HOEFFDING-TYPE DECOMPOSITION FOR U-STATISTICS ON BIPARTITE NETWORKS

Tâm Le Minh[1] & Sophie Donnet[1] & François Massol[2] & Stéphane Robin[3]

[1] *Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, 91120, Palaiseau, France*
[2] *Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France*
[3] *Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, 75005 Paris, France*

Résumé. Les réseaux bipartites sont naturellement représentés par leurs matrices d'adja-cence. On s'intéresse à des réseaux dont les nœuds sont échangeables, c'est-à-dire que leurs matrices d'adjacences sont échangeables ligne-colonne. Nos travaux consistent à utiliser des U-statistiques pour étudier les propriétés topologiques de ces réseaux. En général, les U-statistiques sont les moyennes d'une fonction d'un sous-ensemble sur tous les sous-ensembles d'une population étudiée. Dans le cas des matrices échangeables ligne-colonne, les éléments moyennés possèdent une structure de dépendance complexe. On propose une décomposition à la Hoeffding de ces éléments. On montre comment on peut l'utiliser pour étudier les propriétés des U-statistiques en tant qu'estimateurs. Toujours en utilisant cette décomposition, on construit un estimateur consistant de la variance de ces U-statistiques. Notre objectif est de pouvoir utiliser ces estimateurs dans des problèmes d'inférence statistique sur des données de réseaux bipartites.

Mots-clés. U-statistiques, échangeabilité, réseaux bipartites, décomposition de Hoeffding, estimation de la variance

Abstract. Bipartite networks are naturally represented by their adjacency matrices. We consider node-exchangeable networks, which means their adjacency matrices are row-column exchangeable. We explore the use of U-statistics to investigate properties of these networks. In general, U-statistics are the average of a function of a subsample over all the subsamples of a population. Applied on row-column exchangeable matrices, they become a sum of elements with a complex dependency structure. We derive a Hoeffding-type decomposition of these elements. We show how it allows us to study the properties of U-statistics as estimators. A consistent estimator for the variance of these U-statistics can also be built from this decomposition. Our ultimate goal is to apply these estimators to statistical inference tasks involving bipartite network data.

Keywords. U-statistics, exchangeability, bipartite networks, Hoeffding decomposition, variance estimation

1 Introduction

1.1 Dissociated RCE matrices

Networks are used to represent interactions between the entities of a complex system. The entities are represented by nodes, which are linked by edges when they interact. In bipartite networks, there are two types of nodes and interactions only happen between two nodes of 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). In the adjacency matrix Y of a bipartite network, 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 Y_{ij} can take other values, giving 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 entire network, i.e. the entries of Y. In random graph models, it is common to assume that the nodes of the networks are exchangeable. This means that the distribution of the network does not change if its nodes are permuted. For example, the stochastic blockmodel (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 three node-exchangeable. On the adjacency matrix of a bipartite network, this assumption implies 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.

1.2 U-statistics

U-statistics are a generalization of the empirical mean to functions of more than one variable. 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

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

where $h : \mathbb{R}^k \to \mathbb{R}$ is a symmetric function referred to as the kernel. The case where the $(Y_i)_{i\geq 1}$ are i.i.d. is well-studied, where the U-statistics are known to be asymptotically normal (Hoeffding, 1948) and can be used for inference tasks such as estimation and hypothesis testing.

A network U-statistic averages a function over submatrices of size $p \times q$. Given an infinite adjacency matrix Y from which we observe the first m rows and n columns and given a kernel $h : \mathcal{M}_{p,q}(\mathbb{R}) \to \mathbb{R}$ function of smaller submatrices ($1 \leq p \leq m$, $1 \leq q \leq n$), its expression is

$$U_{m,n}^h(Y) = \binom{m}{p}^{-1} \binom{n}{q}^{-1} \sum_{\substack{1\leq i_1 < ... < i_p \leq m \\ 1\leq j_1 < ... < j_q \leq n}} h(Y_{(i_1,...,i_p;j_1,...,j_q)}),$$

