

Edge coherence in multiplex networks

Swati Chandna

*Department of Economics, Mathematics and Statistics,
Birkbeck, University of London, UK
s.chandna@bbk.ac.uk*

Svante Janson

*Department of Mathematics, Uppsala University, Sweden
svante.janson@math.uu.se*

Sofia C. Olhede

*Institute of Mathematics,
Ecole Polytechnique Fédérale de Lausanne, Switzerland
sofia.olhede@epfl.ch*

Abstract

This paper introduces a nonparametric framework for the setting where multiple networks are observed on the same set of nodes, also known as multiplex networks. Our objective is to provide a simple parameterization which explicitly captures linear dependence between the different layers of networks. For non-Euclidean observations, such as shapes and graphs, the notion of “linear” must be defined appropriately. Taking inspiration from the representation of stochastic processes and the analogy of the multivariate spectral representation of a stochastic process with joint exchangeability of Bernoulli arrays, we introduce the notion of edge coherence as a measure of linear dependence in the graph limit space. Edge coherence is defined for pairs of edges from any two network layers and is the key novel parameter. We illustrate the utility of our approach by eliciting simple models such as a correlated stochastic blockmodel and a correlated inhomogeneous graph limit model.

Keywords: Graph Limits; Multiple Networks; Measures of Dependence

1 Introduction

Graphs describe the interaction of objects. More recent work in statistics has acknowledged that graphs are fundamentally random objects and studied their random variation (Kolaczyk et al., 2020). If we can understand the variability of graphs, then it is natural to ask what different moments may be defined to characterise the dependence between graphs. For example (Kolaczyk et al., 2020) has already addressed what the mean network of a set of networks corresponds to, which was also addressed by (Lunagómez et al., 2020) and others, for example (Nielsen and Witten, 2018; Newman, 2018; Pamfil et al., 2020). The question then naturally arises as to how one summarizes multiple networks when more than one network is under study. It is a question related to characterising multiplex and multilayer networks (Pamfil et al., 2020; MacDonald et al., 2020; Bianconi, 2018), or two graphs that are subject to graph matching (Lyzinski et al., 2015).

There is already a growing popularity of multiplex and dynamic network data structures (e.g. Lei et al., 2020; Lunagómez et al., 2020; Zhang and Cao, 2017, to name a few) where two or more networks are observed. A variety of parametric probabilistic models and estimation methods have been proposed in the literature, e.g. latent space approaches such as Arroyo et al. (2021), Gollini and Murphy (2016) and more recently MacDonald et al. (2020). While it is extremely important to develop models which can sufficiently capture the complex structure observed in such data, it is equally important to quantify dependence between pairs of networks and have a framework which allows for the empirical description of correlated networks.

The closest areas of statistics to take inspiration from when studying dependence measures for graphs are dependence measures for other non-Euclidean objects such as in statistical shape analysis (Dryden and Mardia, 2016), summary statistics of random matrices (Ramsay et al., 1984) as well as representations of stochastic processes (Adler and Taylor, 2009). Most stochastic processes admit a representation in terms of a suitable representation theorem, such as the spectral representation theorem (Adler and Taylor, 2009), or the chaotic representation property (Nualart and Schoutens, 2000). The spectral representation of stochastic processes will inspire our choice of joint representation of multiple networks. In the case of networks we have to make some decisions about a suitable measure of correlation—and determine if we want to make connections with linear filtering of stochastic processes (a superficial similarity that will in general break down as linear operations on adjacency matrices do not produce new adjacency matrices of simple graphs), or whether we want a measure that relates to conditional probability of observing an edge in one graph, given we did so in another graph. We resolve these issues by studying the multivariate Bernoulli distribution with the graph limit representation of networks, (Lovász, 2012).

The key to specifying dependence measures is their ability to characterise the intrinsic structure of the objects they describe. We discuss the notion of dependence from a perspective of defining appropriate fundamental representations, which, we believe is crucial to the study of multiple networks. To specify dependence between two (or more) binary networks, we study the bivariate (or multivariate) Bernoulli distribution (Teugels, 1990) with a view to understand how this, in general, is linked to the notion of graph limits (Lovász, 2012) obtained under the framework of joint exchangeability. For the sake of interpretation, joint exchangeability may be considered analogous to joint locally stationary stochastic processes (Tong, 1973), that require a representation in *the same* locally oscillatory family. This naturally yields a dependence measure that we refer to as *edge coherence* arising as a notion of correlation directly from the graph limit representation of a network. For a pair of networks on the same set of n nodes, edge coherence is an $n \times n$ matrix with the ij th entry quantifying linear dependence between edges in the corresponding adjacency matrices.

Our intuition for multivariate analysis is based on the Gaussian case, which in turn relies on Isserlis' (Isserlis, 1918) formula. This no longer holds for Bernoulli random variables and has a number of consequences. For example, to ensure that d Bernoullis are independent, one has to prescribe that the coefficients of the interaction functions are all zero; unlike in the Gaussian case, this is a hierarchy of interaction coefficients (Dai et al., 2013, Theorem 3.1) involving cross-product ratios, that are directly interpretable in terms of log-odds ratios (Whittaker, 1990). Further, the validity of a multivariate Bernoulli model is much harder to determine (Lovison, 2006; Teugels, 1990; Huber and Maric, 2017), and cannot simply be built up from a bivariate understanding. Our derivations show that the assumption of a common latent vector underlying all graphs, is key to ensuring a simple and valid joint representation, as is the case in the study of non-stationary processes. Further, we establish links with decorated graphs (Lovász and Szegedy, 2010) and directed graphs (Diaconis and Janson, 2008), as such data also corresponds to two Bernoulli random variables per prospective edge.

To illustrate the properties and utility of the proposed framework, we discuss how some simple

network models generate dependence. Specifically, we describe a correlated stochastic blockmodel and more generally, a correlated graph limit model with suitable parameterizations, and note how the commonly used choice of scaling for sparse networks, has important implications. Our proposed approach not only builds on classical multivariate methods for network analysis but also aids in interpretation of linear dependence between graphs under the widely used framework of joint exchangeability.

2 Notation and the Bivariate Bernoulli Distributions

We shall start by assuming that we observe two graphs \mathcal{G}_1 and \mathcal{G}_2 with the same set of nodes denoted as $\{1, \dots, n\}$ or $[n]$. Let $\{A_{ij}^{(1)}\}$ and $\{A_{ij}^{(2)}\}$ denote the two sets of edge variables representing two symmetric arrays and let $\underline{A}_{ij} = \begin{pmatrix} A_{ij}^{(1)} & A_{ij}^{(2)} \end{pmatrix}$, where $(i, j) \in [n] \times [n]$ denote the corresponding two-vector valued observation. In general, we may simultaneously consider d networks and likewise define the corresponding d -variate vector for each (i, j) as $\underline{A}_{ij} = \begin{pmatrix} A_{ij}^{(1)} & A_{ij}^{(2)} & \dots & A_{ij}^{(d)} \end{pmatrix}$.

