Testing independence between networks and nodal attributes via multiscale metrics

Youjin Lee*
Department of Biostatistics, Johns Hopkins School of Public Health
and
Author 2
Department of ZZZ, University of WWW

November 4, 2016

Abstract

Network dependence, which refers to the dependence between network topology and its nodal attributes, often exhibits nonlinear patterns. Unfortunately, without knowledge on metrics defined over network, no statistic has been proposed to test network dependence further beyond globally linear dependence. This paper introduces a multiscale dependence test statistic called Multiscale Network Test (`MNT`), which borrows the idea from diffusion maps and Multiscale Generalized Correlation (`MGC`). Our method can be applied to any exchangeable graph under some mild conditions, without further model assumption nor estimation. We prove the consistency of test statistic and demonstrate superior performance than any other model-based, global test statistics. Simulation in a variety of networks shows outstanding performance of the proposed test, especially in stochastic block model under nonlinear dependency. The application to dMRI networks will be followed.

Keywords: distance correlation, multiscale generalized correlation, diffusion maps, exchangeable graph, stochastic block model

*The authors gratefully acknowledge *please remember to list all relevant funding sources in the unblinded version*

1 Introduction

Network, a collection of nodes and edges between them, has been a celebrated area of study over a field of sociology (Pinquart and Sörensen, 2000; Ellison et al., 2007), information theory (Gross and Acquisti, 2005), biology (Barabasi and Oltvai, 2004; Pujol et al., 2010), statistics (Raftery et al., 2012; Palla et al., 2012), economics (Banerjee et al., 2013), etc. The correlation between network relationship between nodes and their attribute values is a common interest in network analysis. According to an assumption on how they are related each other, there has been lots of efforts to manifest a network as a function of nodal attributes (Wasserman and Pattison, 1996; Howard et al., 2016) or model an outcome of nodal attribute variables through their underlying network structures (Christakis and Fowler, 2007, 2008). However, it is very obscure to determine which one should be put as a dependent variable or even whether networks are truly related to nodal attributes and how they are, if any. Moreover, a fundamental difficulty comes from the empirical fact that network often does not have a natural structure. Thus it is not easy to intuitively come up with how to represent network as a node-specific random variable. 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. This might not be true for always. In this paper we develop a nonparametric test statistic which is also sensitive to nonlinear and local dependence pattern.

Throughout this paper, 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. Even though we assume that $\mathbf{G}$ is undirected and unweighted, we are able to extend all of the theory here to directed and even weighted network. An adjacency matrix of a given network, denoted by $\mathbf{A} = \{A_{ij} : i, j = 1, .., n\}$, is often introduced to formalize the relational data of

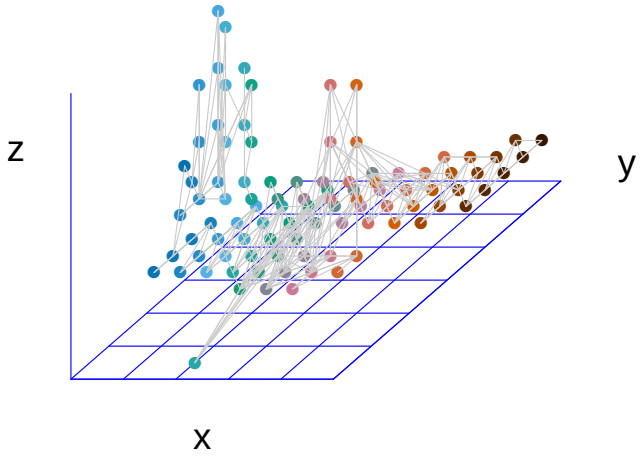

Figure 1: Physical location of one component of human brain network and its tracts that connect one vertex to another.

network, where $A_{ij} = 1$ if node i and node j are adjacent each other and zero otherwise. Let us introduce a m-variate ($m \in \mathbb{N}$) variable for 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 within a school and $\mathbf{X}$ is students' grades or heights, for example; or $\mathbf{G}$ can be a neuronal network in human brain and $\mathbf{X}$ is a few factors of personality. For example, Figure 1 exemplifies one connected network from human brain where dots denote nodes and tracts connected them represent edges. Location over network space, i.e. whether or how much a pair of nodes are closer than the others, is different from actual spatial location, which can be measured by Euclidean distance over three-dimensional xyz space. As correlation between spatial location of subjects and its attributes has been studied, we are going to explore correlation of subjects' attributes to their network location.

The main contribution of this study is to develop multiscale test statistics for network

independence which are robust to both nonlinearity and high dimensionality without modeling nor estimating networks. Having multiscale statistics is not avoidable because we regard location of or distance between nodes over network as a dynamic process. We then choose the optimal scale where distance in network space and distance in attributes maximize their correlation. To parse into this time-dependent distances, we construct a coordinate over network space and test independence to attributes $\mathbf{X}$ at each time in the process. In the next methodology section 2, we are going to introduce a test statistic using multiscale distance metrics which inherit desirable properties. In section 3, numerical 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. 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 week or too strong. Moreover, we could expect that dimension of network spectrum increases as the number of nodes increases. Unfortunately, widely used correlation measures often fail to capture nonlinear associations especially embedded in high-dimensional data set. Székely et al. (2007) extended pairwise constructed generalized correlation coefficient and developed a novel statistic called distance correlation (`dCorr`) as a measure for all types of dependence between two random vectors in any dimension.

Let us first start from a general setting that we are given $n \in \mathbb{N}$ pairs of $i.i.d$ random samples $\left(\mathbf{X}, \mathbf{Y}\right) = \{(\mathbf{x}_i, \mathbf{y}_i) : \mathbf{x}_i \in \mathbb{R}^q, \mathbf{y}_i \in \mathbb{R}^m, i = 1, ..., n\}$. Define $C_{ij} = \| \mathbf{x}_i - \mathbf{x}_j \|$ and $D_{ij} = \| \mathbf{y}_i - \mathbf{y}_j \|$ for $i, j = 1, 2, \ldots, n$, where $\| \cdot \|$ denotes Euclidean distance of any vectors. Distance correlation (`dCorr`) is standardized version of a distance covariance (`dCov`) $\mathcal{V}_n^2$ of

$\mathbf{X}$ and $\mathbf{Y}$, which is the following:

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

where $\tilde{C}$ and $\tilde{D}$ is a doubly-centered C and D respectively, by its column mean and row mean. Distance correlation $\mathcal{R}_n^2(\mathbf{X}, \mathbf{Y})$ is scaled by $\mathcal{V}_n^2(\mathbf{X}, \mathbf{X})$ and $\mathcal{V}_n^2(\mathbf{Y}, \mathbf{Y})$.

$$\mathcal{R}_n^2(\mathbf{X}, \mathbf{Y}) = \frac{\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\mathcal{V}_n^2(\mathbf{X}, \mathbf{X}) \mathcal{V}_n^2(\mathbf{Y}, \mathbf{Y})}} \tag{2}$$