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

1.3 Contribution

$U^h_{m,n}(Y)$ is a sum over elements $X_{\{i_1,...,i_p;j_1,...,j_q\}} := h(Y_{(i_1,...,i_p;j_1,...,j_q)})$ which have a particular dependency structure. The array X verify the following exchangeability property : for any two permutations σ_1 and σ_2, we have

$$(X_{\{i_1,...,i_p;j_1,...,j_q\}})_{\substack{\{i_1,...,i_p\}\subset\mathbb{N}\\ \{j_1,...,j_q\}\subset\mathbb{N}}} \overset{\mathcal{D}}{=} (X_{\{\sigma_1(i_1),...,\sigma_1(i_p);\sigma_2(j_1),...,\sigma_2(j_q)\}})_{\substack{\{i_1,...,i_p\}\subset\mathbb{N}\\ \{j_1,...,j_q\}\subset\mathbb{N}}}. \tag{2}$$

The results on U-statistics studied in the literature, in particular, for U-statistics of jointly exchangeable variables or of separately exchangeable variable (Davezies et al., 2021, Austern and Orbanz, 2022), are not directly applicable to our case. Le Minh (2021) considered U-statistics of RCE matrices, but only for kernels of matrices of size 2×2.

To study the properties of our U-statistics, we find a Hoeffding-type decomposition for $U^h_{m,n}$. We use this decomposition to derive a Central Limit Theorem and a variance estimator.

2 Hoeffding decomposition of subgraph U-statistics

2.1 Aldous-Hoover-Kallenberg (AHK) representation

Corollary 7.23 of Kallenberg (2005) states that for any dissociated RCE matrix Y, there exists ξ_i, η_j and ζ_{ij} arrays of i.i.d. random variables with uniform distribution over $[0,1]$ and a real measurable function f such that for all $1\le i,j<\infty$,

$$Y_{ij} \overset{a.s.}{=} f(\xi_i,\eta_j,\zeta_{ij}). \tag{3}$$

With such a representation, the kernel function taken on a $p\times q$ subgraph can be written

$$h(Y_{(i_1,...,i_p;j_1,...,j_q)}) = h_f((\xi_i)_{i\in\{i_1,...,i_p\}};(\eta_j)_{j\in\{j_1,...,j_q\}};(\zeta_{ij})_{\substack{i\in\{i_1,...,i_p\}\\ j\in\{j_1,...,j_q\}}}),$$

where $h_f = h\circ f$. For any set E, we denote by $\mathcal{P}(E)$ the set of all its subsets and $\mathcal{P}_r(E)$ the sets of all its subsets with cardinal r. For $\mathbf{i}\in\mathcal{P}(\mathbb{N})$ and $\mathbf{j}\in\mathcal{P}(\mathbb{N})$, denote

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

We can then write

$$\mathbb{E}[h(Y_{(i_1,...,i_p;j_1,...,j_q)})\mid\mathcal{A}_{\mathbf{i},\mathbf{j}}] := \mathbb{E}[h_f(Y_{(i_1,...,i_p;j_1,...,j_q)})\mid(\xi_i)_{i\in\mathbf{i}};(\eta_j)_{j\in\mathbf{j}};(\zeta_{ij})_{\substack{i\in\mathbf{i}\\ j\in\mathbf{j}}}].$$

Now, let $0\le r\le p$ and $0\le c\le q$. Let $\mathbf{i}\in\mathcal{P}_r([\![m]\!])$ and $\mathbf{j}\in\mathcal{P}_c([\![n]\!])$, where for any positive integer N, $[\![N]\!] := \{1,...,N\}$. If $\mathbf{i}\subset\{i_1,...,i_p\}$ and $\mathbf{j}\subset\{j_1,...,j_q\}$, the quantities $\mathbb{E}[h(Y_{(i_1,...,i_p;j_1,...,j_q)})\mid\mathcal{A}_{\mathbf{i},\mathbf{j}}]$ only depend on the elements of $\mathbf{i}$ and $\mathbf{j}$ and not on the other elements of $\{i_1,...,i_p\}\setminus\mathbf{i}$ and $\{j_1,...,j_q\}\setminus\mathbf{j}$. For this reason, we can denote

$$\psi^{r,c}_{(\mathbf{i},\mathbf{j})}h(Y) := \mathbb{E}[h(Y_{(i_1,...,i_p;j_1,...,j_q)})\mid\mathcal{A}_{\mathbf{i},\mathbf{j}}].$$