From the characterization of any bivariate Bernoulli vector studied by (Teugels, 1990), we know that three parameters, specifically, two (marginal) success probabilities of observing an edge in either graph that is $E\{A_{ij}^{(1)}\}$ and $E\{A_{ij}^{(2)}\}$, in combination with a co-dependence measure (which may or may not be centered), are required for full specification. If we observe d -variate Bernoullis, then Teugels observes that $2^d - 1$ (centered or non-centred) parameters must be specified to provide the distribution (a parameter per probability of the possible outcomes, with one degree of freedom removed since all distributions are normalized to area 1). To specify co-dependence between the two random variables that are marginally Bernoullis we need to understand an appropriately chosen moment. A simple unnormalized co-dependence measure is obtained by studying the element-wise (Hadamard) matrix product of $A^{(1)}$ and $A^{(2)}$ given by

$$\left(A^{(1)} \circ A^{(2)}\right)_{ij} = A_{ij}^{(1)} A_{ij}^{(2)}, (i, j) \in [n] \times [n].$$

It is clear that the expectation of the Hadamard product corresponds to a joint probability, i.e.

$$E \left\{ \left(A^{(1)} \circ A^{(2)}\right)_{ij} \right\} = \Pr \left\{ A_{ij}^{(1)} = 1 \quad \text{and} \quad A_{ij}^{(2)} = 1 \right\}. \quad (1)$$

This trivial observation yields two (statistical) interpretations of the Hadamard product of adjacency matrices; either as a joint probability of two events occurring (which could easily be manipulated to conditional probabilities, say what is the probability \mathcal{G}_2 has an edge between i and j given \mathcal{G}_1 does) or as an uncentred second moment, naturally transformed into a correlation of edge variables by centering and normalising.

A correlation requires the subtraction of the product of first moments, but both the expectation of the centred and uncentred product naturally satisfy the Cauchy–Schwarz inequality, this leading in each case to a natural choice of normalization. To progress beyond recognizing the moments of (1) we need to introduce a model, and here we shall progress in the framework of scaled graph limits (Bickel and Chen, 2009; Bollobás and Riordan, 2009).

3 Scaled Graph Limits

3.1 Joint exchangeability

A key early assumption to allow for the study of sparse graphs is to posit a model of a scaled graph limit for the observations. To start from our understanding of dense graphs (a graph where the probability of

observing a randomly selected edge remains order one even as we observe larger graphs), let us recall the definition of an exchangeable array (Lauritzen, 2008), namely that an (infinite) array A is jointly exchangeable if for any permutation π it satisfies $(A_{\pi(i)\pi(j)}) \stackrel{d}{=} (A_{ij})$, where $\stackrel{d}{=}$ denotes equivalence in distribution. This definition is normally posed for infinite arrays, as otherwise we need to use the notion of finite exchangeability, rather like a stationary stochastic process is related to a cyclostationary stochastic process (this describing the behaviour of the “boundary” of any array). A consequence of joint exchangeability is the Aldous–Hoover representation theorem (Kallenberg, 2006, Thm 7.22). It states that if (A_{ij}) is jointly exchangeable then there exists a graph limit $f(x, y; \alpha)$, and independent latent variables α and $\{\xi_i\}$, so that

$$A_{ij} \mid \alpha, \xi \stackrel{\text{ind}}{\sim} \text{Bernoulli}\{f(\xi_i, \xi_j; \alpha)\}, \quad i < j, \quad (2)$$

and the array is completed symmetrically setting $A_{ii} = 0$ to correspond to a simple graph. The $\{A_{ij}\}$ are conditionally independent given ξ and α , as we indicate using ind. The assumption of joint exchangeability for a vector-valued binary array which we denote as \mathbf{A} likewise implies distributional invariance under permutation of labels i.e. for any permutation π it follows that

$$(\underline{A}_{\pi(i)\pi(j)}) \stackrel{d}{=} (\underline{A}_{ij}). \quad (3)$$

The Aldous–Hoover theorem (Kallenberg, 2006, Thm 7.22) applies in this setting too, and shows that for an infinite array of d -vectors satisfying (3), there exist independent latent variables α and $\{\xi_i\}$, so that (generalizing (2)) the vectors \underline{A}_{ij} are conditionally independent with d -variate Bernoulli distributions depending on α , ξ_i and ξ_j , i.e.

$$\underline{A}_{ij} \mid \alpha, \xi \stackrel{\text{ind}}{\sim} \text{MultBern}_d\{f(\xi_i, \xi_j; \alpha); f^{(12)}(\xi_i, \xi_j; \alpha), \dots, f^{(d-1d)}(\xi_i, \xi_j; \alpha), \dots\}, \quad (4)$$

where $\text{MultBern}_d\{\}$ is the d -variate Bernoulli distribution (Teugels, 1990), here specified by the $2^d - 1$ cross-moments $f^{(m_1, \dots, m_l)}$, which are enumerated by the set of non-empty subsets $\{m_1, \dots, m_l\}$ of $\{1, \dots, d\}$, and we write

$$\underline{f}(x, y; \alpha) = (f^{(1)}(x, y; \alpha) \quad \dots \quad f^{(d)}(x, y; \alpha))^T, \quad (5)$$

for the vector of the d means. To construct a model for sparse graphs for a finite n -node sample, following (Bickel and Chen, 2009; Bollobás and Riordan, 2009), we modify (2) by introducing a scale factor $\rho_n > 0$, and let the distribution be specified by

$$A_{ij} \mid \alpha, \xi \stackrel{\text{ind}}{\sim} \text{Bernoulli}\{\rho_n f(\xi_i, \xi_j; \alpha)\}, \quad 1 \leq i < j \leq n. \quad (6)$$

Here it is assumed that $f(x, y)$ is a non-negative symmetric function. It is often convenient to further assume that $\|f\|_1 = 1$, since it ensures that ρ_n is a marginal edge probability and makes ρ_n identifiable from the observed graphs. If ρ_n is shrinking with n , then if $f(x, y)$ is bounded, eventually $\rho_n \|f\|_\infty$ will be less than unity; various other fixes can be discussed (cf. e.g. Bollobás et al., 2007, (2.3) and Remark 2.4). Given this, we will simply assume that $f(x, y)$ is bounded and that $\rho_n \|f\|_\infty \leq 1$. In what follows, we shall suppress the indexing on the random variable α . This corresponds to assuming that we observe dissociated random variables.