In addition, a modified distance covariance (`mCov`) $\mathcal{V}_n^*$ and a modified distance correlation (`mCorr`) $\mathcal{R}_n^*$ for testing high dimensional random vectors were also proposed in Székely and Rizzo (2013). However, `dCorr` nor even `MCorr` still perform not very well in the existence of various nonlinear dependency and under existence of outliers [Cencheng]. Out of this concern, Cencheng at al (2016) proposed Multiscale Generalized Correlation (`MGC`) by adding local scale in a sense of nearest neighbors on correlation coefficients. Multiscale version of distance covariance $\{\mathcal{V}^{*2}_n\}_{kl}$ is defined as following :

$$\begin{aligned} \mathcal{V}^{*2}_n(\mathbf{X}, \mathbf{Y})_{kl} &= \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{C}_{ij} \tilde{D}_{ij} I\big(r(C_{ij}) \leq k\big) I\big(r(D_{ij}) \leq l\big) \\ &=: \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{C}^*_{ij} \tilde{D}^*_{ij}, \quad k, l = 1, 2, ..., n, \end{aligned} \tag{3}$$

where $r(C_{ij})$ $(r(D_{ij}))$ denotes a rank $\mathbf{x}_i$ $(\mathbf{y}_i)$ relative to $\mathbf{x}_j$ $(\mathbf{y}_j)$. It basically truncates each pairwise element of distance covariance with respect to rank in terms of (Euclidean) distance. Note that if $k = l = n$, $\mathcal{V}_n^2$ and $\mathcal{V}^{*2}_n$ are equivalent. You can see panels to illustrate both Euclidean distance matrices and truncated, double-centered Euclidean distance matrices in Figure 11. Simply speaking, given an appropriate distance matrix for $\{\mathbf{x}_i : i = 1, 2, \ldots, n\}$ and $\{\mathbf{y}_i : i = 1, 2, \ldots, n\}$ each, we take an double centering and truncate them by column ranking up to k and l in order to obtain $\tilde{C}^*$ or $\tilde{D}^*$ respectively. Since we call all family of $\{\mathcal{R}^{*2}_n\}_{k,l=1,2,...,n}$ as `MGC`, `MGC` is more generalized version of `dCorr`. How to choose the optimal neighborhood scale of (k, l), say (k^*, l^*), is illustrated in [Cencheng] as well as its

superiority and consistency. In the simulation section 3 we are going to show which pattern of underlying dependency particularly `MGC` exhibits improved performance than the global scale. On the other hand, we often do not know the correct optimal scale of (k^*, l^*) given one single observations; while the optimal neighborhood choice can be closely estimated by simulated networks. Following the terminology used by [Cencheng], we call `MGC` when (k^*, l^*) should be estimated upon a single network as `Sample MGC` while `Oracle MGC` denotes `MGC` at near the true scale of (k^*, l^*). We are able to estimate `Oracle MGC` in simulation data only or when underlying network is known.

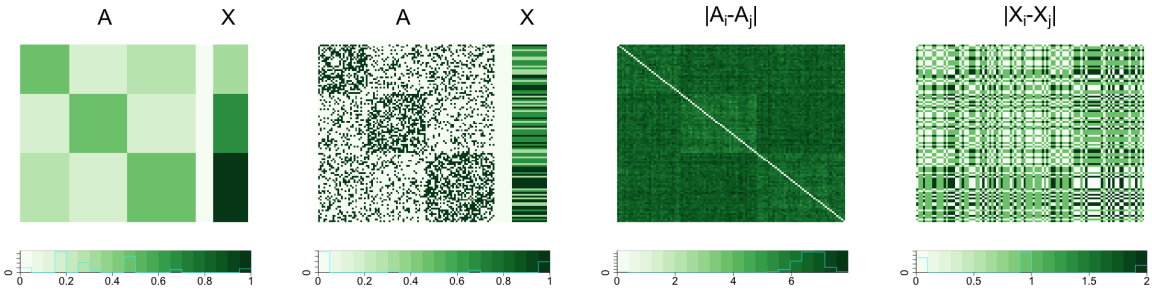

Figure 2: Population probability distribution and realized values of A and X (scaled by 1/3) in the left two panels. Euclidean distance applied to realized A and X are presented in the right two panels. Data generation follows model 26.

When we are given multivariate, real valued nodal attributes $\mathbf{X}$, its Euclidean distance D is easily constructed. However what would be the distance metric over network $\mathbf{G}$ is still remaining question. We are required to find $i.i.d$ node-specific coordinates of which Euclidean distance as C in Figure 11 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 rationale 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 certain family of graphs. Instead of assuming *i.i.d* edges as a directed graph allowing self-loop, we are going to regard a set of edges as *jointly exchangeable* and then exploit *exchangeable representation* theorem, which furnishes representation as conditional *i.i.d* observations. A graph **G** is called exchangeable if and only if its adjacency matrix **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 there exists a random probability measure η on exchangeable random variable **Z** so that a sequence of $Z_1, Z_2, \ldots$ are *i.i.d* conditioned on η *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 $u_i \overset{i.i.d}{\sim} Uniform[0,1], i = 1, 2, ..., n$. Let A be an adjacency matrix of graphon. Then for any $i < j, \quad i, j = 1, 2, ..., n$:

$$Pr\big(A_{ij} = 1 \big| u_i, u_j\big) = g\big(u_i, u_j\big) \tag{4}$$

By *Aldous-Hoover theorem*, we obtain clear representation of exchangeable network through measurable function g, but we are still halfway done in the case of undirected network where $A_{ij} = A_{ji}$ $(i, j = 1, 2, ..., n)$. Under undirected network without self-loop, we

have $\{A_{ij} : i < j\}$ as a function of g.

$$(A_{ij}) = (A_{\sigma(i)\sigma(j)}) \Longleftrightarrow A_{ij} \overset{i.i.d}{\sim} Bern\big(E_U\big[g(u_i, u_j)\big]\big), \quad i < j \tag{5}$$

Networks based on widely used graphical model are exchangeable. One of the most popular models is Stochastic Block Model (SBM) (Holland et al., 1983). In the simplest setting of SBM, we assume that each n nodes of $\mathbf{G}$ 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 latent variable corresponding to block affiliation follow $Z_1, Z_2, ..., Z_n \overset{i.i.d.}{\sim} Multinomial\big(\pi_1, \pi_2, ..., \pi_K\big)$. Then the upper triangular entries of A are independent and identically distributed conditional on $\{\mathbf{Z}\}$:

$$A_{ij} \overset{i.i.d.}{\sim} Bern\big(\sum_{k,l=1}^{K} p_{kl} I\big(Z_i = k, Z_j = l\big)\big), \forall i < j. \tag{6}$$

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 fo A.

$$\begin{aligned} A_{ij}\big|g, W_i, W_j &\overset{ind}{\sim} Bern\big(g(W_i, W_j)\big), \forall i < j \\ A_{ij}\big|g &\overset{i.i.d}{\sim} \int\int Bern\big(g(W_i, W_j)\big) f_W(w_i) f_W(w_j) dw dw, \quad \forall i < j. \end{aligned} \tag{7}$$

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

2.2.1 Exchangeability on point process

Graphon has been studied widely as a limit of random graphs (Lovász and Szegedy, 2006). However, despite its advantage on simple representation, it is either empty or dense.

A precise definition of dense graph and sparse graph is followed by Veitch and Roy (2015).

Definition 2.3 (sparse, not dense, graph)**.** Let $\mathbf{G} = \big(V, E\big)$ be a graph and $|V|$ and $|E|$ denote the number of nodes and edges of G respectively. Then graph G is called sparse or equivalently not dense if and only if $|E|$ is asymptotically $o(|V|^2)$, i.e.

$$\frac{\sqrt{|E|}}{|V|} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \tag{8}$$

Our exchangeable graph 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 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, undirected graph on a point process on $\mathbb{R}_+^2$ can be thought of

$$\mathbf{A} = \sum_{i,j} A_{ij} \delta_{(\theta_i, \theta_j)} \tag{9}$$

where $A_{ij} = A_{ji} \in \{0, 1\}$ with node label space $\theta \in \mathbb{R}_+$, $i, j = 1, 2, ...$, e.g. node i is assumed to be embedded on real line, at $\theta_i \in \mathbb{R}_+$.

