Testing independence in networks via family of network metrics

Abstract

Propelled by increasing demand and supply in network data, investigating whether network structures are associated with the attributes of interest has been an important concern in natural or social science. We consider the problem of network dependence, which refers to any types of dependence between network topology and its nodal attributes and propose the method to test network independence. However due to the interdependency in constructing network, standard independence test cannot be directly applied but instead the former network independence tests have introduced network models primarily focusing on globally persistent dependence. To overcome these challenges, we introduce a nonparametric multiscale test statistic which is robust to both high dimensionality and nonlinearity by utilizing a family of network geometries. Our simulation studies demonstrate the outstanding performance of the method under various circumstances. We also suggest the measure to quantify each node's contribution to detecting dependency, which is applied to the political network in the end.

1 Introduction

Statisticians have long considered the problem of revealing the relationship between two different data sets. Above all determining the existence of any association or any dependence would be the first step in characterizing the relationship. Independence test between two univariate random variables, for example, could start from deriving a correlation coefficient or estimating a coefficient from linear regression model. However as types of data have diversified or dimension of the data has increased, various forms of multivariate independence tests have been suggested (Heller et al., 2012; Székely et al., 2007). We consider independence test upon **non-traditional but ubiquitous dataset of *network* which is very likely to possess the properties of both high dimensionality and nonlinearity**. Network, formally defined as a collection of nodes and edges, has been suffered from a dearth of proper analysis due to its distinct way to be constructed. In this paper we define any kinds of dependency between network topology and nodal attributes as *network dependence* and propose a method to test network independence considering unique properties of network.

The literature on detecting dependency between network and nodal attributes has primarily focused on their relationship explained only by network model under the boundary of model assumption (Wasserman and Pattison, 1996; Howard et al., 2016; Christakis and Fowler, 2007, 2008). A fundamental difficulty of model-based independence tests comes from the fact that not all networks follow the structures described by known network models. Fosdick and Hoff (2015) overcomes this issue by estimating network factors which are believed to embody each node's locations in network space. These factors are in the end used to test independence between network topology and nodal attributes by implementing standard statistical testing method. Through allowing us to choose the dimension of latent factors, they make up the constraints of parametric modeling. However their statistical modeling on networks still rely on the assumption that all the nodes in network would follow the same pattern of dependence – subject to additive and multiplicative effect. In real network data, on the other hand, **it is very common that not all nodes exhibit the same pattern of dependence on their network relationship and also the amount of reliance on network may differ among them.**

Collaborative networks and organization types

- Business
- NGO
- Party
- State
- Science
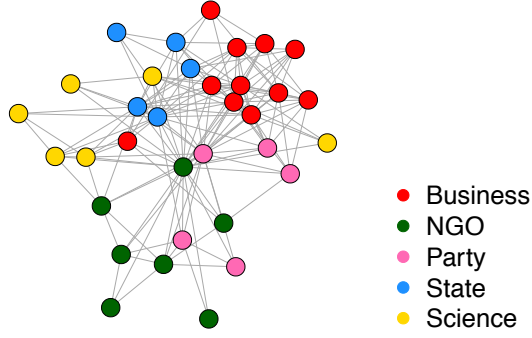
Figure 1: You may conjecture that organizations with the same type are more likely to collaborate each other at first glance; but there has been a lack of statistical method to test if there exists any significant relationship between network topology and node-specific attributes and if any, which node exerts the most dependency on network.

Different from a random vector, network or equivalently graph involves a particular construction which we should take into account. Throughout this paper, we assume that we are given an unweighted and undirected, connected network, equivalently a graph $\mathbf{G}$ without self-loop, comprised of $n(\in \mathbb{N})$ nodes. An adjacency matrix of this given network, denoted by $A = \{A_{ij} : i, j = 1, .., n\}$, is to formalize the relational data of network, where $A_{ij} = 1$ if node i and node j are adjacent each other and zero otherwise. Let us define a m-variate ($m \in \mathbb{N}$) random variable associated with each node, i.e. nodal attributes, $\mathbf{X} \in \mathbb{R}^m$ which we are interested in. Investigating correlation between $\mathbf{G}$ and $\mathbf{X}$, i.e. testing whether their distributions are independent or not is a key focus in our study. An observed network $\mathbf{G}$ can represent social network of students within a school and $\mathbf{X}$ is student-specific grades or heights, for example; or $\mathbf{G}$ can be a collaboration network and $\mathbf{X}$ denotes some properties of participants under collaboration as in Figure 1. We aim to detecting significant association between network topology configured in $\mathbf{G}$ and nodal attributes $\mathbf{X}$ whatever the true nature of network is and however the network is correlated with $\mathbf{X}$. We first have to consider increasing amount of information inherent in network data as the number of nodes increases, which might lead to diverse pattern in dependency as well. In addition, by its definition, an adjacency matrix A involves dependency among its columns and row, so thus we may not utilize an adjacency matrix directly when required independent observations.

3

To overcome these challenges, we propose applying distance-based statistic called multiscale generalized correlation (MGC) (Shen et al., 2016) into testing network indepdnnece which deals with correlation between high dimensional data sets possibly having nonlinear dependence. However we still confront with the difficulty in finding the valid *i.i.d* node-specific coordinates of which Euclidean distance will be an ingredient for the statistics. We claim that graph geometries derived from random walk on graph furnish robust network metrics against to increased variability in edge distributions as well as they satisfy theoretical constraints.

In this paper, beyond the traditional setting of the data in independence test, e.g. between two random vectors, we suggest the method to investigate unique but also ubiquitous data of network and its nodal attributes by utilizing the method testing two random vectors. By suggesting using a family of network metrics which are equipped with favored theoretical and practical properties, we are able to better understand relative locations of each node over network. We are going to elaborate the statistic and demonstrate its validity under some network metrics in Section 2. In Section 3, simulation results demonstrate the best performance of our method compared to the existing under various circumstances. Real data example in section 4 show one of the applications among many.

2 Methodology

2.1 Multiscale Generalized Correlation

Relationship between network and nodal attributes often exhibits local or nonlinear properties having high dimensionality. For example, it is likely that attributes of some subjects may be affected by their network relationship but those of the others are not; or that attributes are correlated on network only when network relationships are too strong. Székely et al. (2007) extended pairwise constructed generalized correlation coefficient and developed distance correlation (dCorr) statistics which guarantee consistency against all types of de-

pendence even under high dimensional data sets. The test setting starts from given $n \in \mathbb{N}$ pairs of $i.i.d$ random samples $(\mathbf{W}, \mathbf{Y}) = \{(\mathbf{w}_i, \mathbf{y}_i) : \mathbf{w}_i \in \mathbb{R}^q, \mathbf{y}_i \in \mathbb{R}^m, i = 1, ..., n\}$. Define $C_{ij} = \parallel \mathbf{w}_i - \mathbf{w}_j \parallel$ and $D_{ij} = \parallel \mathbf{y}_i - \mathbf{y}_j \parallel$ for $i, j = 1, 2, \ldots, n$, where $\parallel \cdot \parallel$ is an Euclidean distance. Distance correlation (`dCorr`) is defined via distance covariance (`dCov`) $\mathcal{V}_n^2$ of $\mathbf{W}$ and $\mathbf{Y}$, which is the following:

$$\mathcal{V}_n^2(\mathbf{W}, \mathbf{Y}) = \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{C}_{ij} \tilde{D}_{ij}. \tag{1}$$

Here $\tilde{C}$ and $\tilde{D}$ is doubly-centered C and D by its column mean and row mean respectively. Distance correlation $\mathcal{R}_n^2(\mathbf{W}, \mathbf{Y})$ is a standardized `dCov` scaled by $\mathcal{V}_n^2(\mathbf{W}, \mathbf{W})$ and $\mathcal{V}_n^2(\mathbf{Y}, \mathbf{Y})$. In addition, a modified distance covariance (`mCov`) $\mathcal{V}_n^*$ and a modified distance correlation (`mCorr`) $\mathcal{R}_n^*$ especially for testing high dimensional random vectors were also proposed in Székely and Rizzo (2013). Unfortunately, these correlation measures with nice closed form often fail to capture nonlinear associations especially embedded in high-dimensional data set (Shen et al., 2016; Heller et al., 2012). Out of this concern, Shen et al. (2016) proposed Multiscale Generalized Correlation (`MGC`) via computing all pairs of local correlation coefficients, excluding distant pairs beyond the certain neighborhood. To be specific, the following set of correlation covariances are calculated:

$$\mathcal{V}_n^{*2}(\mathbf{W}, \mathbf{Y})_{kl} = \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{C}_{ij} \tilde{D}_{ij} I\big(r(C_{ij}) \leq k\big) I\big(r(D_{ij}) \leq l\big) \quad k, l = 1, 2, ..., n, \tag{2}$$