3.2 Multiplex graphs with $d \geq 2$ layers

To elicit a model for d layers of sparse graphs on a finite number of n nodes, we take inspiration from (4) and (6) and assume that for each (i, j) with $1 \leq i < j \leq n$, a vector \underline{A}_{ij} is generated such that

$$\underline{A}_{ij} | \xi \stackrel{\text{ind}}{\sim} \text{MultBern}_d\{\rho_n \underline{f}(\xi_i, \xi_j); \rho_n^2 f^{(12)}(\xi_i, \xi_j), \dots, \rho_n^2 f^{(d-1 d)}(\xi_i, \xi_j), \dots\}, \quad (7)$$

where the d -variate Bernoulli distribution $\text{MultBern}_d\{\}$ is specified by $2^d - 1$ parameters, which we separate into a mean vector $\rho_n \underline{f}(\xi_i, \xi_j)$ and a higher order structure. The choice of scaling of the second order structure, which in general may be denoted as ρ_n^γ for $\gamma \in \mathbb{Q}_{\geq 0}$ is defined for networks with an increasing number of nodes n . By assuming a square power of ρ_n as n increases i.e. with $\gamma = 2$ we expect covariant behavior of the covariance i.e. the joint probability of an edge in (any) two networks scales as the product of edge probabilities in each network. This is natural for Gaussian data but can be problematic for sparse networks. Alternatively, γ may be chosen to be one to allow concurrent edges at the same rate as in the individual networks. Note that if the moments $f^{(m_1, \dots, m_l)}$ are compatible, then so are the scaled moments $\rho_n^l f^{(m_1, \dots, m_l)}$, since we may obtain the sparse version by first generating a dense array satisfying (4) and then obtain the distribution in (7) by thinning, i.e., (in this case) multiply all $A_{ij}^{(m)}$ by independent $\text{Bern}(\rho_n)$ random variables. Such thinning is independent across layers; we could also use a single thinning operation across all layers. This does not matter marginally and will only make a difference in terms of co-dependence. Also, note that the latent variables are chosen to be common across the d graph layers. Let $p_{ij}^{(m)} = p^{(m)}(\xi_i, \xi_j) = \rho_n f^{(m)}(\xi_i, \xi_j)$, denote the edge probability for node pair (i, j) in the m th graph, where $m \in \{1, \dots, d\}$.

The first thing to note in this setting is that for Bernoulli random variables, the distribution is not specified by the d first order moments and $\binom{d}{2}$ covariances, as it is for Gaussians where then any cross-moment can be computed using Isserlis' theorem (Isserlis, 1918). Care of course needs to be taken not to under-specify (or over-specify) a given distribution. For Bernoulli random variables, in addition to the means and covariances, an additional $2^d - 1 - d - \binom{d}{2}$ moments need to be specified. Of course, when $d = 2$, the total number of parameters $2^d - 1 = 3 = d + \binom{d}{2}$ and thus, no additional parameters are required. The second key observation is that (7) specifies a homogeneous scaling across edge probabilities in the d graph layers. Scaled graphons were introduced in (Bollobás and Riordan, 2009) so that the behaviour of finite sized networks could be studied. For very sparse graphs, much of the structural behaviour is determined by their degree of sparsity. We chose to use a single scaling across all graph layers to simplify the modelling, but fully acknowledge that this is a constraint.

Now, from (7) we note that the marginal correlation of $A_{ij}^{(m_1)}$ with $A_{kl}^{(m_2)}$ is specified through forward modelling and using the theorem of total covariance it follows that

$$\text{cov}\{A_{ij}^{(m_1)}, A_{kl}^{(m_2)}\} = \text{cov}_\xi\{E_{A|\xi}\{A_{ij}^{(m_1)}\}, E_{A|\xi}\{A_{kl}^{(m_2)}\}\} + E_\xi\{\text{cov}_{A|\xi}\{A_{ij}^{(m_1)}, A_{kl}^{(m_2)}\}\},$$

where $1 \leq i < j \leq n$ and $1 \leq k < l \leq n$, as well as $1 \leq m_1, m_2 \leq d$. We see that this decomposition directly allows us to characterise two types of dependence; dependence that is intrinsically bivariate specified from the latent variables ξ , but also dependence specified conditionally on the latent variables. If we observe only one array then one can use this decomposition to show that if the array is dissociated then A_{ij} and A_{il} have non-negative covariance bounded above by $(1/2)\text{var}\{A_{ij}\}$; this limits also the correlation of this variable. If the array is associated, then the covariance of A_{ij} and A_{il} is strictly positive (Aldous, 1985; Bassan and Scarsini, 1998). Furthermore, for two arrays, requiring that $\text{cov}_{A|\xi}\{A_{ij}^{(1)}, A_{ij}^{(2)}\} = 0$ is equivalent to requiring $f^{(12)}(x, y) = f^{(1)}(x, y)f^{(2)}(x, y)$. This may be realistic in some scenarios, but as an assumption pushes all correlation between processes into the latent

process. This may constrain what marginal probability functions we can use, not to mention what correlations are achievable.

Why does this matter? The problem of covariance is that while forward modelling of the dependence of A_{ij} is trivial, inverse specification of an arbitrary correlation for marginally Bernoulli variables is not (Chaganty and Joe, 2006), and at times we will wish to achieve a negative correlation between edges, especially if edges require the same resource to form. For the purposes of estimation with only one realization of \underline{A}_{ij} for $1 \leq i < j \leq n$ where $n \in \mathbb{N}$ we can only implement conditional inference, conditionally on the realized value of α , no matter if $n \gg 1$. This is consistent with recent results in estimating subgraph densities from noisy networks (Chang et al., 2020).

We note that the model elicited above can alternatively be specified in terms of centred moments, or directly by the vector of the 2^d (conditional) probabilities as

$$\Pr\{\underline{A}_{ij} = \underline{a} \mid \xi\} = p_{\underline{a}}(\xi_i, \xi_j), \quad \underline{a} = (a_1, \dots, a_d) \in \{0, 1\}^d, \quad (8)$$

where \underline{a} denotes the presence or absence of an edge between nodes i and j in graphs $\{1, \dots, d\}$, see (Teugels, 1990), where conversion formulas are given. The functions $f^{(m_1, \dots, m_l)}(x, y)$ are symmetric in (x, y) i.e. $f^{(m_1, \dots, m_l)}(x, y) = f^{(m_1, \dots, m_l)}(y, x)$, and similarly for $p_{\underline{a}}$ in (8). Moreover, the moments $f^{(m_1, \dots, m_l)}$ must be consistently specified (Huber and Maric, 2017; Chaganty and Joe, 2006, e.g.). Other discussions of the multivariate Bernoulli distribution can be found in (Joe, 1997; Whittaker, 1990).