Definition 2.4 (graphex Kallenberg (1990))**.** Random graphs called *graphex* defined on exchangeable random measures are characterized by $g : \mathbb{R}_+^2 \to [0, 1]$. Conditional on random functions g and unit rate Poisson processed $\theta \times \vartheta$, *graphex* $\mathbf{G}$ with node set $\{\theta\}$ is constructed as:

$$A_{\theta_i \theta_j} \overset{d}{=} g(\vartheta_i, \vartheta_j) \tag{10}$$

and exclude θ_i from $\mathbf{G}$ if θ_i is isolated. Then we can obtain finite, connected subgraphs by restricting $\theta < \nu$ for some $\nu > 0$.

In graphex, joint exchangeability applied to node itself now corresponds to joint exchangeability of a point processed node label θ, not on a node label itself:

Definition 2.5 (Joint exchangeability on point process)**.** Let $h > 0$ and $V_i = [h(i-1), hi]$ for $i \in \mathbb{N}$ then

$$\big(A(V_i \times V_j)\big) \stackrel{d}{=} \big(A(V_{\sigma(i)} \times V_{\sigma(j)}\big) \tag{11}$$

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

Despite its more intricate form mostly due to Poisson process on node, representation of sparse graph as exchangeable formation helps us to demonstrate the validity of our proposed methods in real network data. Exchangeable variables which reflect locations of each node over exchangeable network will be followed.

2.3 Multiscale Distance Metrics

2.3.1 Diffusion maps and diffusion distance

Then which distance metric should be applied for exchangeable graph? That metric must enjoy conditional *i.i.d* from exchangeable graph. Actually despite a surge of research on network representation in terms of a summarizing network factor (Hoff et al., 2002) or some meaningful coefficients, e.g.centrality (Mantzaris et al., 2013; Sporns et al., 2007), there has been no node-specific variable which provides a configuration of node over network space without losing any information. Coifman and Lafon (2006) proposed a meaningful multiscale geometries of data called *diffusion maps* while keeping every information on every local relation over a graph, defined on every discrete time. Distance between a pair of nodes at each time is called *diffusion distance* which is constructed by Euclidean distance of such diffusion maps at each time point. For each time point $t \in \mathbb{N}$, we can define a diffusion distance at time t, C_t, which is also related to a discrete set of real nonzero eigenvalues $\{\lambda_r\}$ and eigenvectors $\{\phi_r\}$ of a transitionmatrix (Coifman and Lafon, 2006; Lafon and Lee, 2006).

$$C_t^2[i,j] := \| \mathbf{U}_t(i) - \mathbf{U}_t(j) \| \quad i,j = 1,2,\ldots,n \tag{12}$$

where

$$\mathbf{U}_t(i) = \left(\lambda_1^t\phi_1(i) \quad \lambda_2^t\phi_2(i) \quad \dots \quad \lambda_q^t\phi_q(i)\right)^t \in \mathbb{R}^q. \tag{13}$$

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. If you see Figure 3, difference between blocks in distance matrix at $t = 3$ looks more distinct than that at $t = 1$. However if the propagation takes enough, e.g. at $t = 50$, it becomes hard to detect the differences as nodes under the peer influence are at the end assimilated. Compared to adjacent relation or geodesic distance, diffusion distance well reflects the connectivity since it takes into account every possible path between two nodes.

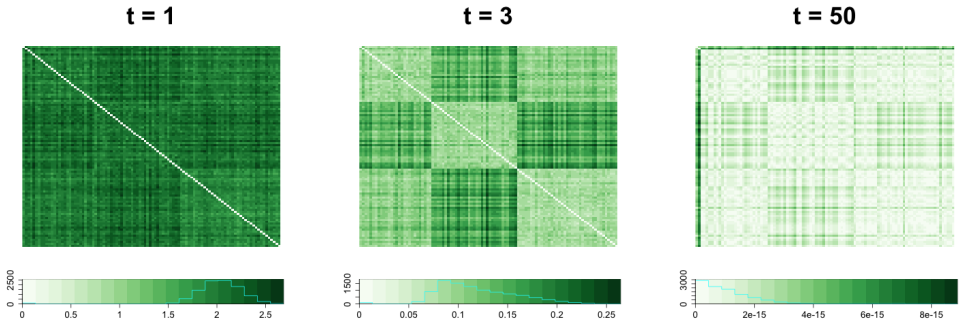

Figure 3: Diffusion distance, i.e. Euclidean distance of diffusion maps at $t = 1$, $t = 3$ and $t = 50$ of sample graph G from Stochastic Block Model provided in simulation (Eq. 26).

Now we have a family of q-variate($q \leq n$) diffusion maps $\{\mathbf{U}_t\}_{t\in\mathbb{N}}$, of which Euclidean distance is diffusion distance. Embedding each node on Euclidean metric is a novel approach in testing independence on network space; there is no estimation nor model assumption involved. However there remains a matter of dependence between observed diffusion maps.

2.3.2 Properties of diffusion maps under exchangeable graphs

We wish that diffusion maps are multivariate configuration of each node whose distance metric well reflects relative location on network space. However, due to the inter-correlated construction of $\mathbf{U}_t$, e.g. ith subject's diffusion depends on others in the given network, it is hard to say that the observed diffusion coordinates of n subjects are independent observations. The following Lemmas show how $\mathbf{U}_t$ retains (conditional) independence privileged to

exchangeable graph.

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 $g : [0, 1]^2 \to [0, 1]$. 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* given a random link function g or equivalently, its underlying distribution.

From the above Lemma 2.1, we can also prove exchangeability and conditional *i.i.d* of diffusion maps family.

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, i.e. any exchangeable random graph from an infinite graph. Then its transition probability so thus diffusion maps at fixed time t also exchangeable, conditional on 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, specifically given random probability measure η on U_t and random link function g on graph.

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 exchangeable graph, which cannot be sparse. If we want to embed a set of nodes in sparse graphs, one more step of conditioning on point process θ is needed, explained in Def. 2.4.

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 $g : [0, 1]^2 \to [0, 1]$ 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 prove exchangeability of a transition matrix in graphex case, which extends to conditional *i.i.d* of its diffusion maps.

We have discussed embedding a vertex $v \in V(\mathbf{G})$ of exchangeable graph into its diffusion map of $\{\mathbf{U}_t\}$. As explained earlier, its Euclidean distance, which is ready to be applied to `MGC` statistics 3, takes into account all possible paths between every pair of nodes and measure the connectivity between them. Unlike in the other metrics in network, i.e. adjacency matrix or geodesic distance, triangle inequality holds in diffusion distance. Proof is provided in Appendix. Thanks to these properties of diffusion maps, we earn better interpretation of its Euclidean distance.

Corollary 2.3.1 (Triangle inequality)**.** For any fixed time $t(\in \mathbb{N})$, let $C_t : V(\mathbf{G})^2 \to \mathbb{R}_+$ be a diffusion distance defined on a pair of nodes in any connected and undirected graph $\mathbf{G}$. Then for any $v, w, z \in V(\mathbf{G})$,

$$C_t(v, z) \leq C_t(v, w) + C_t(w, z) \tag{14}$$

2.3.3 One parameter family of test statistic

If exchangeable observations are applicable to distance-based test statistics, e.g. `MGC`, we are almost able to dispense with obstacles in testing network independence. Assume that we have a finite sample of infinitely exchangeable sequence $(\mathbf{X}, \mathbf{Y}) = \{(\mathbf{x}_i, \mathbf{y}_i) : i = 1, 2, \ldots, n\}$, which is identically distributed as $(\mathbf{x}, \mathbf{y})$ with finite second moment. Then, by the properties of exchangeable sequences, there exist some random measure θ and P such that :

$$\begin{aligned} &x_i|\theta \overset{i.i.d}{\sim} \prod_{i=1}^{n} f_{\mathbf{x}_i|\theta}(\mathbf{x}_i|\theta) \\ &\int \prod_{i=1}^{n} f_{\mathbf{x}_i|\theta}(\mathbf{x}_i|\theta) P(d\theta) \overset{d}{=} \prod_{i=1}^{n} f_{\mathbf{x}}(\mathbf{x}_i) \\ &\mathbf{x}_i \overset{i.i.d}{\sim} f_X(\mathbf{x}) \quad \text{conditioning on the underlying distribution } f_{\mathbf{X}}, \end{aligned} \tag{15}$$

where f is a random, marginal distribution integrated over θ; same arguments hold for exchangeable $\{\mathbf{y}_i\}$.