where $r(C_{ij})$ $(r(D_{ij}))$ denotes a rank of $\mathbf{w}_i$ $(\mathbf{y}_i)$ relative to $\mathbf{w}_j$ $(\mathbf{y}_j)$, i.e. $r(C_{ij}) = k$ means w_i is w_j's k-nearest friend in terms of Euclidean distance. Based on this set of statistics, MGC finds the best statistic which exhibits the largest correlation between the two data sets. It has already been shown that this local scaled statistic performs no worse than `dCorr` and that it results improved sensitivity to nonlinear dependence than the distance correlation. We will demonstrate this in Section 3.
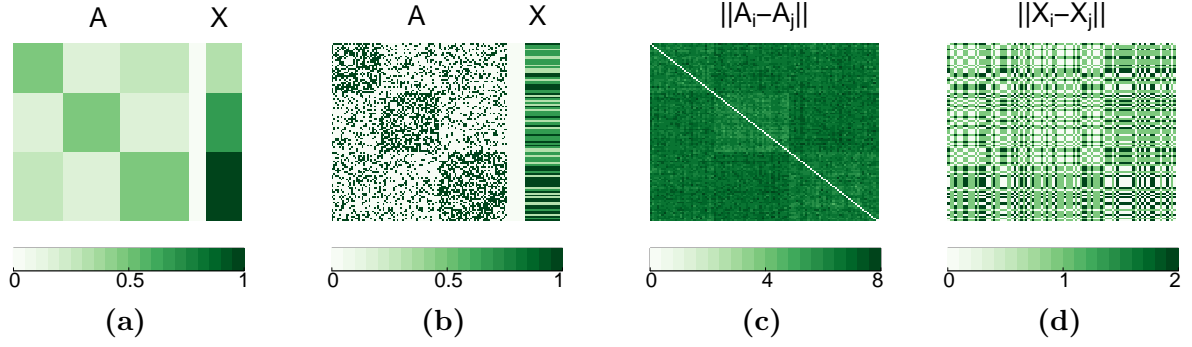
Figure 2: Assume that a set of edges follow certain stochastic block model, also depending on the distribution function of nodal attributes X (a), then with some amount of noise we have a realized adjacency matrix and a set of attribute outcomes (b) of which Euclidean distances ((c) & (d)) are suggested to be used in standard distance-based independence test but neither of them manifests block structures evident in the data generating model.

When we are given multivariate, real valued nodal attributes $\mathbf{X}$, its Euclidean distance is easily constructed. However which random vector we should use as a configuration of network $\mathbf{G}$ so what would be the network-base distance are still remaining questions. To be specific, we are required to find *i.i.d* node-specific coordinates of which Euclidean distance reflects a network-based distance between nodes (We are not always required Euclidean metric (Lyons et al., 2013) but discussion on this is out of scope for this paper). You might first propose directly using a column of an adjacency matrix since an adjacency matrix inherits every edge distribution with certain amount of noise. Then we have a n-pair of observations $\left\{\left(\mathbf{A}_{i\cdot}, \mathbf{X}_i\right) : \mathbf{A}_{i\cdot} = (A_{i1}, ..., A_{in}) \in \mathbb{R}^n, \mathbf{X}_i \in \mathbb{R}^m, i = 1, ..., n\right\}$ as Figure 2. In fact, the main downside of using adjacency matrix is that $\{\mathbf{A_{i\cdot}}\}$ cannot enjoy independent samples in an undirected graph. Even if it is in directed graph, there is a lack of reasonable interpretation behind using Euclidean distance of A as a distance matrix in `MGC` statistics.

2.2   Exchangeable Graph

In order to guarantee the requirement of being *i.i.d* edge distribution, we are going to restrict applicable network to a family of graphs which are *exchangeable*. A graph $\mathbf{G}$ is called exchangeable if and only if its adjacency matrix $\mathbf{A}$ is jointly exchangeable (Orbanz and Roy, 2015).

Definition 2.1 (2-array exchangeability). A random 2-array (A_{ij}) is called jointly exchangeable if

$$(A_{ij}) \stackrel{d}{=} (A_{\sigma(i)\sigma(j)})$$

for every permutation σ of n.

Even though exchangeability itself cannot guarantee being *i.i.d*, thanks to the celebrated *de Finetti* A.1's representation theorem, it has been shown that a sequence of exchangeable variables are *i.i.d* conditioned on its underlying distribution *if and only if* the sequence is exchangeable (Orbanz and Roy, 2015; Caron and Fox, 2014). *Aldous-Hoover theorem* A.2 is the representation theorem of 2-array exchangeable array, which is useful to explain jointly exchangeable adjacent matrix. Exchangeable graph is commonly called *graphon* (Lovász and Szegedy, 2006), which is defined through a random measurable functions (Chan et al., 2013).

Definition 2.2 (graphon). A *graphon* with $n (\in \mathbb{N})$ nodes is defined as a function of a symmetric measurable function $g : [0,1]^2 \to [0,1]$ with input of $u_i \stackrel{i.i.d}{\sim} Uniform[0,1], i = 1, 2, ..., n$. Let A be an adjacency matrix of graphon. Then for any $i < j, \quad i, j = 1, 2, ..., n$:

$$Pr(A_{ij} = 1 | u_i, u_j) = g(u_i, u_j) \tag{3}$$

By *Aldous-Hoover theorem*, we obtain a clear representation of exchangeable network through measurable function g for a half of the edge set under an undirected network where $A_{ij} = A_{ji} \ (i, j = 1, 2, ..., n)$.

$$(A_{ij}) = (A_{\sigma(i)\sigma(j)}) \Longleftrightarrow A_{ij} \stackrel{i.i.d}{\sim} Bern(E_U[g(u_i, u_j)]), \quad i < j \tag{4}$$

Networks based on widely used graphical model are exchangeable. One of the most popular models is Stochastic Block Model (SBM) (Holland et al., 1983) where many of network methodologies have been built on. In the simplest setting of SBM, we assume that each n nodes of **G** belongs to one of $K \in \mathbb{N}(\leq n)$ blocks or groups. Block affiliation is important in that the probability of having edges between a pair of nodes depends on which blocks they are in. Let a latent variable Z indicate each node's block affiliation such that $Z_1, Z_2, ..., Z_n \stackrel{i.i.d.}{\sim}$

$Multinomial\big(\pi_1, \pi_2, ..., \pi_K\big)$. Then the upper triangular entries of A are independent and identically distributed conditional on $\{\mathbf{Z}\}$:

$$A_{ij}\big|Z_i, Z_j \overset{i.i.d.}{\sim} Bern\Big(\sum_{k,l=1}^{K} p_{kl}I\big(Z_i = k, Z_j = l\big)\Big), \forall i < j. \tag{5}$$

The above distribution can also be represented through some random function g : $[0,1]^2 \to [0,1]$, e.g.$g\big(W_i, W_j\big) = \sum_{k,l=1}^{K} p_{kl}I\left(W_i \in \big[\sum_{j=0}^{k-1}\pi_j, \sum_{j=0}^{k}\pi_j\big], W_j \in \big[\sum_{j=0}^{l-1}\pi_j, \sum_{j=0}^{l}\pi_j\big]\right)$ for $W_1, W_2, ..., W_n \overset{i.i.d.}{\sim} Unif[0,1]$, where $\pi_0 = 0$ and $\sum_{j=0}^{K}\pi_j = 1$. Then we can have conditional $i.i.d$ edge distribution given g, restrictive to upper triangular part of A.

$$A_{ij}\big|g, W_i, W_j \overset{ind}{\sim} Bern\big(g(W_i, W_j)\big), \forall i < j$$
$$A_{ij}\big|g \overset{i.i.d}{\sim} \int\int Bern\big(g(W_i, W_j)\big)f_W(w_i)f_W(w_j)dwdw, \quad \forall i < j. \tag{6}$$

Even though this is not the only representation of edge distribution, for any exchangeable graphs, including SBM and also Random Dot Product Graph (RDPG) (Young and Scheinerman, 2007), there must exist a random function g which edges are independent identically distributed conditioning on.

Despite its advantage on simple representation, graphon is either dense, i.e. $|E| = o(|V|^2)$ (Veitch and Roy, 2015), or empty. This implies a graphon often fails to represent real network data where sparsity or scale-free distribution is fairly common. Thus, in addition to graphon, we introduce a concept of *graphex*, first proposed by Veitch and Roy (2015), which is more generalized version of graphon and also includes sparse exchangeable graphs (Caron and Fox, 2014). Caron and Fox (2014) suggested formalizing a network as point process over $\mathbb{R}_+^2$ on the basis of *Kallengerg Representation Theorem* (Kallenberg, 1990). As we were able to conditionally represent $\{A_{ij}\}$ through a random transformation of $i.i.d$ uniform variables, jointly exchangeable point processing network also can be formalized via a random function of $i.i.d$ unit rate Poisson process and of $i.i.d$ uniform variables. To be specific, an undirected graph on a point process on $\mathbb{R}_+^2$ can be thought of having an adjacency matrix defined on

an edge set $\{\theta_i : \theta_i \overset{i.i.d}{\sim} Poisson(1), i = 1, 2, \ldots\}$ (Kallenberg, 1990):

$$A_{\theta_i \theta_j} \overset{i.i.d}{\sim} g(\vartheta_i, \vartheta_j), \quad i < j, \tag{7}$$

where $g : \mathbb{R}_+^2 \to [0, 1]$ is a random function defined on unit-rate Poisson processing $\{\vartheta\}$. In a graphex, joint exchangeability applied to a point processed nodal label θ, not on a node label itself. Despite its more intricate form, representation of sparse graph as exchangeable formation helps us to demonstrate the validity of our proposed methods in real network data.

Definition 2.3 (Joint exchangeability on point process)**.** Let $h > 0$ and $V_i = [h(i - 1), hi]$ for $i \in \mathbb{N}$ then

$$\big(A(V_i \times V_j)\big) \overset{d}{=} \big(A(V_{\sigma(i)} \times V_{\sigma(j)})\big) \tag{8}$$

for any permutation σ of $\mathbb{N}$.