For simplicity, we turn our focus to multiplex networks in the simplest setting of $d = 2$ layers. In this case we have for $1 \leq i < j \leq n$,

$$\underline{A}_{ij} \mid \xi_i, \xi_j \stackrel{\text{ind}}{\sim} \text{MultBern}_2\{\rho_n \underline{f}(\xi_i, \xi_j); \rho_n^2 f^{(12)}(\xi_i, \xi_j)\}, \quad (9)$$

where $\text{MultBern}_2\{\}$ is denoting the two-vector distribution, and $f(x, y) = (f^{(1)}(x, y) \quad f^{(2)}(x, y))^T$. This distribution for a fixed i and j is specified by three moments, chosen as the two means, and the cross-moment as a co-dependence measure. In some settings log odds ratio are advised for this modelling (Whittaker, 1990). Alternatively, we may specify the distribution by four functions $p_{\underline{a}}$ as in (8); these sum to 1, so we only have to specify three of them. There are simple algebraic relations between these functions and the functions in (9) given by

$$\rho_n f^{(1)}(x, y) = p_{10}(x, y) + p_{11}(x, y) \quad (10)$$

$$\rho_n f^{(2)}(x, y) = p_{01}(x, y) + p_{11}(x, y) \quad (11)$$

$$\rho_n^2 f^{(12)}(x, y) = p_{11}(x, y). \quad (12)$$

This representation allows us to further note linkages between (7) and the representation of directed graphs (Diaconis and Janson, 2008). Here directed graphs are parameterized via three independent measurable functions, that we shall denote $g_{00}(x, y; \alpha)$, $g_{01}(x, y; \alpha)$ and $g_{10}(x, y; \alpha)$ where a fourth function is defined by the theorem of total probability implicitly as $g_{11}(x, y; \alpha) = 1 - g_{00}(x, y; \alpha) - g_{01}(x, y; \alpha) - g_{10}(x, y; \alpha)$. These functions correspond to the functions $p_{kl}(x, y; \alpha)$ above; however, the symmetry requirements on the functions $g_{kl}(x, y; \alpha)$ are different than those imposed on $p_{kl}(x, y; \alpha)$, because directed edges correspond to a kind of antisymmetric relations between nodes rather than symmetric relations, (e.g. see Diaconis and Janson (2008) for details).

4 Edge Coherence

We should start by noting that the terminology ‘‘coherence’’ has been used in another context for networks (Ding et al., 2015; Patterson and Bamieh, 2011) and has a different meaning. Our motivation

for re-using the term comes from stochastic processes, especially time series. Some care must be used in understanding the analogy but not over-interpreting it. For a time series, coherence captures linear dependence between zero-mean stochastic processes, and convolutions (“filtering”) in the temporal domain, correspond to linear relationships in the frequency domain. Of course we cannot entertain linear relationships between networks—as linear combinations of adjacency matrices of simple graphs do not yield adjacency matrices of simple graphs. Linear combinations of graph limits with positive weights do yield new graph limit functions, as long as overall scaling is treated carefully and the graph limit does not when scaled exceed unity. To write down an appropriate definition of a dependence measure we need to determine what we are quantifying. The marginal correlation between edges in two networks is given by

$$\sigma_{ij} = \frac{E\{A_{ij}^{(1)} A_{ij}^{(2)}\} - E\{A_{ij}^{(1)}\}E\{A_{ij}^{(2)}\}}{[\text{var}\{A_{ij}^{(1)}\}\text{var}\{A_{ij}^{(2)}\}]^{1/2}}. \quad (13)$$

Of course estimating this quantity will require some additional assumptions, and we need to specify what expectations we are calculating. Calculating the expectations conditionally on ξ , we get the conditional correlation

$$\begin{aligned} r(\xi_i, \xi_j) &= \frac{E_{A|\xi}\{A_{ij}^{(1)} A_{ij}^{(2)}\} - E_{A|\xi}\{A_{ij}^{(1)}\}E_{A|\xi}\{A_{ij}^{(2)}\}}{[\text{var}_{A|\xi}\{A_{ij}^{(1)}\}\text{var}_{A|\xi}\{A_{ij}^{(2)}\}]^{1/2}} \\ &= \frac{\rho_n [f^{(12)}(\xi_i, \xi_j) - f^{(1)}(\xi_i, \xi_j)f^{(2)}(\xi_i, \xi_j)]}{[f^{(1)}(\xi_i, \xi_j) \{1 - \rho_n f^{(1)}(\xi_i, \xi_j)\} f^{(2)}(\xi_i, \xi_j) \{1 - \rho_n f^{(2)}(\xi_i, \xi_j)\}]^{1/2}}. \end{aligned} \quad (14)$$

Note that, in eliciting the multiplex model with two layers in (7), we had chosen the cross-moment $\rho_n^2 f^{(12)}$ as the co-dependence measure. Equivalently, we may use the coherence function as defined above.

We now ask, what does it mean if $r(\xi_i, \xi_j) = 0$ or if $|r(\xi_i, \xi_j)| = 1$? In the former case the conditional (on ξ) probability of an edge in both the layers simply corresponds to the product of the marginal probabilities, as would be expected. On the other hand, perfect correlation between Bernoulli variables means that they coincide. This means that the marginal probabilities are the same, so that $f^{(1)}(\xi_i, \xi_j) = f^{(2)}(\xi_i, \xi_j) = f(\xi_i, \xi_j)$, and, moreover, that the expected value of the Hadamard product is the same as the expected value of either of the two individual adjacencies, i.e., $\rho_n^2 f^{(12)}(\xi_i, \xi_j) = \rho_n f(\xi_i, \xi_j)$, or

$$\rho_n f^{(12)}(\xi_i, \xi_j) = f(\xi_i, \xi_j).$$

Similarly, $r(\xi_i, \xi_j) = -1$ is possible only if $A_{ij}^{(1)} = 1 - A_{ij}^{(2)}$ a.s., and thus $\rho_n (f^{(1)}(\xi_i, \xi_j) + f^{(2)}(\xi_i, \xi_j)) = 1$, which is impossible in the sparse case (for large n and thus small ρ_n).