Lemma 2.4. Under the above conditions, we have

$$\mathcal{V}_n^2(\mathbf{X},\mathbf{Y}) \;\; = \|g^n_{\mathbf{x},\mathbf{y}}(t,s) - g^n_{\mathbf{x}}(t)g^n_{\mathbf{y}}(s)\|^2,$$

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

$$\begin{aligned} g^n_{\mathbf{x},\mathbf{y}}(t,s) &= \tfrac{1}{n}\textstyle\sum_{j=1}^n \exp\{i\,\langle t,\mathbf{x}_j\rangle + i\,\langle s,\mathbf{y}_j\rangle\}, \\ g^n_{\mathbf{x}}(t) &= \tfrac{1}{n}\textstyle\sum_{j=1}^n \exp\{i\,\langle t,\mathbf{x}_j\rangle\}, \\ g^n_{\mathbf{y}}(s) &= \tfrac{1}{n}\textstyle\sum_{j=1}^n \exp\{i\,\langle s,\mathbf{y}_j\rangle\}. \end{aligned}$$

Lemma 2.5. We have

$$\mathcal{V}_n^2(\mathbf{X},\mathbf{Y}) \;\; \longrightarrow \mathcal{V}^2(\mathbf{x},\mathbf{y}) \qquad \text{as } n\to\infty \tag{16}$$

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

$$\mathcal{V}_n^2(\mathbf{X},\mathbf{Y}) \;\; \to 0 \quad \text{as } n\to\infty \tag{17}$$

if and only if $g_{\mathbf{x},\mathbf{y}}(t,s) = g_{\mathbf{x}}(t)g_{\mathbf{y}}(s)$, i.e., $\mathbf{x}$ 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 $(\mathbf{X},\mathbf{Y}) = \{(\mathbf{x}_i,\mathbf{y}_i), i = 1,2,\ldots,n\}$ having finite second moment. Assume $\mathbf{x}_i \overset{i.i.d}{\sim} f_{\mathbf{x}}$ and $\mathbf{y}_i \overset{i.i.d}{\sim} f_{\mathbf{y}}$ given underlying distribution for $i = 1,2,\ldots,n$. Then

$$\mathcal{V}_n^2(\mathbf{X},\mathbf{Y}) \;\; \longrightarrow 0 \quad \text{as } n\to\infty \tag{18}$$

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

dependency of finite moments.

Note that if $\{\mathbf{x}_i : i = 1, 2, \ldots, n\}$ 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. `MGC` is consistent in testing network independence with any exchangeable graph metric and nodal attributes, in particular testing independence between underlying distribution of diffusion maps and nodal attributes $\{(\mathbf{u}_{ti}, \mathbf{x}_i) : i = 1, 2, \ldots, n\}$ such that conditioned on each underlying distribution, $\mathbf{u}_{ti} \overset{i.i.d}{\sim} f_{U(t)}$ and $\mathbf{x}_i \overset{i.i.d}{\sim} f_{\mathbf{X}}$.

$$H_0 \quad : \quad f_{\mathbf{U}^{(t)}.\mathbf{X}} = f_{\mathbf{U}^{(t)}} \cdot f_{\mathbf{X}} \tag{19}$$

In particular, the consistency also holds for the estimated latent positions and adjacency matrix of directed network.

Underlying distribution of $\mathbf{U}_t$, i.e. $\iint f_{\mathbf{U}_t}(\eta, g) P(d\eta) P(dg) = f_{\mathbf{U}^{(t)}}(\mathbf{u})$, given a link function g and a random probability measure η_t of $\mathbf{U}_t$ has been introduced to draw *i.i.d* from exchangeability. Since a family of diffusion maps, $\{\mathbf{U}_t\}_{t \in \mathbb{N}}$ provides a configuration of every node in $\mathbf{G}$, the above hypothesis implies testing independence between the configuration of nodes in network space and in attribute space at each time of diffusion given the underlying distribution of $\mathbf{U}_t$. 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.

2.3.4 Each node's contribution to dependency measure

In the presence of nonlinear dependency or local (in)dependency, some nodes often exerts more reliance on their attributes than the others since the amount of dependence is not consistent over a set of nodes. Like other node-specific measure of its importance, e.g. centrality, each node's leverage on dependency measure can be of interest. Here we actually

quantify each node's contribution to `MGC` statistic so that we know how much its location on network space and its attributes are correlated. Let (k^*, l^*) be the optimal neighborhood choice in distance matrix (C, D). Denote the contribution of node $v \in V(G)$ to the testing statistic by $c(\cdot) : v \to \mathbb{R}$.

$$c(v) \propto \sum_{j=1}^{n} \tilde{C}_{vj} \tilde{D}_{vj} I\big(r(C_{vj}) \leq k^*\big) I\big(r(D_{vj}) \leq l^*\big) \tag{20}$$

which is v^{th} column-sum of the test statistic 3. Since $\mathrm{I}\big(\ r(C_{vj}) \leq k^*\big)$ indicates we only include j^{th} subject's k^* nearest neighbor, and same arguments for $I\big(r(D_{vj}) \leq l^*\big)$, a node-specific attribution $c(v)$ does not take into account element-wise distance product beyond the optimal level of neighborhood.

3 Simulation Study

In simulation studies presented in this paper, 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 model of latent factors, 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 $\mathbf{X}$.

$$\begin{aligned} f(\mathbf{A}, \mathbf{X}, \mathbf{Z}) &= f_{A|Z}(\mathbf{A}|\mathbf{Z}) \cdot f_{Z|X}(\mathbf{Z}|\mathbf{X}) \cdot f_X(\mathbf{X}) \\ &= f_{A|Z}(\mathbf{A}|\mathbf{Z}) \cdot f_{X|Z}(\mathbf{X}|\mathbf{Z}) \cdot f_Z(\mathbf{Z}) \end{aligned} \tag{21}$$

According to the joint model in Eq. 21, edge distribution 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. One example of empirical distributions under both independence and dependence are shown in Figure 4 where the vertical lines indicate 95% quantiles of the null.

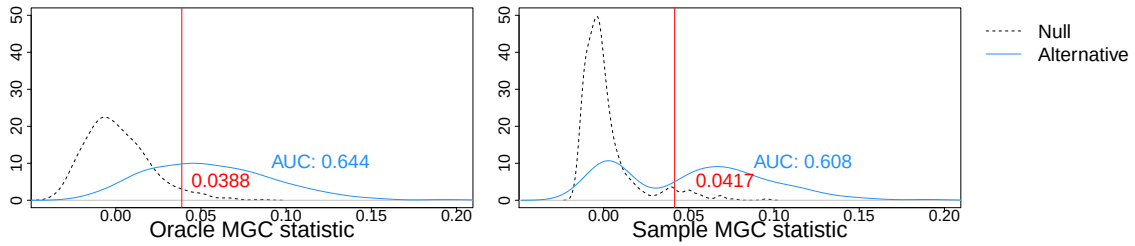

Figure 4: Statistics under null distribution and dependent distribution based on $M = 500$ independently generated SBM presented in model 26. Area under the curve (AUC) beyond 95 % quantiles of the null is equivalent to the empirical power.