## 2.3 Family of Network Distances and Test Statistics

Then which network-specific distance should be applied for an exchangeable graph? Coifman and Lafon (2006) proposed multiscale geometries of data called *diffusion maps*, which inherit every local relation between the nodes when applied in network. The geometries in network are constructed via running random walk on graph by iterating transition matrix of Markov process. Roughly speaking, we go on the journey starting from each node and at each iteration we can cross over one edge. At each iterating time, called *diffusion time*, we calculates the chance to stay between node i and node j considering all possible paths between them, which is assumed to be proportional to the distance. For example, at diffusion time t, we travel from each node to others at most t geodesic distance apart. Corresponding location or coordinates of each node at every iteration can be derived as a family of *diffusion maps*. Diffusion maps are constructed by eigenvectors of Markov matrix of which Euclidean distance is called a diffusion distance. For each diffusion time $t \in \mathbb{N}$, we can define a diffusion distance C_t, using a discrete set of real nonzero eigenvalues $\{\lambda_r\}$ and eigenvectors $\{\phi_r\}$ of a

transition matrix (Coifman and Lafon, 2006; Lafon and Lee, 2006).

$$C_t^2[i,j] := \parallel \mathbf{U}_t(i) - \mathbf{U}_t(j) \parallel \quad i,j = 1, 2, \ldots, n \tag{9}$$

where $\mathbf{U}_t(i) = \left(\lambda_1^t \phi_1(i) \quad \lambda_2^t \phi_2(i) \quad \ldots \quad \lambda_q^t \phi_q(i) \right)^T \in \mathbb{R}^q$ is a diffusion map at time t. As diffusion time t increases, distance matrix C_t is more likely to take into account distance between two nodes which are relatively difficult to reach each other. Figure 3 shows three exemplary distance matrices among whole one-parameter family of distance $\{C_t : t \in \mathbb{N}\}$.

There are two main merits in using a set of diffusion distances under exchangeable graphs: First of all, compared to an adjacency relationship or geodesic distance which are two extremes, diffusion distance well reflects the connectivity since it takes into account every possible path between the two nodes. Second, each diffusion map provides $i.i.d$ multivariate coordinates of nodes so that using diffusion distance in distance-based test statistics is a valid method. The following Lemmas show how $\mathbf{U}_t$ retains (conditional) independence privileged to exchangeable graph.
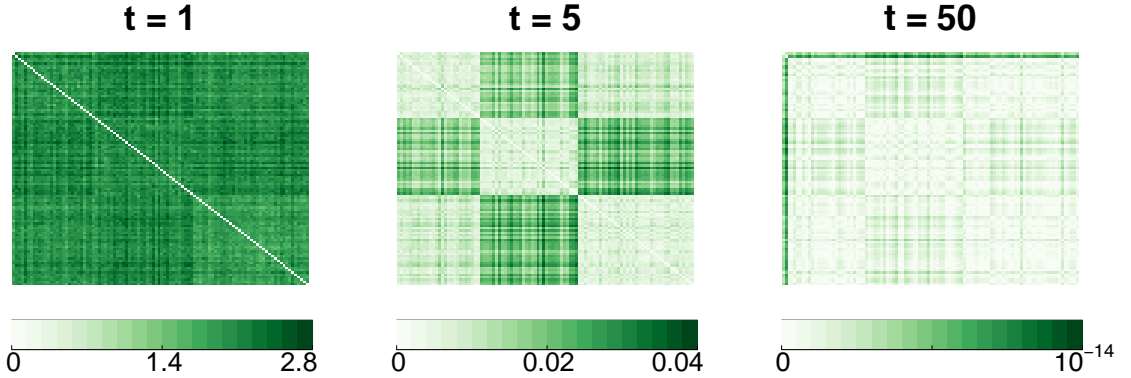
Figure 3: *Diffusion matrix*, as a proposed alternative for Euclidean distance of A, provides **one-parameter family of network-based distances** where at early stage, e.g. at $t = 1$, distance matrix is very similar to Euclidean distance of A but as time goes by the pattern shown in the distance matrix changes, and **at optimal time point $t^* = 5$ distance matrix shows most clear block structures and at the same time it exhibits most dependence to distance matrix of X.**

Lemma 2.1 (Exchangeability and $i.i.d$ of A in graphon). Assume that a connected, undirected and unweighted graph $\mathbf{G}$ is a graphon. Then 2-array of $\{A_{ij} : i = 1, 2, ..., n, i < j\}$

are $i.i.d$ conditioning on some random link function $g : [0,1]^2 \rightarrow [0,1]$. Thus for fixed row (column) of A, $\{A_{i1}, A_{i2}, ..., A_{in}\} \setminus \{A_{ii}\}$, $i \in \{1, 2, ..., n\}$ are conditionally $i.i.d$ given a random link function g or equivalently, its underlying distribution.

Lemma 2.2 (Exchangeability and $i.i.d$ of $\mathbf{U}_t$). Assume that a connected, undirected and unweighted graph $\mathbf{G}$ is a graphon. Then its transition probability so thus diffusion maps at fixed time t also exchangeable, conditional on a link function of graph. Furthermore, by *de Finetti's Theorem* A.1, such diffusion maps at t are conditionally $i.i.d$ given its underlying distribution.

Lemma 2.2 above provides us with $i.i.d$ one-parameter family of $\{\mathbf{U}_t\}_{t \in \mathbb{N}}$ conditional on its underlying distribution. Unfortunately this is a story only applied to an exchangeable graph, which cannot be sparse. In sparse version of exchangeable graphs, i.e. graphex, one more step of conditioning on point process θ is needed.

Lemma 2.3 (Exchangeability and $i.i.d$ of A in graphex). Assume that a connected, undirected and unweighted graph $\mathbf{G}$ is a graphex. Then 2-array of $\{A_{ij} : i = 1, 2, ..., n, i < j\}$ are $i.i.d$ conditioning on some random link function $g : [0,1]^2 \rightarrow [0,1]$ and unit-Poisson process θ. Thus for fixed row (column) of $\mathbf{A}$, $\{A_{i1}, A_{i2}, ..., A_{in}\} \setminus \{A_{ii}\}$, $i \in \{1, 2, ..., n\}$ are conditionally $i.i.d$ on its underlying distribution, specifically conditioning on random link function g and θ.

Similar to Lemma 2.2, we are able to show exchangeability of a transition matrix in graphex case, which leads to conditional $i.i.d$ of its diffusion maps. If exchangeable diffusion maps are applicable to distance-based test statistics, e.g. MGC, we are then able to dispense with obstacles in testing network independence. Assume that we have a finite sample of infinitely exchangeable sequence $(\mathbf{W}, \mathbf{Y}) = \{(\mathbf{w}_i, \mathbf{y}_i) : i = 1, 2, \ldots, n\}$, which is identically distributed as $(\mathbf{w}, \mathbf{y})$ with finite second moment. Then, by the properties of exchangeable

sequences, there exists some random measure θ and its probability measure P such that :

$$w_i|\theta \overset{i.i.d}{\sim} \prod_{i=1}^{n} f_{\mathbf{w}_i|\theta}(\mathbf{w}_i|\theta)$$

$$\int \prod_{i=1}^{n} f_{\mathbf{w}_i|\theta}(\mathbf{w}_i|\theta)P(d\theta) \overset{d}{=} \prod_{i=1}^{n} f_{\mathbf{w}}(\mathbf{w}_i) \tag{10}$$

$$\mathbf{w}_i \overset{i.i.d}{\sim} f_{\mathbf{w}}(\mathbf{w}) \quad \text{conditioning on the underlying distribution } f_{\mathbf{w}},$$

where f is a random, marginal distribution integrated over θ; same arguments hold for exchangeable $\{\mathbf{y}_i\}$. The following two Lemmas serve as the foundation for using exchangeable observations in MGC.

Lemma 2.4. Under the above conditions, we have

$$\mathcal{V}_n^2(\mathbf{W}, \mathbf{Y}) = \|g_{\mathbf{w},\mathbf{y}}^n(t,s) - g_{\mathbf{w}}^n(t)g_{\mathbf{y}}^n(s)\|^2,$$

where g^n is the empirical characteristic function based upon $\{(\mathbf{w}_i, \mathbf{y}_i), i = 1, 2, ..., n\}$

Lemma 2.5. We have

$$\mathcal{V}_n^2(\mathbf{W}, \mathbf{Y}) \longrightarrow \mathcal{V}^2(\mathbf{w}, \mathbf{y}) \qquad \text{as } n \to \infty \tag{11}$$

where $\mathcal{V}^2(\mathbf{w}, \mathbf{y}) := \|g_{\mathbf{w},\mathbf{y}}(t,s) - g_{\mathbf{w}}(t)g_{\mathbf{y}}(s)\|^2$, and $g.$ is a characteristic function, e.g., $g_{\mathbf{w},\mathbf{y}}(t,s) = E\{\exp\{i\langle t, \mathbf{w}\rangle + i\langle s, \mathbf{y}\rangle\}\}$. It follows that

$$\mathcal{V}_n^2(\mathbf{W}, \mathbf{Y}) \to 0 \quad \text{as } n \to \infty \tag{12}$$

if and only if $g_{\mathbf{w},\mathbf{y}}(t,s) = g_{\mathbf{w}}(t)g_{\mathbf{y}}(s)$, i.e., $\mathbf{w}$ is independent of $\mathbf{y}$.

Lemma 2.4 and its following Lemma 2.5 facilitate the use of distance correlation while satisfying *Theorem 2* in Székely et al. (2007).