In general, for a bivariate Bernoulli vector to be well-defined, the moments characterizing it must satisfy certain inequalities. Specifically, we can only specify $(\rho_n, \underline{f}(x, y), r(x, y))$ such that (Chaganty and Joe, 2006)

$$\begin{aligned} &\max \left\{ - \left(\frac{\rho_n^2 f^{(1)}(x, y) f^{(2)}(x, y)}{(1 - \rho_n f^{(1)}(x, y))(1 - \rho_n f^{(2)}(x, y))} \right)^{1/2}, - \left(\frac{(1 - \rho_n f^{(1)}(x, y))(1 - \rho_n f^{(2)}(x, y))}{\rho_n^2 f^{(1)}(x, y) f^{(2)}(x, y)} \right)^{1/2} \right\} \\ &\leq r(x, y) \leq \\ &\min \left\{ \left(\frac{f^{(1)}(x, y)(1 - \rho_n f^{(2)}(x, y))}{f^{(2)}(x, y)(1 - \rho_n f^{(1)}(x, y))} \right)^{1/2}, \left(\frac{(1 - \rho_n f^{(1)}(x, y))f^{(2)}(x, y)}{f^{(1)}(x, y)(1 - \rho_n f^{(2)}(x, y))} \right)^{1/2} \right\}, \quad \forall (x, y) \in [0, 1]^2. \end{aligned} \quad (15)$$

Therefore, not all possible choices can be realized in (9). Negative correlation can be problematic, as observed in other settings, see for example McKenzie (2003) or Andel and Dosla (2010).

For a sparse graph, $r(\xi_i, \xi_j)$ is not an order one quantity, which will continue to decrease with n , that should eventually be estimated as zero. We therefore define the edge coherence function to be given by

$$r_0(\xi_i, \xi_j) = r(\xi_i, \xi_j) / \rho_n, \quad (16)$$

as a proper object behaving consistently with increasing n as long the homogeneous scaling assumption in ρ_n is correct. Alternatively, if cross-moments in (7) were chosen to scale as ρ i.e. $\gamma = 1$, then $r(\xi_i, \xi_j)$ is an $O(1)$ quantity, and we define edge coherence by $r_0(\xi_i, \xi_j) = r(\xi_i, \xi_j)$ instead of (16).

Note that our choice of coherence in (16) is centred, i.e., we chose to remove a product of the means of both arrays $\{A_{ij}^{(1)}\}$ and $\{A_{ij}^{(2)}\}$, and study fluctuations around this product of means, suitably standardized. If $\rho_n \rightarrow 0$ then

$$r_0(\xi_i, \xi_j) = \frac{f^{(12)}(\xi_i, \xi_j) - f^{(1)}(\xi_i, \xi_j)f^{(2)}(\xi_i, \xi_j)}{\{f^{(1)}(\xi_i, \xi_j)f^{(2)}(\xi_i, \xi_j)\}^{1/2}} \{1 + o(1)\}, \quad (17)$$

and apart from the mean correction, this form resembles the coherence for a stochastic process (Adler and Taylor, 2009). There are other ratios we could have defined in (17). Notice that $E_{A|\xi}\{A_{ij}^{(1)}A_{ij}^{(2)}\} = \Pr\{A_{ij}^{(1)} = 1 \cap A_{ij}^{(2)} = 1 | \xi\}$, and so when ρ_n is appreciable, and the marginal probabilities are the same, then (17) before the mean was subtracted can be interpreted as a mean (and sparsity) corrected partial correlation between edges. For stochastic processes, the interpretation of coherence is in terms of linear prediction of zero-mean random variables, which is not the case for networks, where the mean structure is important.

Finally, we might ask, how do we deal with the notion of coherence for more than two graphs? We can collect any of the $\binom{d}{2}$ pairwise measures of dependence and collect them in a data structure. This will not completely characterise the multivariate structure of the d variables, but will provide a simple summary of their dependence structure, like using correlation for non-Gaussian but Euclidean data structures.

5 Models That Generate Correlated Graphs

5.1 Correlated Stochastic Blockmodel

We elicit the correlated stochastic blockmodel (CSBM) for the simplest case of $d = 2$ layers. The general case follows by applying the definitions here to all possible pairs of graphs. Let K denote the number of blocks in each graph and let $z_i \in \{1, \dots, K\}$ where $|\{i, z_i = a\}| = h_a$, provide the block label to which node i belongs. Normally $\{z_i\}$ are unobserved and part of the inference problem is determining how to cluster the observed network (e.g. Olhede and Wolfe (2014)). Let $\Theta^{(1)} = (\theta_{ab}^{(1)})$ and likewise $\Theta^{(2)} = (\theta_{ab}^{(2)})$ where $(a, b) \in [K] \times [K]$, denote the $K \times K$ edge-probability matrix corresponding to $A^{(1)}$ and $A^{(2)}$, respectively, i.e.

$$P(A_{ij}^{(l)} = 1 | z_i = a, z_j = b) = \theta_{ab}^{(l)}, \quad l \in \{1, 2\}. \quad (18)$$

Given nodes i and j such that $z_i = a$ and $z_j = b$, let r_{ab} denote the coherence between blocks a and b , i.e. $r_{ab} = \{r_{ij} | z_i = a, z_j = b\}$, where $r_{ij} \equiv r(\xi_i, \xi_j)$ given by (14) is the conditional correlation between $A_{ij}^{(1)}$ and $A_{ij}^{(2)}$. Then $R = (r_{ab})$, denotes the full $K \times K$ coherence matrix specifying conditional correlation between the two network layers. Thus, given K , the correlated stochastic

blockmodel is parameterized via $(z, \Theta^{(1)}, \Theta^{(2)}, R)$ where following the time series terminology, we refer to $\Theta^{(l)}, l = 1, 2$ as the two auto-interaction parameters and R as the coherence parameter. Alternatively, we may parameterize the model via the uncentred joint moment $\varrho_{ab} = E\{A_{ij}^{(1)} A_{ij}^{(2)}\}$, (e.g. as in Pamfil et al., 2020). This is naturally linked to the coherence parameter since $r_{ab} = (\varrho_{ab} - \theta_{ab}^{(1)} \theta_{ab}^{(2)}) / (\theta_{ab}^{(1)} \{1 - \theta_{ab}^{(1)}\} \theta_{ab}^{(2)} \{1 - \theta_{ab}^{(2)}\})^{1/2}$.