2.2 Hoeffding projections

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 $\mathbf{i} \in \mathcal{P}_r(\llbracket m \rrbracket)$ and $\mathbf{j} \in \mathcal{P}_c(\llbracket n \rrbracket)$, we define by recursion the following quantity

$$p^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y) = \psi^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y) - \sum_{(0,0) \leq (r',c') < (r,c)} \sum_{\substack{\mathbf{i}' \in \mathcal{P}_{r'}(\mathbf{i}) \\ \mathbf{j}' \in \mathcal{P}_{c'}(\mathbf{j})}} p^{r',c'}_{(\mathbf{i}',\mathbf{j}')} h(Y). \tag{4}$$

$p^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y)$ is actually the projection of $h(Y_{(i_1,\ldots,i_p;j_1,\ldots,j_q)})$ on the subspace generated by L_2 functions of all the AHK variables of $\mathcal{A}_{\mathbf{i},\mathbf{j}}$.

Finally, since $\psi^{p,q}_{(i_1,\ldots,i_p;j_1,\ldots,j_q)} h(Y) = h(Y_{(i_1,\ldots,i_p;j_1,\ldots,j_q)})$, (4) yields the decomposition of the kernel function h

$$h(Y_{(i_1,\ldots,i_p;j_1,\ldots,j_q)}) = \sum_{(0,0) \leq (r,c) \leq (p,q)} \sum_{\substack{\mathbf{i} \in \mathcal{P}_r(\{i_1,\ldots,i_p\}) \\ \mathbf{j} \in \mathcal{P}_c(\{j_1,\ldots,j_q\})}} p^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y).$$

This system of projection satisfies properties similar to these of the Hoeffding decomposition for the kernel functions of usual U-statistics on i.i.d. data. In particular, the following orthogonality properties hold.

Proposition 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 $(0,0) \leq (r_1,c_1) < (p_1,q_1)$ and $(0,0) \leq (r_2,c_2) < (p_2,q_2)$ 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*

$$\mathrm{Cov}(p^{r_1,c_1}_{(\mathbf{i}_1,\mathbf{j}_1)} h_1(Y), p^{r_2,c_2}_{(\mathbf{i}_2,\mathbf{j}_2)} h_2(Y)) = 0.$$

2. *Let (r,c) such that $(0,0) \leq (r,c) < (p_1,q_1)$ and $(0,0) \leq (r,c) < (p_2,q_2)$. Let $(\mathbf{i}_1,\mathbf{j}_1)$ and $(\mathbf{i}_2,\mathbf{j}_2)$ two elements of $\mathcal{P}_r(\llbracket m \rrbracket) \times \mathcal{P}_c(\llbracket n \rrbracket)$. If $(\mathbf{i}_1,\mathbf{j}_1) \neq (\mathbf{i}_2,\mathbf{j}_2)$, then*

$$\mathrm{Cov}(p^{r,c}_{(\mathbf{i}_1,\mathbf{j}_1)} h_1(Y), p^{r,c}_{(\mathbf{i}_2,\mathbf{j}_2)} h_2(Y)) = 0.$$

2.3 Decomposition of U-statistics

Having written Hoeffding-type projections of kernel functions, one can decompose the associated U-statistics :

$$U^h_{m,n}(Y) = \sum_{(0,0) \leq (r,c) \leq (p,q)} \binom{p}{r} \binom{q}{c} P^{r,c}_{m,n} h(Y),$$

where for all $0 \leq r \leq p$ and $0 \leq c \leq q$, $P^{r,c}_{m,n} h(Y) := \binom{m}{r}^{-1} \binom{n}{c}^{-1} \sum_{\substack{\mathbf{i} \in \mathcal{P}_r(\llbracket m \rrbracket) \\ \mathbf{j} \in \mathcal{P}_c(\llbracket n \rrbracket)}} p^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y)$ are the U-statistic of kernel functions $Y_{\{\mathbf{i},\mathbf{j}\}} \to p^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y)$ taken on the first $m \times n$ row and column of the matrix Y. Proposition 2.1 ensures that all the $P^{r,c}_{m,n} h(Y)$ are orthogonal and degenerate U-statistics.

