant-herbivore network structure and robustness along land-use intensity gradients in grasslands and forests. *Science advances*, 7(20):eabf3985.

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2):404–409.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Newman, M. E., Watts, D. J., and Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2566–2572.

Newth, D. (2006). The structure of social networks. *Complex Science for a Complex World*, page 71.

Nielsen, A. and Bascompte, J. (2007). Ecological networks, nestedness and sampling effort. *Journal of Ecology*, pages 1134–1141.

Novak, M., Wootton, J. T., Doak, D. F., Emmerson, M., Estes, J. A., and Tinker, M. T. (2011). Predicting community responses to perturbations in the face of imperfect knowledge and network complexity. *Ecology*, 92(4):836–846.

Novella-Fernandez, R., Rodrigo, A., Arnan, X., and Bosch, J. (2019). Interaction strength in plant-pollinator networks: Are we using the right measure? *PloS one*, 14(12):e0225930.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087.

Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'connor, L., and Thuiller, W. (2019). Diversity indices for ecological networks: a unifying framework using hill numbers. *Ecology letters*, 22(4):737–747.

Olesen, J. M., Bascompte, J., Dupont, Y. L., Elberling, H., Rasmussen, C., and Jordano, P. (2011). Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 278(1706):725–732.

Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. (2006). The smallest of all worlds: pollination networks. *Journal of Theoretical Biology*, 240(2):270–276.

Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Sciences*, 104(50):19891–19896.

Ollerton, J., McCollin, D., Fautin, D. G., and Allen, G. R. (2007). Finding nemo: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):591–598.

Orbanz, P. and Roy, D. M. (2014). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence*, 37(2):437–461.

Orsini, C., Dankulov, M. M., Colomer-de Simón, P., Jamakovic, A., Mahadevan, P., Vahdat, A., Bassler, K. E., Toroczkai, Z., Boguná, M., Caldarelli, G., et al. (2015). Quantifying randomness in real networks. *Nature communications*, 6(1):1–10.

Ouadah, S., Latouche, P., and Robin, S. (2022). Motif-based tests for bipartite networks. *Electronic Journal of Statistics*, 16(1):293–330.

Owen, A. B. (2007). The pigeonhole bootstrap. *The Annals of applied statistics*, 1(2):386–411.

Paine, R. T. (1966). Food web complexity and species diversity. *The American Naturalist*, 100(910):65–75.

Passmore, H., Bruna, E., Heredia, S., and Vasconcelos, H. (2012). Resilient networks of ant-plant mutualists in amazonian forest fragments. *PLoS ONE*, 7(8).

Patterson, B. D. and Atmar, W. (1986). Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological journal of the Linnean society*, 28(1-2):65–82.

Pauw, A. and Stanway, R. (2015). Unrivalled specialization in a pollination network from south africa reveals that specialization increases with latitude only in the southern hemisphere. *Journal of Biogeography*, 42(4):652–661.

Payrató-Borràs, C., Hernández, L., and Moreno, Y. (2019). Breaking the spell of nestedness: The entropic origin of nestedness in mutualistic systems. *Physical Review X*, 9(3):031024.

Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., Maglianesi, M. A., Melián, C. J., Pitteloud, C., Roslin, T., et al. (2018). Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93(2):785–800.

Pérez-Matus, A., Ospina-Alvarez, A., Camus, P. A., Carrasco, S. A., Fernandez, M., Gelcich, S., Godoy, N., Ojeda, F. P., Pardo, L. M., Rozbaczylo, N., et al. (2017). Temperate rocky subtidal reef community reveals human impacts across the entire food web. *Marine Ecology Progress Series*, 567:1–16.

Perna, A. and Latty, T. (2014). Animal transportation networks. *Journal of The Royal Society Interface*, 11(100):20140334.

Picard, F., Daudin, J.-J., Koskas, M., Schbath, S., and Robin, S. (2008). Assessing the exceptionality of network motifs. *Journal of Computational Biology*, 15(1):1–20.

Pocock, M. J., Evans, D. M., and Memmott, J. (2012). The robustness and restoration of a network of ecological networks. *Science*, 335(6071):973–977.

Podani, J. and Schmera, D. (2012). A comparative evaluation of pairwise nestedness measures. *Ecography*, 35(10):889–900.

Poisot, T., Cirtwill, A. R., Cazelles, K., Gravel, D., Fortin, M.-J., and Stouffer, D. B. (2016). The structure of probabilistic networks. *Methods in Ecology and Evolution*, 7(3):303–312.