3.1 Stochastic Block Model

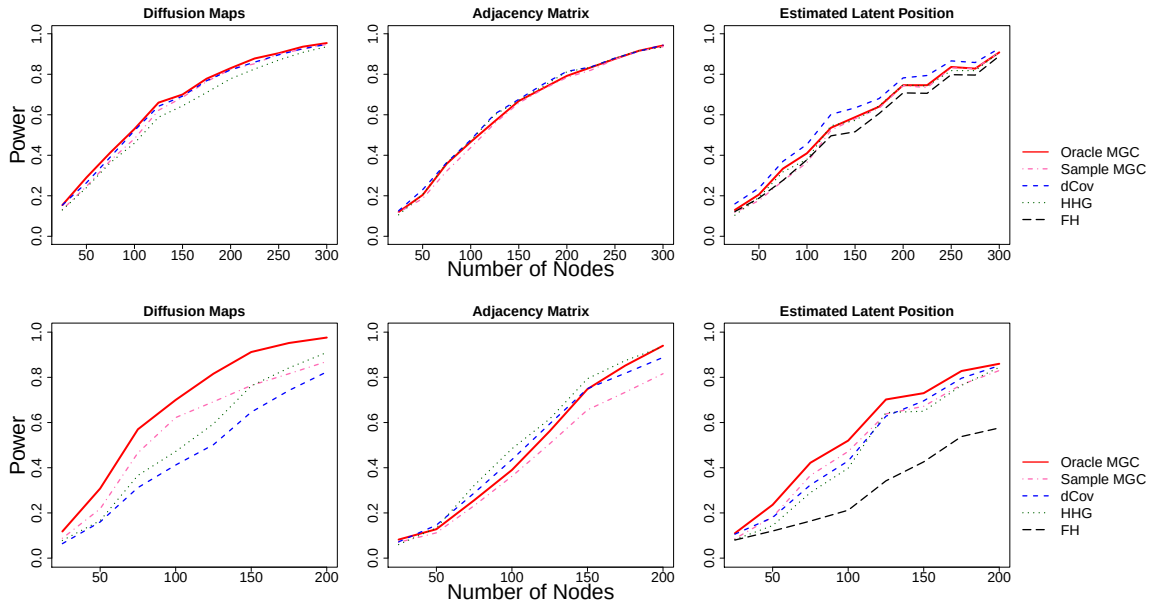

Figure 5: Empirical power based on $M = 500$ independently generated SBM presented in model 25 (top) and model 26 (bottom) using diffusion maps (left) and Euclidean of adjacency matrix (middle) and estimated latent position(right). The most right figure contains the results of `FH` test as well.

We already mentioned that Stochastic Block Model (SBM) is one of the most popular and also useful network generative model, especially as a tool for community detection (Karrer and Newman, 2011). We present the SBM with $K = 2$ blocks (model 25) and $K = 3$ blocks (model 26) where block affiliation for each node is correlated with its attributes X. Figure 5 describes empirical power of `MGC`, `dCov`, and `HHG` based on diffusion distance, Euclidean distance of adjacent matrix and Euclidean distance latent factors. In two block SBM, the performance of `MGC` is very similar to others while it is most outstanding in three block SBM. This comparative advantage of `MGC` is attributed to its ability to capture non-linear dependence. Note that expectation of having edge between a pair of nodes is a monotonic function of (Euclidean) distance of their attributes in model 25 but non-monotonic in model 26 as formalized below.

$$\begin{aligned} E(A_{ij}|X_i, X_j) &= 0.6I(|X_i - X_j| = 0) + 0.4I(|X_i - X_i| = 1) \\ E(A_{ij}|X_i, X_j) &= 0.5I(|X_i - X_j| = 0) + 0.2I(|X_i - X_j| = 1) + 0.3I(|X_i - X_j| = 2) \end{aligned} \tag{22}$$

Due to non-linear network dependence on attributes in the latter case, optimal neighborhood choice of (k^*, l^*) results in not global scale. Figure 13 in Appendix demonstrates that neighborhood choice of (k, l) in `MGC` statistic achieves its optimal in local scale. Roughly speaking, considering a pair of nodes in the nearest neighbor, in term of attribute values of X, e.g. (1,2) or (2,3), exhibits most significant dependence. When you only take into account these subsets of observations, you can actually represent $E(A_{ij}|X_i, X_j)$ as a monotonic function of $|X_i - X_j|$ (equivalently $\| \mathbf{X}_i - \mathbf{X}_j \|$ in multivariate case). On the other hand, if you keep linear dependence even in three block SBM, discrepancy between local and global scale of (k, l) diminishes so thus global distance-based tests do as good as `MGC`, which you can find in **Three Block SBM 2** and Figure 14, Appendix A.

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 the value of $\theta \in (0, 1)$. When $\theta > 0.2$, linear dependency of edge distribution upon nodal

attribute X is lost.

$$Power(\theta) = E(A_{ij}|X_i, X_j) = 0.5I(|X_i - X_j| = 0) + 0.2I(|X_i - X_j| = 1) + \theta I(|X_i - X_j| = 2) \tag{23}$$

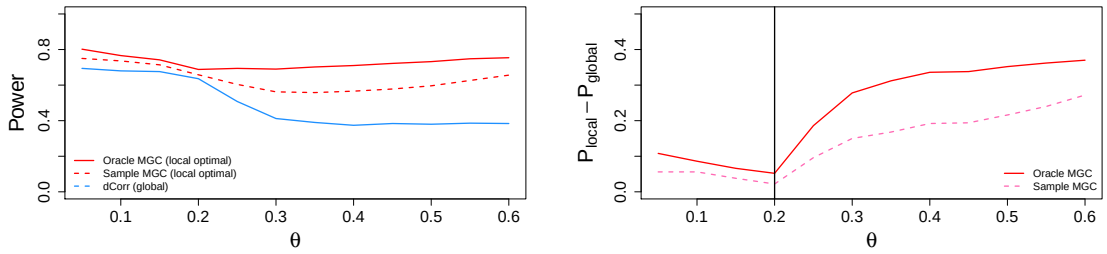

Figure 6: Change of empirical power across θ in both local and global scale of distance correlation (left). Change of difference between these two powers in Oracle and Sample `MGC`. Superiority of optimal local scale become evident from $\theta > 0.2$, when distribution of edges have non-linear dependence on X.

If you see Figure 6, power of `dCorr` starts to drop from $\theta = 0.2$ while that of `MGC` almost stays clam, which implies `MGC` keeps its sensitivity even under nonlinear dependency compared 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, this block model is limited by homogeneous distribution within block and provides a poor fit to networks with hubs or highly varying node degrees within blocks or communities. Instead, the Degree-Corrected Stochastic Block model (DCSBM) proposed by Karrer and Newman (2011) varies distribution of node degree within a block, preserving the overall block community structure. Consider two block SBM having more variability induced by θ (model 28).