3 Central Limit Theorem

For the following sections, we will be investigating the asymptotic properties of $U^h_{m,n}(Y)$, we let the dimensions m and n of the network grow to infinity. Let $(m_N, n_N)_{N\geq 1}$ be a sequence such that $\frac{m_N}{N} \xrightarrow[N\to\infty]{} \rho$ and $\frac{n_N}{N} \xrightarrow[N\to\infty]{} 1-\rho$, where $\rho \in]0,1[$. Denote $U^h_N(Y) := U^h_{m_N,n_N}(Y)$ and $P^{r,c}_N h := P^{r,c}_{m_N,n_N} h$. When this is unambiguous, we will simply write U^h_N, $P^{r,c}_N h$, $p^{r,c}_{(\mathbf{i},\mathbf{j})} h$ and $\psi^{r,c}_{(\mathbf{i},\mathbf{j})} h$ instead of $U^h_N(Y)$, $P^{r,c}_N h(Y)$, $p^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y)$ and $\psi^{r,c}_{(\mathbf{i},\mathbf{j})} h(Y)$.

The Hoeffding decomposition of U^h_N can be used to prove a Central Limit Theorem for U^h_N. Denote $U^h_\infty := P^{0,0}_N h = \mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})]$ and $v^{r,c}_h := \mathbb{V}[\psi^{r,c}_{([\![r]\!],[\![c]\!])} h]$. We have

$$\begin{aligned}
\sqrt{N}(U^h_N - U^h_\infty) &= \sqrt{N}\left(pP^{1,0}_N h + qP^{0,1}_N h + \sum_{\substack{(0,0)<(r,c)\leq(p,q) \\ (r,c)\neq(1,0)\neq(0,1)}} \binom{p}{r}\binom{q}{c} P^{r,c}_N h \right) \\
&= \frac{\sqrt{N}p}{m_N} \sum_{i=1}^{m_N} p^{1,0}_{(\{i\},\emptyset)} h + \frac{\sqrt{N}q}{n_N} \sum_{j=1}^{n_N} p^{0,1}_{(\emptyset,\{j\})} h + o_P(1) \\
&= \frac{p}{\sqrt{\rho m_N}} \sum_{i=1}^{m_N} p^{1,0}_{(\{i\},\emptyset)} h + \frac{q}{\sqrt{(1-\rho)n_N}} \sum_{j=1}^{n_N} p^{0,1}_{(\emptyset,\{j\})} h + o_P(1).
\end{aligned}$$

The classical Central Limit Theorem for i.i.d. variables applies independently to the two main terms of this decomposition, so their sum converges to the sum of two independent Gaussian variables with respective variances $\frac{p^2}{\rho} v^{1,0}_h$ and $\frac{q^2}{1-\rho} v^{0,1}_h$.

Theorem 3.1. *Let Y be a dissociated RCE matrix. Let h be a $p \times q$ kernel function such that $\mathbb{E}[h(Y_{(1,\dots,p;1,\dots,q)})^2] < \infty$. Set*

$$V^h = \frac{p^2}{\rho} v^{1,0}_h + \frac{q^2}{1-\rho} v^{0,1}_h.$$

If $V^h > 0$, then

$$\sqrt{N}(U^h_N - U^h_\infty) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, V^h).$$

This theorem can be extended to functions of U-statistics using the delta method (see for example Chapter 3 of Van der Vaart, 2000).

Corollary 3.2. *Let $h_1, \dots, h_D$ be D linearly independent kernel functions such that Theorem 3.1 applies to each. 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 and ∇g its gradient. If $V^\delta := \nabla g(\theta)^T \Sigma \nabla g(\theta) > 0$, then*

$$\sqrt{N}\left(g(U^{h_1}_N, \dots, U^{h_D}_N) - g(\theta) \right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}\left(0, V^\delta\right),$$

where

$$\Sigma = \left(C^{h_k,h_\ell} \right)_{1\leq k,\ell\leq D},$$

with $C^{h_k,h_\ell} = \lim_{N\to+\infty} N\mathrm{Cov}(U^{h_k}_N, U^{h_\ell}_N)$ for all $1 \leq k,\ell \leq D$ (and $C^{h_k,h_k} = V^{h_k}$).