Theorem 2.6. Suppose that we are given n pairs of exchangeable observations $(\mathbf{W}, \mathbf{Y}) = \{(\mathbf{w}_i, \mathbf{y}_i); i = 1, 2, \ldots, n\}$ having finite second moment. Assume $\mathbf{w}_i \overset{i.i.d}{\sim} f_{\mathbf{w}}$ and $\mathbf{y}_i \overset{i.i.d}{\sim} f_{\mathbf{y}}$

given underlying distribution for $i = 1, 2, \ldots, n$. Then

$$\mathcal{V}_n^2(\mathbf{W}, \mathbf{Y}) \longrightarrow 0 \quad \text{as } n \to \infty \tag{13}$$

if and only if $\mathbf{w}$ is independent of $\mathbf{y}$. Moreover, `dCorr` and `MGC` are consistent for testing dependence between $\mathbf{w}$ and $\mathbf{y}$, i.e., the testing power converges to 1 asymptotically for any dependency of finite second moment.

Note that if $\{\mathbf{w}_i : i = 1, 2, \ldots, n\}$ are *i.i.d*, they are exchangeable. Thus estimated latent factors, which are assumed *i.i.d* by Fosdick and Hoff (2015) can also be applied to Theorem 2.6. We already have shown that even under undirected network, diffusion maps remain exchangeable at each diffusion time point t.

Theorem 2.7. Then `MGC` is consistent in testing network independence through testing $H_0 : f_{\mathbf{U}^{(t)} . \mathbf{X}} = f_{\mathbf{U}^{(t)}} \cdot f_{\mathbf{X}}$. In particular, the consistency also holds for the estimated latent positions and adjacency matrix of directed network instead of diffusion maps.

Moreover, we are not going to state otherwise but if you assume to be given a unit-Poisson process $\{\theta_i\}_{i=1}^n$, you can lead to the same results for sparse graphex as Theorem 2.6.

2.4 Measure for Node Contribution

On the other hand, in the presence of nonlinear dependency, some nodes often exert more reliance on their attributes than the others. Like other node-specific measure of importance, e.g. centrality, the amount of each node's leverage on dependence can be of interest. Here we suggest the measure of node's contribution to detecting dependence by utilizing `MGC` statistic. Let (k^*, l^*) be the optimal neighborhood choice in distance matrix (C, D) respectively. Denote the contribution of node $v \in V(G)$ to the testing statistic by $c(\cdot) : v \to \mathbb{R}$

$$c(v) \propto \sum_{j=1}^n \tilde{C}_{jv} \tilde{D}_{jv} I\big(r(C_{jv}) \leq k^*\big) I\big(r(D_{jv}) \leq l^*\big), \tag{14}$$

which is proportional to v^{th} column-sum of the pre-summed test statistic 2. Note that the deviation of non-negative `MGC` statistic from zero implies departure from the independence and also note that we truncate the `dCov` statistics by column entry's rank. Thus $\tilde{C}_{jv}\tilde{D}_{jv}$ would not be truncated if node j ($\in \{1, 2, \ldots, n\} \setminus v$) is important to node v and its larger, positive value would contribute to $\mathcal{V}_n^{*2}$ more. The statistic $c(v)$ comes out from these observations. We illustrate how this contribution measure works in the next section 3.4.

3 Simulation Study

In the simulation studies, we make a comparison between empirical testing power across various multivariate independence test statistics: `MGC`, `dCorr(mCorr)`, Heller-Heller-Gorfine (`HHG`) (Heller et al., 2012), and likelihood ratio test of Fosdick and Hoff (`FH`). For computing statistical power, we used type I error $\alpha = 0.05$ and obtain p-values of each sample network via permutations. For fair comparison between these testing methods, we also present an additive and multiplicative effect model which `FH` mostly watch for. All the simulation models are illustrated by joint distribution of adjacent matrix $\mathbf{A}$, nodal attributes $\mathbf{X}$, and latent variable $\mathbf{Z}$, which explains dependence structure between $\mathbf{A}$ and $\mathbf{X}$.

$$
\begin{aligned}
f(\mathbf{A}, \mathbf{X}, \mathbf{Z}) &= f_{A|Z}(\mathbf{A}|\mathbf{Z}) \cdot f_{Z|X}(\mathbf{Z}|\mathbf{X}) \cdot f_X(\mathbf{X}) \\
&= f_{A|Z}(\mathbf{A}|\mathbf{Z}) \cdot f_{X|Z}(\mathbf{X}|\mathbf{Z}) \cdot f_Z(\mathbf{Z})
\end{aligned}
\tag{15}
$$

According to the joint model 15, distribution of edges and nodal attributes are correlated only through a node-specific latent variable $\mathbf{Z}$ no matter whether $\mathbf{X}$ is modeled via $\mathbf{Z}$ or vice versa. For each simulated network, empirical power will be derived by comparing observed statistic to the empirical distribution under the null.

3.1 Stochastic block model

We already mentioned that Stochastic Block Model (SBM) is one of the most popular and also useful network generative model. We present the SBM with $K = 3$ blocks (model 19)
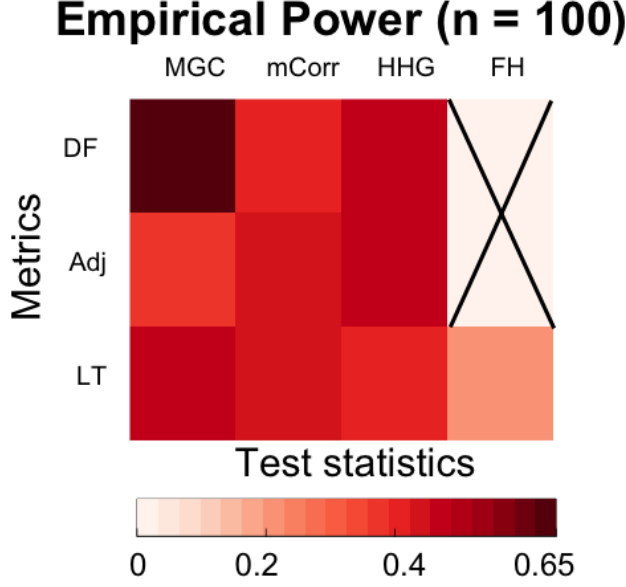
Figure 4: This power heatmap illustrates the superior power of multiscale generalized correlation (`MGC`) under diffusion distance matrix (`DF`) in three SBM (model 19), compared to under adjacency matrix distance (`Adj`) or latent factor distance (`LT`). **This demonstrates one exemplary network where `MGC` statistic along with a family of diffusion distances catches non monotonic correlations efficiently than the other statistics and metrics.**

where block affiliation for each node is correlated with its attributes X. Figure 4 illustrates superior performance of `MGC` as a test statistic combined with diffusion maps (`DF`) as a network metric. To simply represent model 19,

$$E(A_{ij}|X_i, X_j) = 0.5I(|X_i - X_j| = 0) + 0.2I(|X_i - X_j| = 1) + 0.3I(|X_i - X_j| = 2). \quad (16)$$

When $X_i = X_j$, these two nodes are most likely to have an edge but when X_i and X_j differ by one, they are even less likely to have an edge, with probability of 0.2, than the most different pairs of nodes. This actually describes **nonlinear dependence where `MGC` is believed to work better than the distance correlation**. To deep into studying performance of local optimal scaled `MGC` in the presence of both linear dependency and non-linear dependency, we control these two phases through changing a value of $\theta \in (0, 1)$. When $\theta > 0.2$, linear dependency of edge distribution upon nodal attribute X is lost.

$$E(A_{ij}|X_i, X_j) = 0.5I(|X_i - X_j| = 0) + 0.2I(|X_i - X_j| = 1) + \theta I(|X_i - X_j| = 2) \quad (17)$$

If you see Figure 5, power of `dCorr` starts to drop from $\theta = 0.2$ while that of `MGC` almost stays clam.This implies `MGC` keeps its sensitivity even under nonlinear dependency compared

15
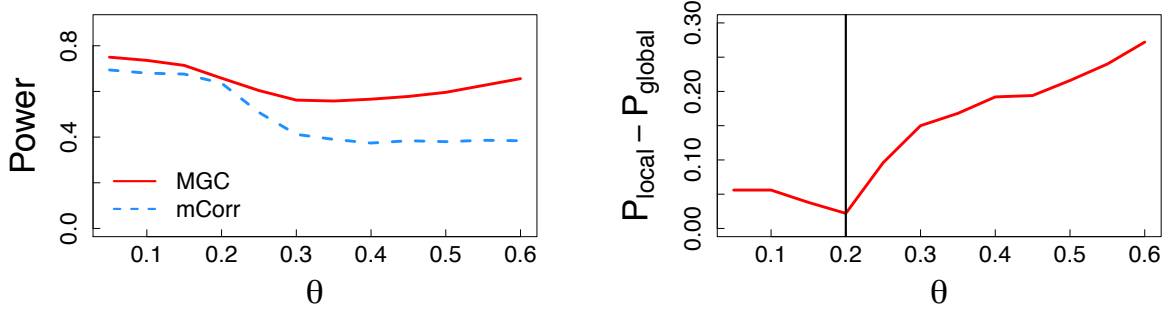
Figure 5: X-axis of θ controls the existence/amount of nonlinear dependency and in this particular case nonlinearity exists when $\theta > 0.2$ and gets larger as it increases. You can see the discrepancy in power between global and local scale tests also gets larger accordingly, **mostly due to decreasing power of global test but relatively stable power of `MGC` under nonlinear dependency** as presented in the left panel.

to `dCorr`.