Given the model parameters as described above, a pair of correlated SBM networks may be generated easily following (Lunn and Davies, 1998). For simplicity, assume that $\Theta^{(1)} = \Theta^{(2)} = (\theta_{ab})$. Then $(A^{(1)}, A^{(2)})$ are networks from the corresponding bivariate SBM, if for $i < j$ independently, we set

$$\begin{aligned} A_{ij}^{(1)} &= \{1 - V_{ij}^{(1)}\} U_{ij}^{(1)} + V_{ij}^{(1)} W_{ij} \\ A_{ij}^{(2)} &= \{1 - V_{ij}^{(2)}\} U_{ij}^{(2)} + V_{ij}^{(2)} W_{ij} \\ U_{ij}^{(k)} | z &\stackrel{d}{=} \text{Bern}\{\theta_{z_i z_j}\}, \quad V_{ij}^{(k)} | z \stackrel{d}{=} \text{Bern}\{r_{z_i z_j}\}, \quad k = 1, 2, \\ W_{ij} | z &\stackrel{d}{=} \text{Bern}\{\theta_{z_i z_j}\}, \end{aligned} \tag{19}$$

where $U_{ij} = (U_{ij}^{(1)}, U_{ij}^{(2)})^T$ and $V_{ij} = (V_{ij}^{(1)}, V_{ij}^{(2)})^T$ each contain two (independent) Bernoullis and W is a scalar Bernoulli. The label vector $z = (z_1, \dots, z_n)^T$ follows a multinomial distribution (as in the standard one-dimensional SBM). Note that we can only generate Bernoullis for $V_{ij}^{(k)}$ if the specified mean r_{ab} is non-negative. Now, clearly, conditionally on $z_i = a$ and $z_j = b$, the above mechanism leads to

$$\begin{aligned} E\{A_{ij}^{(1)} | z\} &= \{1 - r_{ab}\} \theta_{ab} + r_{ab} \theta_{ab} = \theta_{ab} \\ E\{A_{ij}^{(2)} | z\} &= \{1 - r_{ab}\} \theta_{ab} + r_{ab} \theta_{ab} = \theta_{ab} \\ \text{cov}\{A_{ij}^{(1)}, A_{ij}^{(2)} | z\} &= \text{cov}\{\{1 - V_{ij}^{(1)}\} U_{ij}^{(1)} + V_{ij}^{(1)} W_{ij}, \{1 - V_{ij}^{(2)}\} U_{ij}^{(2)} + V_{ij}^{(2)} W_{ij}\}. \end{aligned}$$

To understand second moments we note that

$$\begin{aligned} \text{cov}\{A_{ij}^{(1)}, A_{ij}^{(2)} | z\} &= \text{var}\{W_{ij}\} E\{V_{ij}^{(1)}\} E\{V_{ij}^{(2)}\} \\ &= \theta_{ab} \{1 - \theta_{ab}\} r_{ab}^2. \end{aligned}$$

This allows us to generate positively correlated random variables, as if we assume that $z_i = a$ and $z_j = b$, then $\text{corr}\{A_{ij}^{(1)}, A_{ij}^{(2)}\} = r_{ab}^2$. Generating negatively correlated Bernoullis is relatively challenging but possible, and requires a different scheme to realize (e.g. Andel and Dosla, 2010), already adressed in time series.

5.2 Correlated Homogeneous Graph Limit Model

The correlated stochastic blockmodel clearly provides a simplified description for a pair of networks via a block-constant coherence matrix R . In this coherence matrix, the diagonal and off-diagonal blocks correspond to coherence between graphs, for pairs of node within blocks and across blocks, respectively. The easiest extension to graph limits is to directly follow the stochastic blockmodel and replace the blockmodel specification by the graph limit functions $f^{(1)}(x, y)$ and $f^{(2)}(x, y)$ and a scale factor ρ_n as in (6), coupled with a coherence function $r_0(x, y)$ as defined by (16). As in our specification of the correlated stochastic blockmodel, assume that $f^{(1)}(x, y) = f^{(2)}(x, y) = f(x, y)$, so that marginally $A_{ij}^{(l)} | \xi \sim \text{Bern}\{\rho_n f(\xi_i, \xi_j)\}$, $l = 1, 2$. Then given a (positive) correlation function $r(x, y)$, the corresponding correlated pair of networks follow via the scheme given in (19), except with ξ

instead of z ; $\rho_n f(\xi_i, \xi_j)$ instead of $\theta_{z_i z_j}$ and $\rho_n \sqrt{r(\xi_i, \xi_j)}$ instead of $r_{z_i z_j}$. This then produces a 2-layer correlated homogeneous graph limit model with the appropriate graph limits (please refer to calculations of the previous section).

5.3 Blockmodel to graph limits

The general 2-layer correlated stochastic blockmodel may be studied as it is or via a graph limit parameterization. To construct the corresponding graph limit objects we proceed as follows. Define $H_0 = 0$, and iteratively the set of constants $H_a = \sum_{k=1}^a h_k$, $a \in \{1, \dots, K\}$. Then for $l \in \{1, 2\}$, we define

$$f^{(l)}(x, y) = \sum_{a,b=1}^k \theta_{ab}^{(l)} \mathbf{1}(x \in (H_{a-1}/n, H_a/n]) \mathbf{1}(y \in (H_{b-1}/n, H_b/n]), \quad (20)$$

as the (marginal) graph limit associated with graph l . The Hadamard product has expectation ϱ_{ab} . To ensure a well-defined coherence quantity, we assume that $0 \leq \varrho_{ab}$. This cross-moment leads us to the cross-graph limit, which is given by

$$f^{(12)}(x, y) = \sum_{a,b=1}^k \varrho_{ab} \mathbf{1}(x \in (H_{a-1}/n, H_a/n]) \mathbf{1}(y \in (H_{b-1}/n, H_b/n]).$$

This corresponds to a bivariate representation of the blockmodel. To go from ϱ_{ab} to the coherence we define

$$r(x, y) = \sum_{a,b=1}^k \frac{\varrho_{ab} - \theta_{ab}^{(1)} \theta_{ab}^{(2)}}{\sqrt{\theta_{ab}^{(1)} \theta_{ab}^{(2)} (1 - \theta_{ab}^{(1)}) (1 - \theta_{ab}^{(2)})}} \mathbf{1}(x \in (H_{a-1}/n, H_a/n]) \mathbf{1}(y \in (H_{b-1}/n, H_b/n]).$$