4 Variance estimator

In the previous section, we have established the asymptotic normality of U-statistics of RCE dissociated matrices. However, to apply our results to statistical inference tasks, the analytical calculation of V^h is often tedious and depends from the kernel function h. Now, we introduce an estimator of the variance of U_N^h which is kernel-agnostic. The idea behind our estimator is to exploit the expression of V^h found in Theorem 3.1, based on the Hoeffding projections. It consists in two variance terms $v_h^{1,0} = \mathbb{V}[\psi^{1,0}_{(\{1\},\emptyset)}h]$ and $v_h^{0,1} = \mathbb{V}[\psi^{0,1}_{(\emptyset,\{1\})}h]$, which are the variances of specific conditional expectations. We derive estimators for these conditional expectations and we plug them in an empirical variance estimator. We show that the resulting estimator is consistent.

4.1 Estimators of the conditional expectations

Let $\widehat{\mu}_N^{(i)}$ be the average of the kernel function applied on the $p \times q$ subgraphs containing the row i. Symmetrically, let $\widehat{\nu}_N^{(j)}$ be the average of the kernel function applied on the $p \times q$ subgraphs containing the column j.

$$\widehat{\mu}_N^{(i)} := \binom{m_N - 1}{p-1}^{-1} \binom{n_N}{q}^{-1} \sum_{\substack{\{i_2,\dots,i_p\} \in [\![m_N]\!] \setminus \{i\} \\ \{j_1,\dots,j_q\} \in [\![n_N]\!]}} h(Y_{(i,i_2,\dots,i_p;j_1,\dots,j_q)}),$$

$$\widehat{\nu}_N^{(j)} := \binom{m_N}{p}^{-1} \binom{n_N - 1}{q-1}^{-1} \sum_{\substack{\{i_1,\dots,i_p\} \in [\![m_N]\!] \\ \{j_2,\dots,j_q\} \in [\![n_N]\!] \setminus \{j\}}} h(Y_{(i_1,\dots,i_p;j,j_2,\dots,j_q)}).$$

Proposition 4.1. *If Y is a RCE matrix, then :*

- $\mathbb{E}[\widehat{\mu}_N^{(i)} \mid \xi_i] = \psi^{1,0}_{(\{i\},\emptyset)}h$ *and* $\mathbb{E}[\widehat{\nu}_N^{(j)} \mid \eta_j] = \psi^{0,1}_{(\emptyset,\{j\})}h$,
- $\widehat{\mu}_N^{(i)} \xrightarrow[N\to\infty]{a.s.,L_1} \psi^{1,0}_{(\{i\},\emptyset)}h$ *and* $\widehat{\nu}_N^{(j)} \xrightarrow[N\to\infty]{a.s.,L_1} \psi^{0,1}_{(\emptyset,\{j\})}h$.

As a consequence, $\widehat{\mu}_N^{(i)}$ and $\widehat{\nu}_N^{(j)}$ are conditionally unbiased and consistent estimators for $\psi^{1,0}_{(\{i\},\emptyset)}h$ and $\psi^{0,1}_{(\emptyset,\{j\})}h$.

In particular, the second property is a strong law of large numbers and comes from the fact that $\widehat{\mu}_N^{(i)}$ and $\widehat{\nu}_N^{(j)}$ are backward martingales with respect to the decreasing filtrations $\mathcal{F}_N^{\mu(i)}(Y) := \sigma((\widehat{\mu}_K^{(i)}(Y))_{K\geq N})$ and $\mathcal{F}_N^{\nu(j)}(Y) := \sigma((\widehat{\nu}_K^{(j)}(Y))_{K\geq N})$.

4.2 Estimator of the variance

Since we have defined estimators for $\psi^{1,0}_{(\{i\},\emptyset)}h$ and $\psi^{0,1}_{(\emptyset,\{j\})}h$, we can give natural plug-in estimators for $v_h^{1,0} = \mathbb{V}[\psi^{1,0}_{(\{1\},\emptyset)}h]$ and $v_h^{0,1} = \mathbb{V}[\psi^{0,1}_{(\emptyset,\{j\})}h]$:

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

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, an estimator for V^h is

$$\widehat{V}_N := \frac{p^2}{\rho} \widehat{v}_N^{1,0} + \frac{q^2}{1-\rho} \widehat{v}_N^{0,1}.$$

The following theorem shows that $\widehat{V}_N$ is a consistent estimator for V^h. It follows from the fact that first, $\mathbb{E}[\widehat{v}_N^{1,0}] = v_h^{1,0} + O\left(N^{-1}\right)$ and $\mathbb{E}[\widehat{v}_N^{0,1}] = v_h^{0,1} + O\left(N^{-1}\right)$ and second, $\mathbb{V}[\widehat{v}_N^{1,0}] = O\left(N^{-1}\right)$ and $\mathbb{V}[\widehat{v}_N^{0,1}] = O\left(N^{-1}\right)$.