3.2 Degree-corrected two block model

The SBM connotes that all nodes within the same block have the same expected degree. However the network model of homogeneous distribution within block leads to a poor fit to networks with highly varying node degrees even within block. Out of this concern, the Degree-Corrected Stochastic Block model (DCSBM) varies distribution of node degree within a block, preserving the overall block community structure (Karrer and Newman, 2011). Consider two block SBM having more variability induced by V (model 20). Here we have an edge distribution as a linear function of Euclidean distance of X. However its variance becomes inflated due to $\tau > 0$. Relatively poor performance of an adjacency metric as presented in Figure **??** can be attributed to large variability in it. Even in this case, `MGC` achieves almost same level of power across increasing τs. On the other hand, node-specific latent factors look more sensitive in `FH` model, but still diffusion maps metrics work better.

16
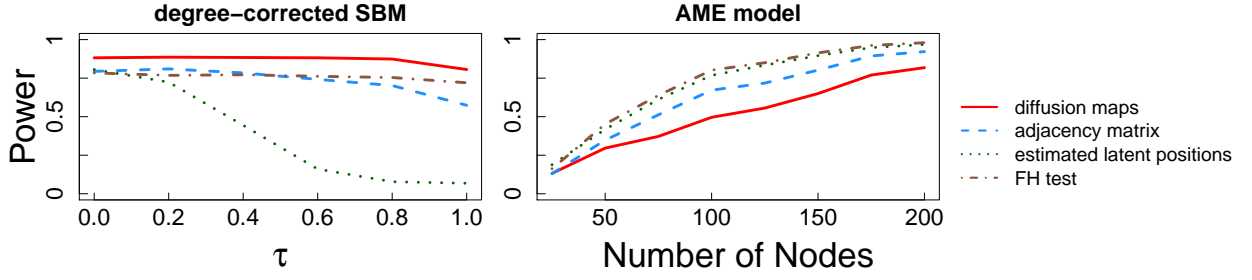
Figure 6: (a) In degree-corrected SBM where the variability in degree distribution increases as τ increases, testing power of diffusion maps are more likely to be robust against increasing variability compared to other network metrics, e.g. adjacency matrix or latent positions. `FH` test statistics allowing different dimensions of network factors perform consistently well but still have less power than `MGC`. (b) `MGC` utilizing diffusion distances loses some power under additive and multiplicative model which favors estimated latent position metrics, but `MGC` does as good as `FH` tests under latent factor metrics which closes to the truth. This reveals the flexibility in distance-based matrix in `MGC` statistics, which can be chosen depending on model fit or preliminary knowledge.

3.3   Additive and multiplicative graph model

Hoff et al. (2002) proposed an approach of modeling network as an additive and multiplicative function of node-specific latent factors. Whereas Fosdick and Hoff (2015) embedded nodes into latent factors assuming that their network model is *correct*, a family of diffusion maps are nonparametric version of embedding nodes into multivariate variable without losing any information on adjacent relationship. Thus in the model 21, where logit of A obeys the presumed model, the estimated latent factors would be very close to the truth, much closer than embedding made from diffusion maps. Even though we rarely see the network nearly follows the model in reality, If we should, using network factors as independent observations from graph $\mathbf{G}$ and applying them to `MGC` performs not very worse than `FH` statistic (Figure **??**). In other words, if network really fits well to the network model with node-specific latent factors as covariates, then it would safe to use those factors in the `MGC` statistic directly. Since they assume $i.i.d$ generative model for factors, there is nothing wrong with applying `MGC` using $i.i.d$ observations of estimated factors.

17

3.4 Node Contribution Test
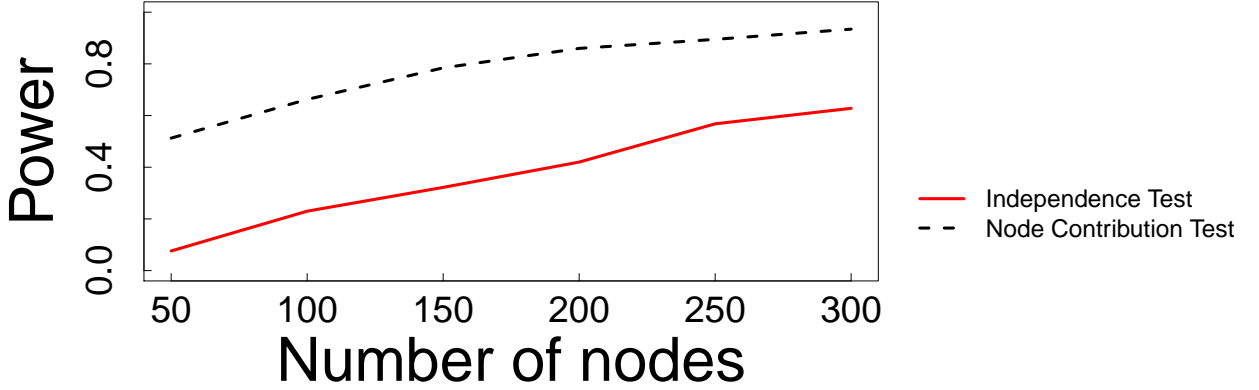
Figure 7: This plot describes that both power of `MGC` and the rate of correctly-ranked node contribution increase as the number of nodes increases when only half of the nodes for each simulation actually are set to be dependent on network, **which validates the use of node contribution measure in independence test.**

To examine the effectiveness of node contribution measure in testing dependency as presented in the statistic 14, we deliberately simulate the network and its nodal attributes as half of the nodes are independent while the other half are dependent on network (model 22). As an ad hoc test of node contribution, we rank the nodes in terms of decreasing order of $c(v)$ and count the ratio of dependent samples's ranks within the number of dependent nodes. If it works perfectly, all dependent nodes would take higher rank than every independent node so thus the rate equals to one. We call this rate as *inclusion rate*:

$$\text{inclustion rate}\big(c(v)\big) = \sum_{v \in V(\mathbf{G})} \big\{ rank_{c(v)}\big(v\big) \leq m \big\}/m, \tag{18}$$

where $m(\leq |V(\mathbf{G})|)$ is the number of nodes under network dependence. We set $m = n/2$ out of $n = |V(\mathbf{G})|$.

4 Real Data Examples

Collaborative networks and organization types
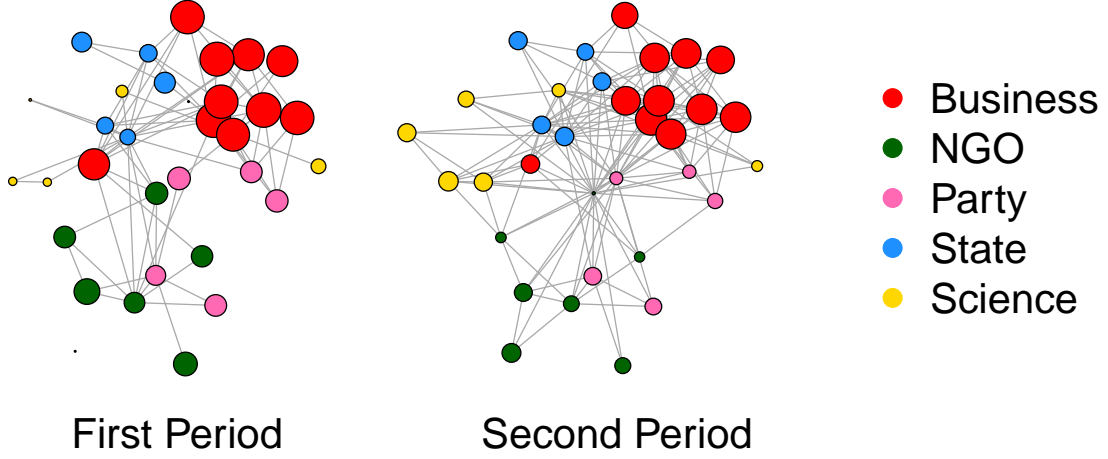
Figure 8: Both panels depict the collaborative networks during the two time periods having significant network dependency in types of organizations. **Using MGC statistics, we are not only able to test network independence but also calculate each node's amount of contribution to detecting dependence, which is proportional to node size here.** You can tell that the tendency to collaborate within the same type is strongest among the business group while scientist relatively collaborates less with any others, especially in the first period.