As in the previous two block SBM, in this simulation scheme, we also observe monotonic edge distribution function of Euclidean distance. However its variance becomes inflated due to θ (model **??**). Relatively poor performance of an adjacency metric as presented in Figure 7

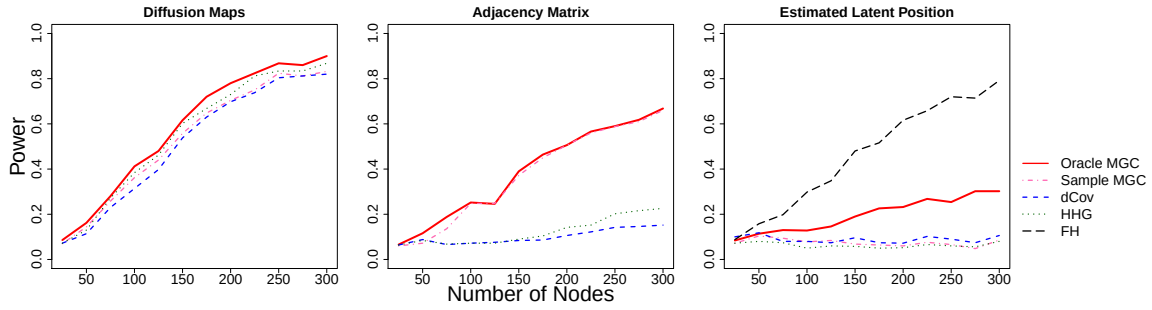

Figure 7: Empirical power based on $M = 500$ independently generated SBM presented in model 28 using diffusion maps (left) and Euclidean of adjacency matrix (middle) and estimated latent position(right). The most right figure contains the results of `FH` test as well.

can be attributed to large variability in it. Even in this case, multiscale statistic performs better than the other global ones. On the other hand, node-specific latent factors look more sensitive in `FH` model shown in the right panel, but still diffusion maps metrics work better.

3.3 Additive and multiplicative graph model

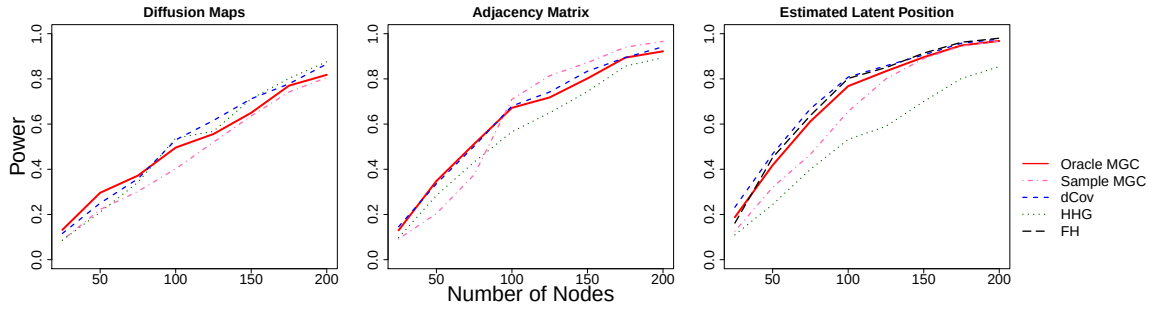

Figure 8: Empirical power based on $M = 500$ independently generated additive and multiplicative graph model presented in model 29 using diffusion maps (left) and Euclidean of adjacency matrix (middle) and estimated latent position(right). The most right figure contains the results of `FH` test as well.

Hoff et al. (2002) proposed an approach of jointly modeling network and its attributes, where networks possess additional structure via sender-specific(or row-specific) and receiver-specific(or column-specific) latent factors. Whereas Fosdick and Hoff (2015) embedded nodes into network factors assuming that additive an multiplicative 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 29, where logit of A is an additive and multiplicative function of their estimated factors, estimated latent position would be very close to the truth, much closer than embedding made from diffusion maps. However we rarely see the network which is fitted to 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 8). 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 testing independence 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.

3.3.1 Node Contribution Test

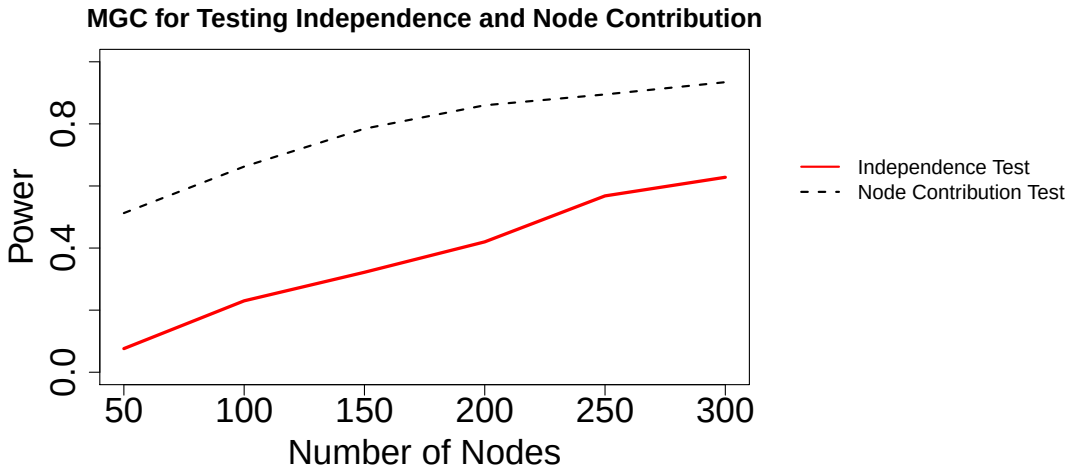

Figure 9: Change of empirical power and inclusion rate at each total sample size n. You can see that inclusion rate of $c(v)$ increases as empirical power of `MGC` increases.

To examine the effectiveness of node contribution measure in testing dependency as presented in Eq. 20, 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 30). 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*:

$$\text{inclustion rate}\big(c(v)\big) = \#\big\{rank_{c(v)}\big(u\big) \leq m\big\}/m, \tag{24}$$

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

4 Real Data Examples

4.1 MRI

needs contexts

We look into one of the connected components with $n = 95$ nodes of whole disconnected network.

$\{\mathbf{U}_t \in \mathbb{R}^{95}\}_{t\in\mathbb{N}}$ $\mathbf{X} = (x, y, z) \in \mathbb{R}^3$

It turns out that $q = 95$, i.e. full rank of transition matrix. In other words we are testing independence between 95-dimensional diffusion maps and 3-dimensional nodal attributes at each $t \in \mathbb{N}$.

test independence between brain network and its 3-dimensional locations; test independence between functional location and physical location.

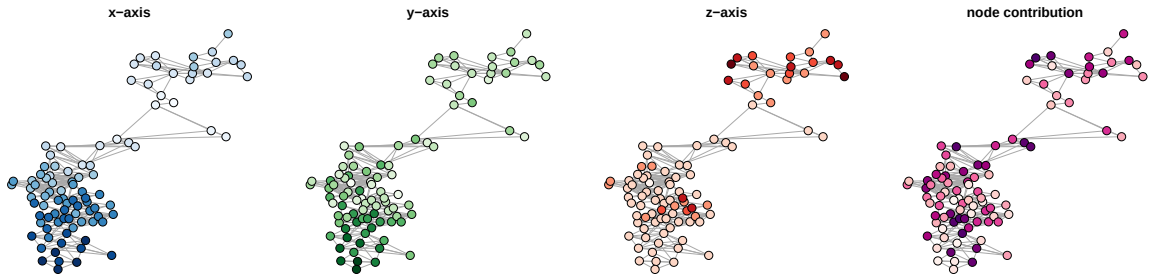

Figure 10: Subnetwork of MRI network. Darker colored nodes indicate higher positioned node in terms of x-axis(left), y-axis(middle), and z-axis(right).