Theorem 4.2. *We have $\widehat{v}_N^{1,0} \xrightarrow[N\to\infty]{\mathbb{P}} v_h^{1,0}$ and $\widehat{v}_N^{0,1} \xrightarrow[N\to\infty]{\mathbb{P}} v_h^{0,1}$. As a consequence, $\widehat{V}_N \xrightarrow[N\to\infty]{\mathbb{P}} V^h$.*

With Theorem 4.2, it is possible to use $\widehat{V}_N$ for statistical inference tasks when plugged-in in place of V^h, an asymptotic normality result similar to Theorem 3.1 holds.

Corollary 4.3. *If $V > 0$, then*

$$\sqrt{\frac{N}{\widehat{V}_N}} (U_N^h - U_\infty^h) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

Our approach for building an estimator for V^h in the case of a simple U-statistic can be extended in the case of functions of several U-statistics. One can estimate the elements C^{h_k,h_ℓ} of the covariance matrix Σ of Corollary 3.2 in a similar way. With the estimators for the conditional expectations of h_k $\widehat{\mu}_N^{h_k,(i)}$ and $\widehat{\nu}_N^{h_k,(j)}$, and of h_ℓ, $\widehat{\mu}_N^{h_\ell,(i)}$ $\widehat{\nu}_N^{h_\ell,(j)}$. We define the estimator for the covariance term C^{h_k,h_ℓ} as :

$$\widehat{C}_N^{h_k,h_\ell} := \frac{p^2}{\rho} \widehat{c}_N^{h_k,h_\ell;1,0} + \frac{q^2}{1-\rho} \widehat{c}_N^{h_k,h_\ell;0,1},$$

where

$$\widehat{c}_N^{h_k,h_\ell;1,0} := \binom{m_N}{2}^{-1} \sum_{1 \le i_1 < i_2 \le m_N} \frac{(\widehat{\mu}_N^{h_k,(i_1)} - \widehat{\mu}_N^{h_k,(i_2)})(\widehat{\mu}_N^{h_\ell,(i_1)} - \widehat{\mu}_N^{h_\ell,(i_2)})}{2},$$

and

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

$\widehat{C}_N^{h_k,h_\ell}$ has analogous properties to $\widehat{V}_N^{h_k}$, which is actually a subcase when $h_k = h_\ell$. Therefore, $\widehat{\Sigma}_N := (\widehat{C}_N^{h_k,h_\ell})_{1 \le k,\ell \le D}$ is an estimator for Σ of Corollary 3.2. Define

$$\widehat{V}_N^\delta := \nabla g(U_N^{h_1}, ..., U_N^{h_D})^T \widehat{\Sigma}_N \nabla g(U_N^{h_1}, ..., U_N^{h_D}),$$

an estimator of the asymptotic variance of Corollary 3.2. Since we have $\widehat{\Sigma}_N \xrightarrow[N\to\infty]{} \Sigma$ and g differentiable at θ, we also have $\widehat{V}_N^\delta \xrightarrow[N\to\infty]{} V^\delta = \nabla g(\theta)^T \Sigma \nabla g(\theta)$ and we can formulate the following extension to Corollary 4.3.

Corollary 4.4. *If $V^\delta > 0$, then*

$$\sqrt{\frac{N}{\widehat{V}_N^\delta}} \left(g(U_N^{h_1}, ..., U_N^{h_D}) - g(\theta) \right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0,1).$$

5 Applications

5.1 RCE dissociated random graph models

A RCE graph model is dissociated if and only if it can be written as a W-graph (Diaconis and Janson, 2008, Lovàsz and Szegedy, 2010), i.e. it is defined by a distribution $\mathcal{W}$ depending on two parameters in $[0,1]$ such that for $1 \leq i,j < \infty$:

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

All RCE dissociated graph models are special cases of the W-graph model. More particularly, for binary graphs, their W-graph formulation can be written

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

where $w : [0,1]^2 \to [0,1]$ is some function. For example, one can easily see that the latent block model (a bipartite version of the stochastic block model defined by Govaert and Nadif, 2003) corresponds to the case where w is block constant. Likewise, the bipartite expected degree distribution model (Ouadah et al., 2002) corresponds to the case where w is of product form, i.e. $w(\xi_i, \eta_j) = \int w(\xi_i, \eta) d\eta \times \int w(\xi, \eta_j) d\xi$.