In the field of political science, who exerts more powerful impacts than others over political network and which factors impact on the power differentials are one of the interests (Ingold and Leifeld, 2014). Minhas et al. (2016) made an inference from political networks (Cranmer et al., 2016) via the additive and multiplicative effects (AME). As mentioned in Section 3.3, the AME model estimates the latent factors and uses them to test independence with the nodal attributes. Among diverse attributes that Cranmer et al. (2016) provided, we focus on the types of organizations and how 34 political organizations having different types are participating policy network. We changed a given directed network into undirected network and use a dissimilarity matrix for distance matrix of the attributes, i.e., $\| \mathbf{X}_i - \mathbf{X}_j \| = 0$ if and only if node i and node j are from the same type and one otherwise. Two collaboration networks comprised of the same set of nodes across two time periods are

provided (Ingold and Leifeld, 2014). Figure 8 and Figure 9 illustrates these two networks and shows each node's reliance on its organization type when collaborating. For two periods, the network independence test statistics of `MGC` and `dCorr` using diffusion distance matrices result in significant p-values across diffusion times from $t = 1$ to $t = 10$. The conclusion from the `FH` test is also the same.
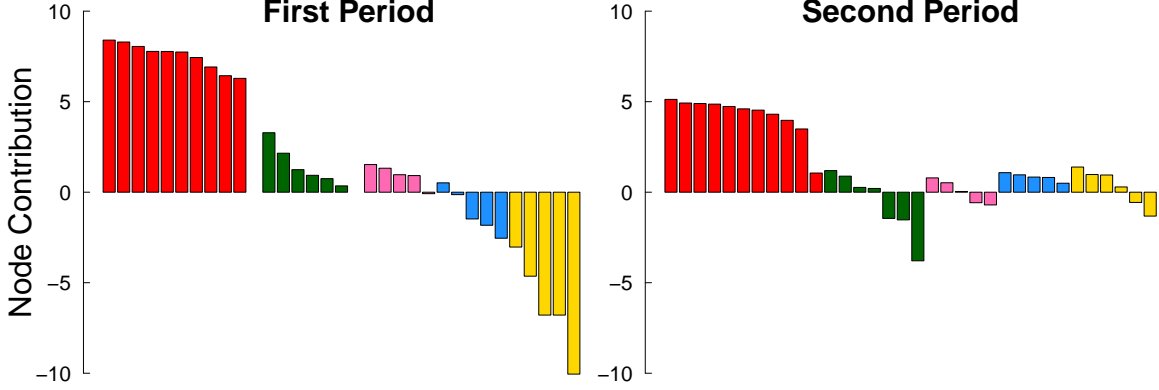
Figure 9: In the first period, we have two extreme cases among the business group and science group, which reflects our observations in Figure 8. Generally organizations cooperate more actively between different types in the second period but still their collaboration network is highly dependent on their organization types.

5 Discussions

In this paper, we have shown that `MGC`, merged with a family of diffusion distance, provides us powerful independence test statistics in network. Having multiscale statistics, i.e. one parameter family of statistics, is not avoidable because **we regard distance between nodes over network as a dynamic process.** Through simulation studies, we demonstrated that our methods work better that the others especially under nonlinear dependency, and we are able to measure each node's contribution to detecting dependence. Deriving the contributions is particularly important in nonlinear network dependence since we assume possibly different amount of the dependencies among the nodes.

However obtaining a full family of statistics are computationally infeasible. Also we did

not suggest any theoretically supported tools to select one metrics among them so thus we have one single statistic. As an ad hoc, we selected an *optimal* diffusion time t with highest power from $t = 1$ to $t = 10$ for our simulation since we could observe a stabilized empirical power within this period. In the real data study we have significant p-values across diffusion time $t = 1$ to $t = 10$. **Developing the adaptive method to find this optimal t when dependence is maximized would be a natural next step.** Despite these shortcomings, a range of applications based on our work are very diverse. Even though we specifically constraint the statistic into testing independence between network and nodal attributes, **we are able to implement independence testing between two networks of same size by using diffusion distance of each network.** This type of test will be useful when we want to investigate whether a pair of networks are topologically or structurally independent. In this possible extension, we could also enjoy both `MGC` and a family of diffusion distances which are specialized in preserving local relationship between the two data sets.

A appendix

A.1 simulation schemes

- **Three Block SBM**

$$
X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Multi(1/3, 1/3, 1/3), \quad i = 1, \ldots, n
$$

$$
Z_i | X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Multi(0.5, 0.25, 0.25)I(x = 1) + Multi(0.25, 0.5, 0.25)I(x = 2)
$$
$$
+ Multi(0.25, 0.25, 0.5)I(x = 3), \quad i = 1, \ldots, n
$$

$$
A_{ij} | Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij} | z_i, z_j) \overset{d}{=} Bern(0.5)I(|z_i - z_j| = 0) + Bern(0.2)I(|z_i - z_j| = 1)
$$
$$
+ Bern(0.3)I(|z_i - z_j| = 2), \quad i, j = 1, \ldots n; \quad i < j.
$$

$$(19)$$

- **Increasing variance in DCSBM**

$$X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Bern(0.5), \quad i = 1, \ldots, n$$

$$Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Bern(0.6)I(x = 0) + Bern(0.4)I(x = 1), \quad i = 1, \ldots, n$$

$$V_i \overset{i.i.d}{\sim} Uniform(1 - \tau, 1 + \tau), \quad i = 1, \ldots, n; \quad \tau = 0, 0.2, \ldots, 1 \qquad (20)$$

$$A_{ij}|Z_i, Z_j, V_i, V_j \overset{i.i.d}{\sim} f_{A|Z,V}(a_{ij}|z_i, z_j, v_i, v_j) \overset{d}{=} Bern(0.2 \cdot v_i v_j)I(|z_i - z_j| = 0)$$

$$+ Bern(0.05 \cdot v_i v_j)I(|z_i - z_j| = 1), \quad i, j = 1, \ldots, n; i < j.$$

- **Additive and Multiplicative Graph Model**

$$Z_i \overset{i.i.d}{\sim} f_Z(z) \overset{d}{=} Uniform[0, 1]. \quad i = 1, \ldots, n$$

$$X_i|Z_i \overset{i.i.d}{\sim} f_{X|Z}(x|z) \overset{d}{=} Normal(Z_i, 1), \quad i = 1, \ldots, n$$

$$A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \overset{d}{=} Bern\big((1 - z_i)^2 \times (1 - z_j)^2\big), \quad i, j = 1, \ldots, n; i < j.$$

$$(21)$$

- **Node Contribution**

$$X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Bern(0.5) \quad i = 1, \ldots, n/2, \ldots, n$$

$$Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Bern(0.6)I(x = 0) + Bern(0.4)I(x = 1), \quad i = 1, \ldots, n/2, \ldots, n$$

$$A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j)$$

$$\overset{d}{=} \begin{cases} Bern(0.4)I(|z_i - z_j| = 0) + Bern(0.1)I(|z_i - z_j| > 0) & i = 1, \ldots, n/2 \\ Bern(0.25) & i = 1 + n/2, \ldots, n \end{cases}$$

$$(22)$$

A.2   Algorithms

Algorithm 1 Mutiscale representation of nodes in network

Input: Transition probability matrix P of network G and a set of time points $\{t_i : t_i \in \mathbb{N}\}$ of diffusion time.
Output: A list of diffusion maps at each time point.

1: **function** DMAP ($n \times n$ transition matrix P, time points $\{t_1, t_2, \ldots, t_K\}$)
2:     $\pi := \texttt{statdistr}(P)$ $\triangleright$ stationary distribution of P
3:     $\Pi := \texttt{Diag}(\pi)$ $\triangleright$ Diagonal matrix with diagonal element of π
4:     $Q := \Pi^{1/2} P \Pi^{-1/2}$
5:     $\lambda := \texttt{eigenvalue}(Q)$ $\triangleright$ a real-valued vector with length of $q (\leq n)$.
6:     $\Lambda := \texttt{Diag}(\lambda)$
7:     $\Psi := \texttt{eigenfunction}(Q)$ $\triangleright$ $n \times q$ real-valued matrix
8:     $\Phi := \Pi^{-1/2} \Psi$ $\triangleright$ $n \times q$ real-valued eigenfunction matrix of P
9:     **for** $t_i : i = 1$ **do** K
10:        $\texttt{Maps}[i] := \Phi \Lambda^{t_i}$
11:     **end for**
12:     $\texttt{Maps} = \text{list}(\texttt{Maps}[1], \texttt{Maps}[2], \ldots, \texttt{Maps}[K])$
       return Maps
13: **end function**

Algorithm 2 Multiscale Generalized Correlation (MGC) test statistics with diffusion maps as a network-based distance.

Input: A connected, undirected network G with its nodal attributes $\mathbf{X}$.
Output: A list of ((a) p-value of sample MGC, (b) estimated sample MGC statistic, (c) p-value map for all local correlations, (d) a set of estimated optimal neighborhood scales $\{(k^*, l^*)\}$) for each diffusion maps.

1: **function** NETWORKTEST (G, $\mathbf{X}$, $\mathbf{T}$:= (diffusion time points $\{t_1, t_2, \ldots, t_K\}$))
2:     $A := \texttt{get.adjacency}(G)$ $\triangleright$ obtain an adjacency matrix of network G
3:     $P := A \ / \ \texttt{rowSums}(A)$
4:     $U := \texttt{dmap}(P, \mathbf{T})$ $\triangleright$ a list of diffusion maps in each time point
5:     **for** $t_i : i = 1$ **do** K
6:        $C := \texttt{dist}(U[i])$ $\triangleright$ distance matrix of diffusion maps at time t_i
7:        $D := \texttt{dist}(X)$ $\triangleright$ distance matrix of nodal attributes
8:        $\texttt{MGC}[i] = \texttt{MGCPermutationTest}(C, D)$
9:     **end for**
10:     $\texttt{MGC} = \text{list}(\texttt{MGC}[1], \texttt{MGC}[2], \ldots, \texttt{MGC}[K])$
       return MGC
11: **end function**

Algorithm 3 Node-specific contribution to detecting dependency via `MGC` statistic

Input: Distance metric of graph G, C, and attributes X, D, and (one of) the estimated optimal scales $\{k^*, l^*\}$

Output: unstandardized contributions of each node in network $\{c(v)\}$

 1: **function** CONTRIBUTION (C, D , $\{(k^*, l^*)\}$)
 2:     $\tilde{C} := \texttt{DoubleCentering}(C)$
 3:     $\tilde{D} := \texttt{DoubleCentering}(D)$
 4:     $\texttt{Rank}(M_{ij}) :=$ (rank of node j with respect to node i)
 5:     **for** $v = 1$ **do** $|V(G)|$ ▷ iterate over every each node
 6:         $c(v) = 0$
 7:         **for** $j = 1$ **do** n
 8:             $c(v) = c(v) + \tilde{C}_{vj} \tilde{D}_{vj} I(\texttt{Rank}(C_{vj}) \leq k^*, \texttt{Rank}(D_{vj}) \leq l^*)$
 9:         **end for**
10:     **end for**
11:     $\texttt{cset} := \{c(v) : v = 1, 2, \ldots, |V(G)|\}$
        return `cset`
12: **end function**

A.3 Lemmas and Theorems

Theorem A.1 (*de Finetti's Theorem*)**.** Let $\mathbf{X}_1, \mathbf{X}_2, ...$ be an infinite sequence of random variables with values in a space of $\mathbf{X}$. The sequence $\mathbf{X}_1, \mathbf{X}_2, ...$ is exchangeable *if and only if* there is a random probability measure η on $\mathbf{X}$ such that the X_i are conditionally i.i.d. given η.