5 Discussions

Throughout this study, we demonstrate that multiscale network test statistic to test independence between network topology and its nodal attributes performs well in diverse settings, being supported by thorough theory on distance correlation and diffusion maps. These two tools are both specialized in detecting local and nonlinear dependence patterns, which other existing methods are lack of. However, testing independence is often the very first step in investigating relationship between network topology and nodal attributes in our interest. It is more likely that we want to know more than binary decision of rejecting or not rejecting the hypothesis. Multiscale test statistics attributed to both neighborhood choice $\{(k, l)\}$ and time spent in diffusion processes $\{t\}$ provides us a hint on latent dependence structure as well. On the other hand, due to the ambiguousness of saying *optimal*, our work has some limitations; we do not suggest any theoretically supported tools to select the *optimal* time to obtain p-values of test statistics. Future research can be focused on restoring true dependence pattern or estimating *optimal* scale from a family of statistics. In addition, obtaining a full family of statistics are also computationally infeasible; at every Markov process, we chose the optimal region of neighborhood. As an ad hoc, we selected optimal t with highest power or lowest p-values from 1 to 10 for our simulation 3. Despite a few shortcomings, multiscale version of test would be very useful when indirect networking through diffusion process is not ignorable or cluster membership significantly affects attributes. Furthermore even though we specifically constraint the statistic into testing independence between network and nodal attributes, we are able to implement independence testing of two networks with same size by inputting diffusion distance at each time point from each of network in Eq. 3. This type of test will be useful when we want to show a pair of networks are topologically or structurally independent. For example, we might wonder if network on *Facebook* and network induced by club activity within class or school are independent or not; if DNA methylation network is correlated with gene expression network (Bartlett et al., 2014) so that the behavior of interest measured by DNA methylation network can be matched to that noticed by gene expression network, etc. In a broad sense, we suggest a proper metric applied to network and justify its use in measuring correlations between *network vs. attributes* or possibly *network*

vs. network. Even though we have fully presented procedures, e.g. testing on diffusion maps from $t = 1$ to $t = 10$ and obtaining a family of all local scales for each, depending on the previous knowledge or on the results of model fitting, you may shorten the steps as well. We also presented the case where additive and multiplicative models work pretty well and how to modify the statistic in this case. Likewise the application and variation of multiscale network test is almost limitless.

A appendix

A.1 simulation schemes

- **Two Block SBM**

$$\begin{aligned}
&X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Bern(0.5), \quad i = 1, \dots, n \\
&Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Bern(0.6)I(x = 0) + Bern(0.4)I(x = 1), \quad i = 1, \dots, n \\
&A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \overset{d}{=} Bern(0.4)I(|z_i - z_j| = 0) + Bern(0.1)I(|z_i - z_j| > 0) \\
&\qquad i, j = 1, \dots n; i < j.
\end{aligned} \tag{25}$$

- **Three Block SBM 1**

$$\begin{aligned}
&X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Multi(1/3, 1/3, 1/3), i = 1, \dots, n \\
&Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Multi(0.5, 0.25, 0.25)I(x = 1) + Multi(0.25, 0.5, 0.25)I(x = 2) \\
&\qquad + Multi(0.25, 0.25, 0.5)I(x = 3), \quad i = 1, \dots, n \\
&A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \overset{d}{=} Bern(0.5)I(|z_i - z_j| = 0) + Bern(0.2)I(|z_i - z_j| = 1) \\
&\qquad + Bern(0.3)I(|z_i - z_j| = 2), \quad i, j = 1, \dots n; i < j.
\end{aligned} \tag{26}$$

- **Three Block SBM 2**

$$
\begin{aligned}
&X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Multi(1/3, 1/3, 1/3), \quad i = 1, \ldots, n \\
&Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Multi(0.5, 0.25, 0.25)I(x = 1) + Multi(0.25, 0.5, 0.25)I(x = 2) \\
&\qquad + Multi(0.25, 0.25, 0.5), \quad i = 1, \ldots, n \\
&A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \overset{d}{=} Bern(0.5)I(|z_i - z_j| = 0) \\
&\qquad + Bern(0.2)I(|z_i - z_j| > 1) \quad i, j = 1, \ldots n; i < j \\
&A_{ji} = A_{ij}; A_{ii} = 0, \quad i, j = 1, \ldots n
\end{aligned}
\tag{27}
$$

- **Degree-corrected SBM**

$$
\begin{aligned}
&\theta_i \overset{i.i.d}{\sim} Uniform(0, 2), i = 1, \ldots, n \\
&A_{ij}|\mathbf{Z}, \theta \overset{i.i.d}{\sim} f_{A|Z,\theta}(a_{ij}|z_i, z_j, \theta_i, \theta_j) \overset{d}{=} Bern(0.2 \cdot \theta_i\theta_j)I(|z_i - z_j| = 0) \\
&\qquad + Bern(0.05 \cdot \theta_i\theta_j)I(|z_i - z_j| = 1), \quad i, j = 1, \ldots, n; i < j.
\end{aligned}
\tag{28}
$$

- **Additive and Multiplicative Graph Model**

$$
\begin{aligned}
&Z_i \overset{i.i.d}{\sim} f_Z(z) \overset{d}{=} Uniform[0, 1]. \quad i = 1, \ldots, n \\
&X_i|Z_i \overset{i.i.d}{\sim} f_{X|Z}(x|z) \overset{d}{=} Normal(Z_i, 1), \quad i = 1, \ldots, n \\
&A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \overset{d}{=} Bern\big((1 - z_i)^2 \times (1 - z_j)^2\big), \quad i, j = 1, \ldots, n; i < j.
\end{aligned}
\tag{29}
$$