5.4 Correlated Inhomogeneous Graph Limit Model

Given a pair of graphs $(\mathcal{G}_1, \mathcal{G}_2)$, observed on the same set of nodes, a common question of interest is the similarity between them, or more generally, whether the presence of an edge between nodes i and j in \mathcal{G}_1 implies the corresponding edge in \mathcal{G}_2 . With this in mind, and proceeding in an analogy with what is known in signal processing, as an “input–output” relationship, we consider the following approach to modeling relationship between two graphs:

$$A_{ij}^{(2)} | \alpha, \xi, A_{ij}^{(1)} \sim \text{Bern} \left\{ h(\xi_i, \xi_j) A_{ij}^{(1)} \right\}, \quad 1 \leq j < i \leq n, \quad (21)$$

where we may interpret $A_{ij}^{(1)}$ as an input to obtain an output $A_{ij}^{(2)}$. This may also be viewed as “modulation” as it is similar to amplitude and frequency modulation in signal processing, being a multiplicative process. Note that (21) implies that $A_{ij}^{(2)} = 1 \implies A_{ij}^{(1)} = 1$, so $\mathcal{G}_2 \subseteq \mathcal{G}_1$.

Let $f^{(1)}(x, y)$ and $f^{(2)}(x, y)$ denote graph limit functions corresponding to $A^{(1)}$ and $A^{(2)}$ (e.g. (6)) and let $r(x, y)$ denote the limiting correlation function. Let $\rho_n f^{(1)}(\xi_i, \xi_j) = p_{ij}^{(1)}$. To parameterize $A_{ij}^{(2)}$ given $A_{ij}^{(1)}$, we calculate its expectation using the law of iterated expectation, which is given by

$$\begin{aligned} E\{A_{ij}^{(2)} | \xi\} &= E_{A_{ij}^{(1)} | \xi} \{E\{A_{ij}^{(2)} | \xi, A_{ij}^{(1)}\}\} = E_{A_{ij}^{(1)} | \xi} \{h(\xi_i, \xi_j) A_{ij}^{(1)}\} \\ &= E_{A_{ij}^{(1)} | \xi} \{\rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j)\} = \rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j) \\ &= h(\xi_i, \xi_j) p_{ij}^{(1)} = p_{ij}^{(2)}, \end{aligned}$$

where we defined $p_{ij}^{(2)}$ as a pointwise product of h and $p_{ij}^{(1)}$. Furthermore we find that the expected value of the Hadamard product under (21) is given by

$$\begin{aligned} E\left\{\left(A^{(1)} \circ A^{(2)}\right)_{ij} \mid \xi\right\} &= E\{h(\xi_i, \xi_j)A_{ij}^{(1)} \mid \xi\} = \rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j) \equiv p_{ij}^{(12)}, \\ &= \Pr\{A_{ij}^{(1)} = 1 \cap A_{ij}^{(2)} = 1\}. \end{aligned} \quad (22)$$

Clearly, $p_{ij}^{(12)} = p_{ij}^{(2)}$, and thus, the probability of an edge in \mathcal{G}_2 is equal to the probability of the corresponding edge occurring in both \mathcal{G}_1 and \mathcal{G}_2 . Thus, $A_{ij}^{(2)} = 1$ with a high probability implies a high probability of the edge occurring in both the layers or $A_{ij}^{(1)} = 1$. Also, clearly $p_{ij}^{(1)} = 0$, implies $p_{ij}^{(2)} = 0$. Further implications can be understood by noting the the probabilities associated with the four possible outcomes, as follows

$$\begin{aligned} \Pr\{A_{ij}^{(1)} = 0 \cap A_{ij}^{(2)} = 1\} &= \Pr\{A_{ij}^{(2)} = 1\} - \Pr\{A_{ij}^{(1)} = 1 \cap A_{ij}^{(2)} = 1\} \\ &= \rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j) - \rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j) \\ &= 0 \\ \Pr\{A_{ij}^{(1)} = 1 \cap A_{ij}^{(2)} = 0\} &= \Pr\{A_{ij}^{(1)} = 1\} - \Pr\{A_{ij}^{(1)} = 1 \cap A_{ij}^{(2)} = 1\} \\ &= \rho_n f^{(1)}(\xi_i, \xi_j) - \rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j) \\ &= \rho_n f^{(1)}(\xi_i, \xi_j) \{1 - h(\xi_i, \xi_j)\} \\ \Pr\{A_{ij}^{(1)} = 0 \cap A_{ij}^{(2)} = 0\} &= 1 - \rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j) - \rho_n f^{(1)}(\xi_i, \xi_j) \{1 - h(\xi_i, \xi_j)\} \\ &= 1 - \rho_n f^{(1)}(\xi_i, \xi_j). \end{aligned}$$

Note that the scaling in this model is not homogeneous as the probability of an edge in the second graph will scale like the probability of having the corresponding edge in both the graphs, which the generating mechanism in a sense is regulating. The conditional correlation under this model, takes the form

$$\begin{aligned} r(\xi_i, \xi_j) &= \frac{p_{ij}^{(12)} - p_{ij}^{(1)} p_{ij}^{(2)}}{[\text{var}\{A_{ij}^{(1)}\} \text{var}\{A_{ij}^{(2)}\}]^{1/2}}, \quad 1 \leq j < i \leq n, \\ &= \left[\frac{h(\xi_i, \xi_j) \{1 - \rho_n f^{(1)}(\xi_i, \xi_j)\}}{1 - \rho_n h(\xi_i, \xi_j) f^{(1)}(\xi_i, \xi_j)} \right]^{1/2}. \end{aligned} \quad (23)$$

Clearly, as $\rho_n \rightarrow 0$, this is a finite order one quantity with the leading term given by $h(\xi_i, \xi_j)^{1/2}$. Thus, coherence $r_0(\xi_i, \xi_j)$ under this model is set equal to the conditional correlation $r(\xi_i, \xi_j)$. This is in contrast to what we observed under the general model described in Section 3 and shows that the choice of scaling has important implications. Naturally still the Cauchy–Schwartz inequality applies in this setting and so the magnitude of $r(\xi_i, \xi_j)$ is bounded.