Theorem A.2 (*Aldous Hoover Theorem*)**.** Let $\mathbf{A} = \{A_{ij}\}, 1 \le i, j \le \infty$ be a jointly exchangeable binary array if and only if there exists a random measurable function $f : [0, 1]^3 \to \mathbf{A}$ such that

$$\left(A_{ij}\right) \overset{d}{=} \left(f\left(U_i, U_j, U_{ij}\right)\right), \tag{23}$$

where $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij})_{i,j>i \in \mathbb{N}}$ with $U_{ij} = U_{ji}$ are a sequence and matrix, respectively, of *i.i.d* Uniform[0,1] random variables.

***Proof of spectral decomposition of diffusion distance*.** Diffusion distance of $\mathbf{G}$ can be represented via a spectral decomposition of its transition matrix P (Coifman and Lafon, 2006; Lafon and Lee, 2006). Recall that diffusion distance at time t, C_t, is a functional L^2 distance, weighted by $1/\pi$ in the statistic 9. Since an adjacency matrix A does not guarantee a symmetry of P, define a symmetric kernel $Q = \Pi^{1/2} P \Pi^{-1/2}$, where Π is a $n \times n$ diagonal matrix of which ith diagonal element is $\pi(i)$. Under compactness of P, Q has a discrete set of real nonzero eigenvalues $\{\lambda_r\}_{r=\{1,2,...,q\}}$ and a set of their corresponding orthonormal eigenvectors $\{\psi_r\}_{r=\{1,2,...,q\}}$, i.e. $Q[i, j] = \sum_{r=1}^{q} \lambda_r \psi_r(i) \psi_r(j)$ $(1 \le q \le n)$. Returning to the transition probability between Node i and Node j,

$$\begin{aligned}
P[i, j] &= \sqrt{\pi(j)/\pi(i)} Q[i, j] \\
&= \sum_{r=1}^{q} \lambda_r \{\psi_r(i)/\sqrt{\pi(i)}\} \{\psi_r(j)\sqrt{\pi(j)}\} \\
&:= \sum_{r=1}^{q} \lambda_r \phi_r(i) \{\psi_r(j)\sqrt{\pi(j)}\}
\end{aligned} \tag{24}$$

where $\phi_r(i) := \psi_r(i)/\sqrt{\pi(i)}$. Since $\sum_{r=1}^{q} \psi_r^2(j) = 1$ for all $j \in \{1, 2, ..., n\}$, we can represent the diffusion distance at time t as:

$$C_t^2[i, j] = \sum_{r=1}^{n} \lambda_r^{2t} \big(\phi_r(i) - \phi_r(j)\big)^2 \tag{25}$$

That is,

$$C_t[i, j] = \| \mathbf{U}_t(i) - \mathbf{U}_t(j) \| \tag{26}$$

where

$$\mathbf{U}_t(i) = \Big(\lambda_1^t \phi_1(i) \quad \lambda_2^t \phi_2(i) \quad \ldots \quad \lambda_q^t \phi_q(i)\Big)^T \in \mathbb{R}^q. \tag{27}$$

$\square$

Proof of Lemma 2.1 . By *Aldous-Hoover Theorem* A.2, a random array (A_{ij}) is jointly exchangeable *if and only if* it can be represented as follows :

There is a random function $g : [0, 1]^2 \to [0, 1]$ such that

$$(A_{ij}) \overset{d}{=} Bern(g(W_i, W_j)), \quad i, j = 1, \ldots, n; i < j \tag{28}$$

where $W_i \overset{i.i.d.}{\sim} Uniform(0, 1)$. Thus if $\mathbf{A}$ is an adjacency matrix of an undirected, exchangeable network, for any $i, j = 1, ..., n; i < j$:

$$
\begin{aligned}
f_A\big(a_{ij}\big) &= \int f_A\big(a_{ij}\big|w_i, w_j\big) f_W(w_i) f_W(w_j) dw dw \\
&= \int_0^1 \int_0^1 g(w_i, w_j)^{a_{ij}} \big(1 - g(w_i, w_j)\big)^{1-a_{ij}} dw_i dw_j
\end{aligned}
\tag{29}
$$

Then within each row, adjacent elements are independent and also identically distributed except a diagonal element.

$\square$

Proof of Lemma 2.2. We have shown that for fixed time t, diffusion distance is defined as an Euclidean distance of diffusion maps. Diffusion map at time t is represented as follows

:

$$\mathbf{U}_t(i) = \begin{pmatrix} \lambda_1^t \phi_1(i) & \lambda_2^t \phi_2(i) & \cdots & \lambda_q^t \phi_q(i) \end{pmatrix} \in \mathbb{R}^q. \tag{30}$$

where $\Phi = \Pi^{-1/2}\Psi$ and $Q = \Psi\Lambda\Psi^T = \Pi^{1/2}P\Pi^{-1/2}$. Thus $P\Pi^{-1/2}\Psi = \Pi^{-1/2}\Psi\Lambda$. Then for any rth row ($r \in \{1, 2, ..., q\}$, ($q \leq n$)), we can see that $P\phi_r = \lambda_r\phi_r$ where $\phi_r = \begin{pmatrix} \psi_r(1)/\sqrt{\pi(1)} & \psi_r(2)/\sqrt{\pi(2)} & \cdots & \psi_r(n)/\sqrt{\pi(n)} \end{pmatrix}$. Therefore to guarantee exchangeability (or $i.i.d$) of $\mathbf{U}_t$, it suffices to show exchangeability (or $i.i.d$) of P.

Assume joint exchangeability of $\mathbf{G}$, i.e. $(A_{ij}) \stackrel{d}{=} (A_{\sigma(i)\sigma(j)})$. Since A_{ij} is binary, $A_{ij}/\sum_j A_{ij} = A_{ij}/(1 + \sum_{l \neq j} A_{il})$. Moreover, A_{ij} and $(1 + \sum_{l \neq j} A_{il})$ are independent given its link function g, and $A_{\sigma(i)\sigma(j)}$ and $(1 + \sum_{l \neq j} A_{\sigma(i)\sigma(l)})$ are independent also given g. Then the following joint exchangeability of transition probability holds for $i \neq j; i, j = 1, 2, \ldots, n$:

$$(P_{ij}) = \left(\frac{A_{ij}}{1 - A_{ij} + \sum_{j=1}^{n} A_{ij}} \right) \stackrel{d}{=} \left(\frac{A_{\sigma(i)\sigma(j)}}{1 - A_{\sigma(i)\sigma(j)} + \sum_{\sigma(j)=1}^{n} A_{\sigma(i)\sigma(j)}} \right) = (P_{\sigma(i)\sigma(j)}) \tag{31}$$

When $i = j$, $P_{ij} = P_{\sigma(i)\sigma(j)} = 0$ for $i = 1, 2, \ldots, n$. Thus, transition probability is also exchangeable. This results exchangeable eigenfunctions $\{\Phi(1), \Phi(2), , ..., \Phi(n)\}$ where $\Phi(i) := \begin{pmatrix} \phi_1(i) & \phi_2(i) & \cdots & \phi_q(i) \end{pmatrix}^T$, $i = 1, 2, \ldots, n$. Thus diffusion maps at fixed t, $\mathbf{U}_t = \begin{pmatrix} \Lambda^t\Phi(1) & \Lambda^t\Phi(2) & \cdots & \Lambda^t\Phi(n) \end{pmatrix}$ are exchangeable. Furthermore by *de Finetti's Theorem* A.1, we can say that $\mathbf{U}(t) = \{\mathbf{U}_t(1), \mathbf{U}_t(2), \ldots, \mathbf{U}_t(n)\}$ are conditionally independent on their underlying distribution. $\square$

Proof of Lemma *2.3.* Based on Kallenberg and Exchangeable Graph (KEG) frameworks, introduced in Veitch and Roy (2015), a random array (A_{ij}) is jointly exchangeable in terms of Poisson process *if and only if* it can be represented as follows : there is a random function $g : \mathbb{R}_+^2 \to [0, 1]$ such that

$$(A_{ij}) \stackrel{d}{=} (A_{v_i,v_j}) \stackrel{d}{=} Bern(g(\vartheta_i, \vartheta_j)), \quad i, j = 1, \ldots, n; \quad i < j, \tag{32}$$

where $v_i \overset{i.i.d.}{\sim} Poisson(1), \vartheta_i \overset{i.i.d.}{\sim} Poisson(1), v_i \leq \nu$, for some pre-specified $\nu > 0$ so that finite-sized graph can deliberately include nodes only if they participate in at least one edges. Thus if $\mathbf{A}$ is an adjacency matrix of such undirected, exchangeable network, for any $i, j = 1, ..., n; i < j$:

$$
\begin{aligned}
f_{A|V}\big(a_{ij}\big|v_i, v_j\big) &= \int f_{A|V,\vartheta}\big(a_{ij}\big|v_i, v_j, \vartheta_i, \vartheta_j\big) f_\vartheta(\vartheta_i) f_\vartheta(\vartheta_j) d\vartheta_i d\vartheta_j \\
&= \int_0^\nu \int_0^\nu g(\vartheta_i, \vartheta_j)^{a_{ij}} \big(1 - g(\vartheta_i, \vartheta_j)\big)^{1-a_{ij}} \cdot dPois_1(\vartheta_i) \cdot dPois_1(\vartheta_j) d\vartheta_i d\vartheta_j.
\end{aligned}
$$