Pornon, A., Andalo, C., Burrus, M., and Escaravage, N. (2017). Dna metabarcoding data unveils invisible pollination networks. *Scientific Reports*, 7(1):16828.

Pržulj, N., Corneil, D. G., and Jurisica, I. (2004). Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18):3508–3515.

Przytycka, T. M. (2006). An important connection between network motifs and parsimony models. In *Annual International Conference on Research in Computational Molecular Biology*, pages 321–335. Springer.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191.

Ramírez, S. R., Eltz, T., Fujiwara, M. K., Gerlach, G., Goldman-Huertas, B., Tsutsui, N. D., and Pierce, N. E. (2011). Asynchronous diversification in a specialized plant-pollinator mutualism. *Science*, 333(6050):1742–1746.

Ramos-Jiliberto, R., Domínguez, D., Espinoza, C., Lopez, G., Valdovinos, F. S., Bustamante, R. O., and Medel, R. (2010). Topological change of andean plant–pollinator networks along an altitudinal gradient. *Ecological Complexity*, 7(1):86–90.

Rankin, R. W., Mann, J., Singh, L., Patterson, E. M., Krzyszczyk, E., and Bejder, L. (2016). The role of weighted and topological network information to understand animal social networks: a null model approach. *Animal Behaviour*, 113:215–228.

Redhead, J. W., Woodcock, B. A., Pocock, M. J., Pywell, R. F., Vanbergen, A. J., and Oliver, T. H. (2018). Potential landscape-scale pollinator networks across great britain: structure, stability and influence of agricultural land cover. *Ecology Letters*, 21(12):1821–1832.

Reinert, G. and Röllin, A. (2010). Random subgraph counts and u-statistics: multivariate normal approximation via exchangeable pairs and embedding. *Journal of applied probability*, 47(2):378–393.

Reitzner, M. and Schulte, M. (2013). Central limit theorems for u-statistics of poisson point processes. *The Annals of Probability*, 41(6):3879–3909.

Rivera-Hutinel, A., Bustamante, R., Marín, V., and Medel, R. (2012). Effects of sampling completeness on the structure of plant–pollinator networks. *Ecology*, 93(7):1593–1603.

Robertson, C. (1928). *Flowers and insects; lists of visitors of four hundred and fifty-three flowers.* C. Robertson.

Rubin, H. and Vitale, R. (1980). Asymptotic distribution of symmetric statistics. *The Annals of Statistics*, pages 165–170.

Saavedra, S., Gilarranz, L. J., Rohr, R. P., Schnabel, M., Uzzi, B., and Bascompte, J. (2014). Stock fluctuations are correlated and amplified across networks of interlocking directorates. *EPJ Data Science*, 3(1):1–11.

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., Fuller, G., and Woelm, F. (2020). *The Sustainable Development Goals and COVID-19. Sustainable Development Report 2020.* Cambridge University Press, Cambridge.

Schleuning, M., Blüthgen, N., Flörchinger, M., Braun, J., Schaefer, H. M., and Böhning-Gaese, K. (2011). Specialization and interaction strength in a tropical plant–frugivore network differ among forest strata. *Ecology*, 92(1):26–36.

Schmidt, M. N. and Mørup, M. (2013). Nonparametric bayesian modeling of complex networks: An introduction. *IEEE Signal Processing Magazine*, 30(3):110–128.

Schucany, W. R. and Bankson, D. M. (1989). Small sample variance estimators for u-statistics. *Australian Journal of Statistics*, 31(3):417–426.

Sciarra, C., Chiarotti, G., Ridolfi, L., and Laio, F. (2021). A network approach to rank countries chasing sustainable development. *Scientific Reports*, 11(1):1–12.

Scott, D. and Huggins, R. (1985). A law of the iterated logarithm for weakly exchangeable arrays. *Mathematical Proceedings of the Cambridge Philosophical Society*, 98(3):541–545.

Sebastián-González, E., Dalsgaard, B., Sandel, B., and Guimaraes Jr, P. R. (2015). Macroecological trends in nestedness and modularity of seed-dispersal networks: human impact matters. *Global Ecology and Biogeography*, 24(3):293–303.

Sen, P. K. (1960). On some convergence properties of ustatistics. *Calcutta Statistical Association Bulletin*, 10(1-2):1–18.

Sen, P. K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. *The Annals of Statistics*, pages 316–329.

Seo, E. and Hutchinson, R. (2018). Predicting links in plant-pollinator interaction networks using latent factor models with implicit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Shao, Q.-M. and Zhang, Z.-S. (2019). Berry–esseen bounds of normal and nonnormal approximation for unbounded exchangeable pairs. *The Annals of Probability*, 47(1):61–108.

Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, 31(1):64–68.

Silverman, B. W. (1976). Limit theorems for dissociated random variables. *Advances in Applied Probability*, 8(4):806–819.

Simmons, B. I., Cirtwill, A. R., Baker, N. J., Wauchope, H. S., Dicks, L. V., Stouffer, D. B., and Sutherland, W. J. (2019). Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*, 128(2):154–170.

Smith, E. P. and van Belle, G. (1984). Nonparametric estimation of species richness. *Biometrics*, pages 119–129.

Snijders, T. A. (1981). The degree variance: an index of graph heterogeneity. *Social networks*, 3(3):163–174.

Snijders, T. A. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, 14(1):75–100.

Stang, M., Klinkhamer, P. G., and Van Der Meijden, E. (2006). Size constraints and flower abundance determine the number of interactions in a plant–flower visitor web. *Oikos*, 112(1):111–121.

Staniczenko, P., Kopp, J. C., and Allesina, S. (2013). The ghost of nestedness in ecological networks. *Nature communications*, 4(1):1–6.

Stark, D. (2001). Compound poisson approximations of subgraph counts in random graphs. *Random Structures & Algorithms*, 18(1):39–60.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, pages 583–602.

Stouffer, D. B., Camacho, J., Jiang, W., and Nunes Amaral, L. A. (2007). Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):1931–1940.

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212.

Strona, G., Nappo, D., Boccacci, F., Fattorini, S., and San-Miguel-Ayanz, J. (2014). A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications*, 5.

Stumpf, M. P. and Porter, M. A. (2012). Critical truths about power laws. *Science*, 335(6069):665–666.

Sueur, C., Romano, V., Sosa, S., and Puga-Gonzalez, I. (2019). Mechanisms of network evolution: a focus on socioecological factors, intermediary mechanisms, and selection pressures. *Primates*, 60:167–181.

Sugiura, S. (2007). Structure of a herbivore–parasitoid community: are parasitoids shared by different herbivore guilds? *Basic and Applied Ecology*, 8(6):544–551.

Sundaresan, S. R., Fischhoff, I. R., Dushoff, J., and Rubenstein, D. I. (2007). Network metrics reveal differences in social organization between two fission–fusion species, grevy's zebra and onager. *Oecologia*, 151:140–149.

Tajoli, L., Airoldi, F., and Piccardi, C. (2021). The network of international trade in services. *Applied Network Science*, 6(1):1–25.

Tantardini, M., Ieva, F., Tajoli, L., and Piccardi, C. (2019). Comparing methods for comparing networks. *Scientific reports*, 9(1):1–19.

Thébault, E. (2013). Identifying compartments in presence–absence matrices and bipartite networks: insights into modularity measures. *Journal of Biogeography*, 40(4):759–768.

Thébault, E. and Fontaine, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856.

Thomas, M., Verzelen, N., Barbillon, P., Coomes, O. T., Caillon, S., McKey, D., Elias, M., Garine, E., Raimond, C., Dounias, E., et al. (2015). A network-based method to detect patterns of local crop biodiversity: validation at the species and infra-species levels. In *Advances in Ecological Research*, volume 53, pages 259–320. Elsevier.

Thompson, J. N. (2005). *The geographic mosaic of coevolution*. University of Chicago Press.

Timóteo, S., Correia, M., Rodríguez-Echeverría, S., Freitas, H., and Heleno, R. (2018). Multilayer networks reveal the spatial structure of seed-dispersal interactions across the great rift landscapes. *Nature Communications*, 9(1):140.

Todeschini, A., Miscouridou, X., and Caron, F. (2020). Exchangeable random measures for sparse and modular graphs with overlapping communities. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(2).

Trøjelsgaard, K. and Olesen, J. M. (2013). Macroecology of pollination networks. *Global Ecology and Biogeography*, 22(2):149–162.

Tsiotas, D. (2019). Detecting different topologies immanent in scale-free networks with the same degree distribution. *Proceedings of the National Academy of Sciences*, 116(14):6701–6706.

Tylianakis, J. M. and Morris, R. J. (2017). Ecological networks across environmental gradients. *Annual Review of Ecology, Evolution, and Systematics*, 48.

UN General Assembly (2017). Work of the statistical commission pertaining to the 2030 agenda for sustainable development. Technical report, Resolution A/RES/71/313, United Nation Publications, New York.

Valdovinos, F. S. (2019). Mutualistic networks: moving closer to a predictive theory. *Ecology letters*, 22(9):1517–1534.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge University Press.