6 Discussion

Introducing correlation between graphs is a natural next step in modelling graphs, thereby acknowledging dependence also in this non-Euclidean setting. This clearly goes beyond the mean of a graph, a topic with significant interest (Lunagómez et al., 2020). Graphs can naturally also be dependent without being perfectly replicated. In graph theory the notion of graph limits were partially introduced to use tools

from analysis in a combinatorics setting. Our framework shows how simple ‘linear’ operations may be interpreted in the space of graph limits. Further, we note that marginally the edge variables in a single graph are correlated when one of the indices are fixed to be the same. Edges in exchangeable models, and scaled exchangeable arrays, can only exhibit positive correlation inside a single exchangeable graph (Aldous, 1985, p.133). Once we move between two “layers” or graphs on the same nodes, negative correlation can be generated. One simple way of doing so would be to take $f^{(1)}(x, y) = g(x, y)$ and $f^{(2)}(x, y) = 1 - c \cdot g(x, y)$ for some positive c and function $g(x, y)$, (e.g. Chaganty and Joe, 2006). Of course as the dimension of the Bernoulli vector increases data structures become more difficult to characterise (Huber and Maric, 2017). Given this, for multiplex networks with more than two layers, one may proceed by applying the proposed framework to all possible pairs of network layers. Thus, reducing a d -dimensional model to $d(d - 1)/2$ two-dimensional models parameterized by edge coherence. Edge coherence between a pair of networks is a weighted graph itself. It summarizes their dependence structure and may be used for key applications such as detecting changes in second-order dependence structure across network layers observed over time or space.

7 Acknowledgements

This work was supported by the European Research Council under Grant CoG 2015-682172NETS, within the Seventh European Union Framework Program and the Knut and Alice Wallenberg Foundation.

References

- Adler, R. J. and Taylor, J. E. (2009). *Random fields and geometry*. Springer Science & Business Media.
- Aldous, D. (1985). Exchangeability and related topics. In *Ecole d’Ete St Flour 1983*. Springer Verlag, Berlin, Germany.
- Andel, J. and Dosla, S. (2010). Bernoulli processes with non-positive correlations. *Communications in Statistics—Theory and Methods*, 39(17):3191–3201.
- Arroyo, J., Athreya, A., Cape, J., Chen, G., Priebe, C. E., and Vogelstein, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *Journal of Machine Learning Research*, 22(142):1–49.
- Bassan, B. and Scarsini, M. (1998). Finite row-column exchangeable arrays. *Commentationes Mathematicae Universitatis Carolinae*, 39(1):137–145.
- Bianconi, G. (2018). *Multilayer networks: structure and function*. Oxford University Press.
- 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.
- Bollobás, B., Janson, S., and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31:3–122.
- Bollobás, B. and Riordan, O. (2009). Metrics for sparse graphs. *Surveys in Combinatorics 2009*, 211–287. *LMS Lecture Notes Series*, 365.

- Chaganty, N. R. and Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika*, 93(1):197–206.
- Chang, J., Kolaczyk, E. D., and Yao, Q. (2020). Estimation of subgraph densities in noisy networks. *Journal of the American Statistical Association*, pages 1–14.
- Dai, B., Ding, S., and Wahba, G. (2013). Multivariate Bernoulli distribution. *Bernoulli*, 19(4):1465–1483.
- Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII.*, pages 33–61.
- Ding, Q., Sun, W., and Chen, F. (2015). Network coherence in the web graphs. *Communications in Nonlinear Science and Numerical Simulation*, 27(1-3):228–236.
- Dryden, I. L. and Mardia, K. V. (2016). *Statistical shape analysis: with applications in R*, volume 995. John Wiley & Sons.
- Gollini, I. and Murphy, T. B. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265.
- Huber, M. and Maric, N. (2017). Bernoulli correlations and cut polytopes. *arXiv preprint arXiv:1706.06182*.
- Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139.
- Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.
- Kallenberg, O. (2006). *Probabilistic symmetries and invariance principles*. Springer Science & Business Media, Berlin, Germany.
- Kolaczyk, E. D., Lin, L., Rosenberg, S., Walters, J., and Xu, J. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *The Annals of Statistics*, 48(1):514–538.
- Lauritzen, S. L. (2008). Exchangeable Rasch matrices. *Rend. Mat. Appl.*, 28(7):83–95.
- Lei, J., Chen, K., and Lynch, B. (2020). Consistent community detection in multi-layer network data. *Biometrika*, 107(1):61–73.
- Lovász, L. (2012). *Large networks and graph limits*, volume 60. American Mathematical Soc.
- Lovász, L. and Szegedy, B. (2010). Limits of compact decorated graphs. *arXiv preprint arXiv:1010.5155*.
- Lovison, G. (2006). A matrix-valued Bernoulli distribution. *Journal of Multivariate Analysis*, 97(7):1573–1585.
- Lunagómez, S., Olhede, S. C., and Wolfe, P. J. (2020). Modeling network populations via graph distances. *Journal of the American Statistical Association*, pages 1–18.
- Lunn, A. D. and Davies, S. J. (1998). A note on generating correlated binary variables. *Biometrika*, 85(2):487–490.

- Lyzinski, V., Sussman, D. L., Fishkind, D. E., Pao, H., Chen, L., Vogelstein, J. T., Park, Y., and Priebe, C. E. (2015). Spectral clustering for divide-and-conquer graph matching. *Parallel Computing*, 47:70–87.
- MacDonald, P. W., Levina, E., and Zhu, J. (2020). Latent space models for multiplex networks with shared structure. *arXiv preprint arXiv:2012.14409*.
- McKenzie, E. (2003). Ch. 16. discrete variate time series. *Handbook of statistics*, 21:573–606.
- Newman, M. E. J. (2018). Network structure from rich but noisy data. *Nature Physics*, 14(6):542–545.
- Nielsen, A. M. and Witten, D. (2018). The multiple random dot product graph model. *arXiv preprint arXiv:1811.12172*.
- Nualart, D. and Schoutens, W. (2000). Chaotic and predictable representations for Lévy processes. *Stochastic processes and their applications*, 90(1):109–122.
- Olhede, S. C. and Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41):14722–14727.
- Pamfil, A. R., Howison, S. D., and Porter, M. A. (2020). Inference of edge correlations in multilayer networks. *Physical Review E*, 102(6):062307.
- Patterson, S. and Bamieh, B. (2011). Network coherence in fractal graphs. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 6445–6450. IEEE.
- Ramsay, J. O., Berge, J. T., and Styán, G. P. H. (1984). Matrix correlation. *Psychometrika*, 49(3):403–423.
- Teugels, J. L. (1990). Some representations of the multivariate Bernoulli and binomial distributions. *Journal of Multivariate Analysis*, 32(2):256–268.
- Tong, H. (1973). Some comments on spectral representations of non-stationary stochastic processes. *Journal of Applied Probability*, pages 881–885.
- Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Wiley Publishing.
- Zhang, J. and Cao, J. (2017). Finding common modules in a time-varying network with application to the drosophila melanogaster gene regulation network. *Journal of the American Statistical Association*, 112(519):994–1008.