(33)

where $dPois_1(\cdot)$ is a probability distribution function of Poisson process with rate of 1. Thus given $\{\mathbf{V}\}$, edge probability except self-loop within each row (or column) is conditionally $i.i.d$ given a link function g and Poisson process V. $\qquad\square$

Proof of Lemma 2.4 *Convergence of empirical characteristic functions of exchangeable variables.* This follows exactly the same as *Theorem 1* in Székely et al. (2007). Note that this Lemma always holds without any assumption on $\{(\mathbf{w}_i, \mathbf{y}_j), i = 1, 2, ..., n\}$, e.g., it holds without assuming exchangeability, nor identically distributed, nor finite second moments. $\qquad\square$

Proof of Lemma 2.5 *Empirical characteristic function of exchangeable variables.* It suffices to prove the first argument 11 since the second argument 12 immediately follows from the first one by the property of characteristic functions. Proving the first one is equivalent to *Theorem 2* in Székely et al. (2007). However, they required $\{(\mathbf{w}_i, \mathbf{y}_i)\}$ to be independently identically distributed as $(\mathbf{w}, \mathbf{y})$ with finite second moments; here we have exchangeable $\{(\mathbf{w}_i, \mathbf{y}_i)\}$ instead.

Followed by *de Finetti's Theorem* A.1, if and only if $\{w_i\}$ are (infinitely) exchangeable, there exists an underlying distribution $f_{\mathbf{w}}$ of $\mathbf{w}$ such that $\mathbf{w}_i \overset{i.i.d}{\sim} f_{\mathbf{w}}$. By the same logic there exists a random, we have an underlying distribution $f_{\mathbf{y}}$ where $\mathbf{y}_i \overset{i.i.d}{\sim} f_{\mathbf{y}}$. Let $(\mathbf{w}_i, \mathbf{y}_i) \overset{i.i.d}{\sim} f_{\mathbf{w}, \mathbf{y}}$. Then under the assumption of finite second moment of the underlying distributions and measurable, conditioned random functions, we have a strong large number for V-statistics

followed by Székely et al. (2007), i.e.,

$$\int_{D(\delta)} \|g_{\mathbf{w},\mathbf{y}}^n(t,s) - g_{\mathbf{w}}^n(t)g_{\mathbf{y}}^n(s)\|^2 dh \xrightarrow{n\to\infty} \int_{D(\delta)} \|g_{\mathbf{w},\mathbf{y}}(t,s) - g_{\mathbf{w}}(t)g_{\mathbf{y}}(s)\|^2 dh, \qquad (34)$$

where $D(\delta) = \{(t,s) : \delta \leq |t|_p \leq 1/\delta, \delta \leq |s|_q \leq 1/\delta\}$, and $h(t,s)$ is the weight function chosen in Székely et al. (2007). $\qquad\square$

Proof of Theorem 2.6 *Consistency of* `dCorr` *applied to exchangeable variables.* Under the exchangeability and finite moments assumptions of underlying distribution, it follows from Lemma 2.4 and 2.5 that $\mathcal{V}_n^2(\mathbf{W},\mathbf{Y}) \xrightarrow{n\to\infty} 0$ if and only if underlying distribution of $\{\mathbf{w}_i\}$, $\mathbf{w}$ is independent from underlying distribution of $\{\mathbf{y}_i\}$, $\mathbf{y}$. Therefore, the `dCorr` or `mCorr` converges to 0 if and only if underlying distributions are independent; and its testing power converges to 1 under any joint distribution of finite moments. Since the multiscale generalized correlation based on any consistent global correlation is also consistent (Shen et al., 2016), `MGC` statistic constructed by `dCorr` or `mCorr` is also consistent in testing dependence. $\qquad\square$

Proof of Theorem 2.7 *Consistency of* `MGC` *applied to exchangeable variables.* Suppose that we have undirected, connected network $\mathbf{G}$ having a family of diffusion maps $\{\mathbf{u}_t(i) : i = 1,\ldots,n; \; t \in \mathbb{N}\}$ and with nodal attributes $\{\mathbf{x}_i : i = 1,\ldots,n\}$. We have shown in the Lemma 2.2 that $\{\mathbf{u}_t\}$ are exchangeable for each $t \in \mathbb{N}$. Thus there exists an underlying distribution of $\mathbf{u}_t$ such that $\mathbf{u}_t(i) \stackrel{i.i.d}{\sim} f_{\mathbf{u}^{(t)}}$ for each of $t = 1, 2, \ldots$; and we have $\mathbf{x}_i \stackrel{i.i.d}{\sim} f_{\mathbf{X}}$. Under the assumption of finite second moment of $\mathbf{u}^{(t)}$ and $\mathbf{x}$, `MGC` statistics constructed by $\{(\mathbf{u}_t(i), \mathbf{x}_i) : i = 1, 2, \ldots, n\}$ yield a consistent testing which determines the independence between underlying distributions of $\mathbf{u}^{(t)}$ and $\mathbf{x}$, followed by Lemma 2.5. From the same setting of network $\mathbf{G}$, we have estimated *i.i.d* node-specific network factors $\{\mathbf{F}_i\}$, we have n-pair of *i.i.d* $\{(\mathbf{F}_i, \mathbf{x}_i)\}$ and they can be applied it to `MGC` without assuming conditional underlying distribution. In case of using adjacency matrix directly into test, we must assume the adjacency matrix comes from directed network $\mathbf{G}$, i.e. $A_{ij} \stackrel{i.i.d}{\sim} f_A$ for all $i, j = 1, 2, \ldots, n$; otherwise, each column is dependent on one another. $\qquad\square$

References

Caron, F. and E. B. Fox (2014). Sparse graphs using exchangeable random measures. *arXiv preprint arXiv:1401.1137*.

Chan, S. H., T. B. Costa, and E. M. Airoldi (2013). Estimation of exchangeable graph models by stochastic blockmodel approximation. In *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, pp. 293–296. IEEE.

Christakis, N. A. and J. H. Fowler (2007). The spread of obesity in a large social network over 32 years. *New England journal of medicine 357*(4), 370–379.

Christakis, N. A. and J. H. Fowler (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine 358*(21), 2249–2258.

Coifman, R. R. and S. Lafon (2006). Diffusion maps. *Applied and computational harmonic analysis 21*(1), 5–30.

Cranmer, S. J., P. Leifeld, S. D. McClurg, and M. Rolfe (2016). Navigating the range of statistical tools for inferential network analysis. *American Journal of Political Science*.

Fosdick, B. K. and P. D. Hoff (2015). Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association 110*(511), 1047–1056.

Heller, R., Y. Heller, and M. Gorfine (2012). A consistent multivariate test of association based on ranks of distances. *Biometrika*, ass070.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association 97*(460), 1090–1098.

Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social networks 5*(2), 109–137.

Howard, M., E. Cox Pahnke, W. Boeker, et al. (2016). Understanding network formation in strategy research: Exponential random graph models. *Strategic Management Journal 37*(1), 22–44.

Ingold, K. and P. Leifeld (2014). Structural and institutional determinants of influence reputation: a comparison of collaborative and adversarial policy networks in decision making and implementation. *Journal of Public Administration Research and Theory*, muu043.

Kallenberg, O. (1990). Exchangeable random measures in the plane. *Journal of Theoretical Probability 3*(1), 81–136.

Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E 83*(1), 016107.

Lafon, S. and A. B. Lee (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE transactions on pattern analysis and machine intelligence 28*(9), 1393–1403.

Lovász, L. and B. Szegedy (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B 96*(6), 933–957.

Lyons, R. et al. (2013). Distance covariance in metric spaces. *The Annals of Probability 41*(5), 3284–3305.

Minhas, S., P. D. Hoff, and M. D. Ward (2016). Inferential approaches for network analyses: Amen for latent factor models. *arXiv preprint arXiv:1611.00460*.

Orbanz, P. and D. M. Roy (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence 37*(2), 437–461.

Shen, C., C. E. Priebe, M. Maggioni, and J. T. Vogelstein (2016). Discovering relationships across disparate data modalities. *arXiv preprint arXiv:1609.05148*.

Székely, G. J. and M. L. Rizzo (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis 117*, 193–213.

Székely, G. J., M. L. Rizzo, N. K. Bakirov, et al. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics 35*(6), 2769–2794.

Veitch, V. and D. M. Roy (2015). The class of random graphs arising from exchangeable random measures. *arXiv preprint arXiv:1512.03099*.

Wasserman, S. and P. Pattison (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika 61*(3), 401–425.

Young, S. J. and E. R. Scheinerman (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pp. 138–149. Springer.

SUPPLEMENTARY MATERIAL

All of the R functions and simulation data in RData format are provided in https://github.com/neurodata/Multiscale-Network-Test.