Van Kleunen, L. B., Peterson, K. A., Hayden, M. T., Keyes, A., Schwartz, A. J., Li, H., and Dee, L. E. (2023). Decision-making under uncertainty for species introductions into ecological networks. *Ecology Letters*.

van Zwet, W. R. (1984). A berry-esseen bound for symmetric statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 66(3):425–440.

Vázquez, D. P. and Aizen, M. A. (2004). Asymmetric specialization: a pervasive feature of plant–pollinator interactions. *Ecology*, 85(5):1251–1257.

Vázquez, D. P., Melián, C. J., Williams, N. M., Blüthgen, N., Krasnov, B. R., and Poulin, R. (2007). Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116(7):1120–1127.

Veitch, V. and Roy, D. M. (2015). The class of random graphs arising from exchangeable random measures. *arXiv preprint arXiv:1512.03099*.

Veitch, V., Sharma, E., Naulet, Z., and Roy, D. M. (2017). Exchangeable modelling of relational data: checking sparsity, train-test splitting, and sparse exchangeable poisson matrix factorization. *arXiv preprint arXiv:1712.02311*.

Vila, M., Bartomeus, I., Dietzsch, A. C., Petanidou, T., Steffan-Dewenter, I., Stout, J. C., and Tscheulin, T. (2009). Invasive plant integration into native plant–pollinator networks across europe. *Proceedings of the Royal Society B: Biological Sciences*, 276(1674):3887–3893.

Vilhena, D. A. and Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature communications*, 6(1):6848.

Villa-Galaviz, E., Boege, K., and del Val, E. (2012). Resilience in plant-herbivore networks during secondary succession. *PLoS One*, 7(12):e53009.

Vitousek, P. M. (1994). Beyond global warming: ecology and global change. *Ecology*, 75(7):1861–1876.

Vizentin-Bugoni, J., Maruyama, P. K., Debastiani, V. J., Duarte, L. d. S., Dalsgaard, B., and Sazima, M. (2016). Influences of sampling effort on detected patterns and structuring processes of a neotropical plant–hummingbird network. *Journal of Animal Ecology*, 85(1):262–272.

Walther, B. A. and Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, 28(6):815–829.

Wang, Y., Pelekis, C., and Ramon, J. (2015). U-statistics on network-structured data with kernels of degree larger than one. In *Statistically Sound Data Mining*, pages 37–48. PMLR.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of mathematical psychology*, 44(1):92–107.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

Wells, K. and O'Hara, R. B. (2013). Species interactions: estimating per-individual interaction strength and covariates before simplifying data into per-species ecological networks. *Methods in Ecology and Evolution*, 4(1):1–8.

Welti, E., Helzer, C., and Joern, A. (2017). Impacts of plant diversity on arthropod communities and plant–herbivore network architecture. *Ecosphere*, 8(10):e01983.

White, H. C., Boorman, S. A., and Breiger, R. L. (1976). Social structure from multiple networks. i. blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–780.

Wilber, M. Q., Johnson, P. T., and Briggs, C. J. (2017). When can we infer mechanism from parasite aggregation? a constraint-based approach to disease ecology. *Ecology*, 98(3):688–702.

Williamson, S. A. (2016). Nonparametric network models for link prediction. *The Journal of Machine Learning Research*, 17(1):7102–7121.

Wolfe, P. J. and Olhede, S. C. (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*.

Wright, D. H. and Reeves, J. H. (1992). On the meaning and measurement of nestedness of species assemblages. *Oecologia*, 92:416–428.

Yang, J., Han, C., and Airoldi, E. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Artificial Intelligence and Statistics*, pages 1060–1067. PMLR.

Yoshihara, K.-i. (1976). Limiting behavior of u-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35(3):237–252.

Young, S. J. and Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 138–149. Springer.

Zamora-López, G. (2009). *Linking structure and function of complex cortical networks*. PhD thesis, Universität Potsdam.

Zhang, Z.-S. (2022). Berry–esseen bounds for generalized u-statistics. *Electronic Journal of Probability*, 27:1–36.

Zhao, L., Beverlin, B., Netoff, T., and Nykamp, D. Q. (2011). Synchronization from second order network connectivity statistics. *Frontiers in computational neuroscience*, 5:28.

Zhao, L. and Chen, X. (1990). Normal approximation for finite-population u-statistics. *Acta mathematicae applicatae Sinica*, 6(3):263–272.

Zhou, T., Ren, J., Medo, M., and Zhang, Y.-C. (2007). Bipartite network projection and personal recommendation. *Physical review E*, 76(4):046115.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France