5.2 Example of kernels : Motif counts

A motif is a small subgraphs with a specific pattern. The motif occurrences in the complete network can be counted. Motif counts are known to characterize random networks. They are useful statistics for random graphs as they provide information on the network local structure. Their asymptotic properties are widely studied and a large numbers of studies use motif counts to perform statistical tests. Our framework is particularly well adapted to the use of motif counts for statistical tests as they are closely related to U-statistics with kernel functions of the same size than the motifs. More precisely, in a network of size $m_N \times n_N$, if M_N is the number of occurrences of a certain motif and C_N is the number of different possible positions for this motif, then the relative frequency M_N/C_N can be expressed as a U-statistic.

For example, the motif of Figure 1 has size 2×2 and can be counted with the kernel h_1 with expression

$$\begin{aligned} h_1(Y_{(i_1,i_2;j_1,j_2)}) =& \frac{1}{4}\Big(Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_1}(1-Y_{i_2j_2}) + Y_{i_1j_1}Y_{i_1j_2}Y_{i_2j_2}(1-Y_{i_2j_1}) \\ &+ Y_{i_1j_1}Y_{i_2j_1}Y_{i_2j_2}(1-Y_{i_1j_2}) + Y_{i_1j_2}Y_{i_2j_1}Y_{i_2j_2}(1-Y_{i_1j_1}) \Big). \end{aligned}$$

Its associated U-statistic $U_N^{h_1}$ is equal to M_N/C_N, where M_N is the number of occurrences of this motif and $C_N = 4\binom{m_N}{2}\binom{n_N}{2}$. The value of the expected relative frequency is $U_\infty^{h_1} = \mathbb{E}[h_1(Y_{(i_1,i_2;j_1,j_2)})]$, depending from the model. With the variance estimator defined in the previous section, Corollary 4.3 directly applies to $U_N^{h_1}$, so we have

$$\sqrt{\frac{N}{\widehat{V}_N^{h_1}}}(U_N^{h_1} - U_\infty^{h_1}) \xrightarrow[N\to\infty]{} \mathcal{N}(0,1),$$

which can be used to build confidence intervals for $U_N^{h_1}$ and perform statistical tests.

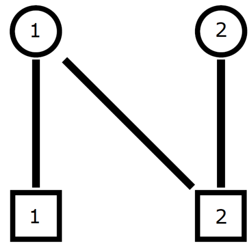

Figure 1: Motif counted by $U_N^{h_1}$.

References

Austern, M., and Orbanz, P. (2022). Limit theorems for distributions invariant under groups of transformations. *The Annals of Statistics*, 50(4), 1960-1991.

Davezies, L., D'Haultfœuille, X., and Guyonvarch, Y. (2021). Empirical process results for exchangeable arrays. *The Annals of Statistics*, 49(2), 845-862.

Diaconis, P., and Janson, S. (2007). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII*, 28(1), 33-61.

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).

Govaert, G., and Nadif, M. (2003). Clustering with block mixture models. *Pattern Recognition*, 36(2), 463-473.

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3), 293-325.

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.

Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles* (Vol. 9). New York: Springer.

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. (2021). U-statistics on bipartite exchangeable network. *arXiv preprint arXiv:2103.12597.*

Le Minh, T. (2023). Hoeffding-type decomposition for U-statistics on bipartite networks. *arXiv preprint arXiv:2308.14518.*

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404-409.

Ouadah, S., Latouche, P., and Robin, S. (2022). Motif-based tests for bipartite networks. *Electronic Journal of Statistics*, 16(1), 293-330.

Silverman, B. W. (1976). Limit theorems for dissociated random variables. *Advances in Applied Probability*, 8(4), 806-819.

Young, S. J., and Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph: 5th International Workshop, WAW 2007, San Diego, CA, USA, December 11-12, 2007. Proceedings* (pp. 138-149). Springer Berlin Heidelberg.

Zhou, T., Ren, J., Medo, M., and Zhang, Y. C. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, 76(4), 046115.