A.2 supplementary figures

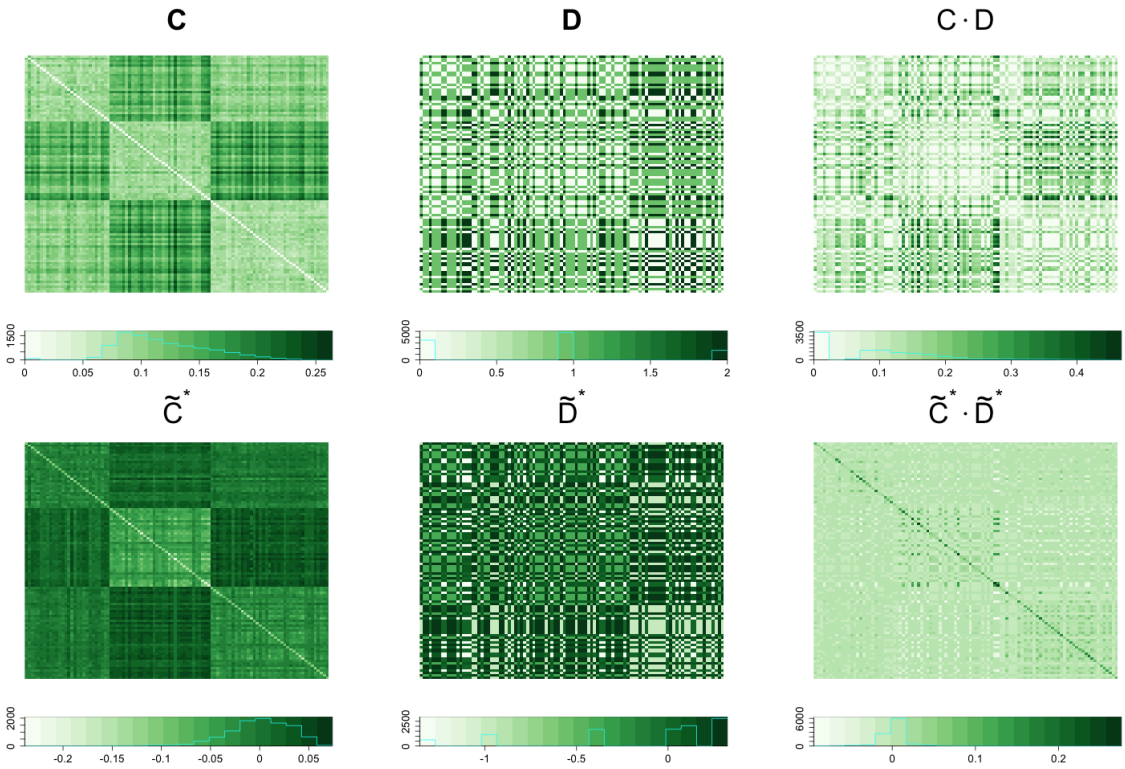

Figure 11: (a) Top : Euclidean distance of diffusion maps at time $t = 3$, C; Euclidean distance of X, D; element-wise product of C and D. (b) Bottom : truncated double-centered $\tilde{C}$ by k^*th nearest neighbor in C; truncated double-centered $\tilde{D}$ by l^*th nearest neighbor in D; element-wise product of two truncated matrices $\tilde{C}^*$ and $\tilde{D}^*$.

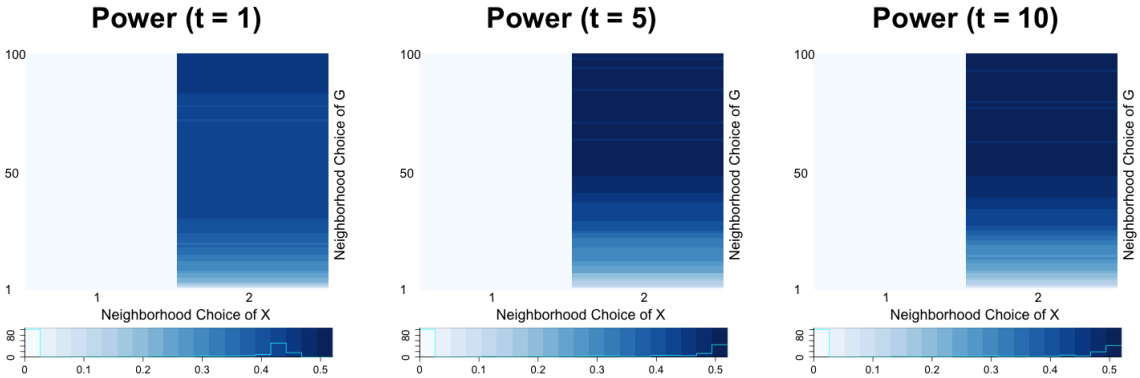

Figure 12: Empirical power heatmap at diffusion time point $t = 1$(left), $t = 5$(middle), and $t = 10$(right) based on $M = 500$ independently generated SBM presented in model 25

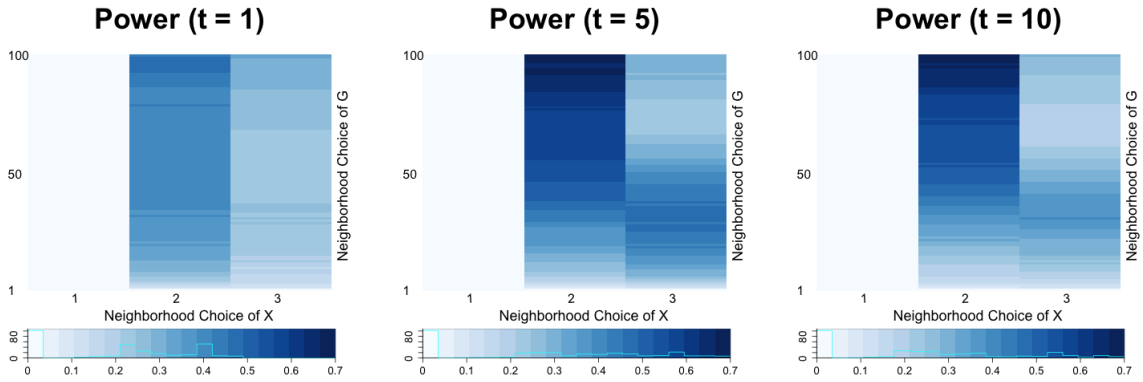

Figure 13: Empirical power heatmap at diffusion time point $t = 1$(left), $t = 5$(middle), and $t = 10$(right) based on $M = 500$ independently generated SBM presented in model 26

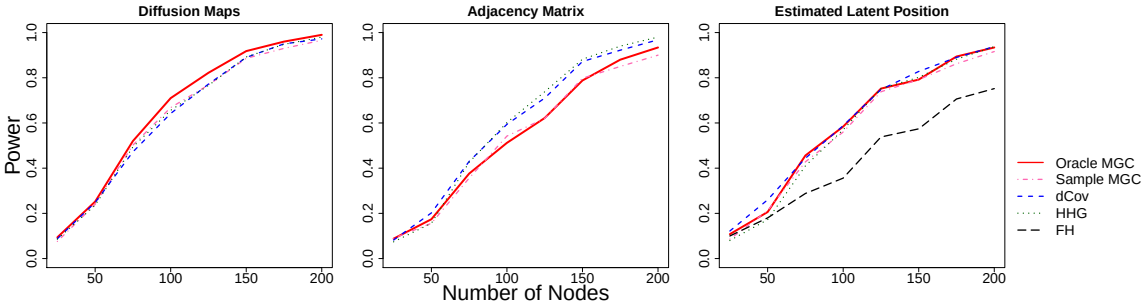

Figure 14: Empirical power based on $M = 500$ of three SBM under linear dependence using diffusion maps (left), Euclidean of adjacency matrix (middle) and estimated latent position(right). The most right figure contains the results of `FH` test as well.

A.3 Node Contribution

$$
\begin{aligned}
&X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Bern(0.5) \quad i = 1, \ldots, 100, \ldots, 200 \\
&Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Bern(0.6)I(x = 0) + Bern(0.4)I(x = 1), \quad i = 1, \ldots, 100, \ldots, n \\
&A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \\
&\qquad \overset{d}{=} \begin{cases} Bern(0.4)I(|z_i - z_j| = 0) + Bern(0.1)I(|z_i - z_j| > 0) & i = 1, \ldots, n/2 \\ Bern(0.25) & i = 1 + n/2, \ldots, n \end{cases}
\end{aligned} \tag{30}
$$

A.4 Monotonicity and Power

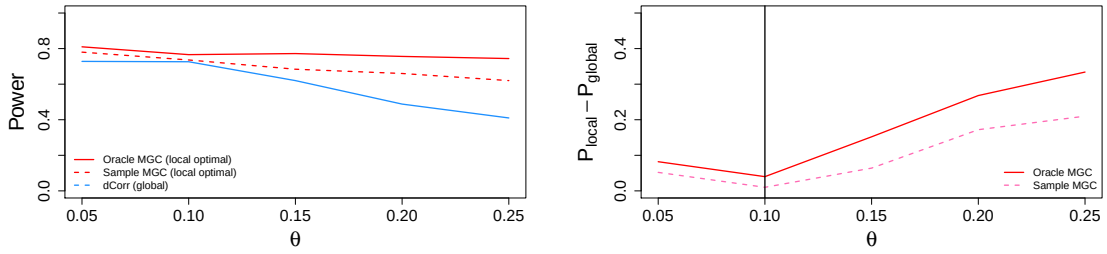

Figure 15: Change of empirical power across θ in $Power(\theta) = E(A_{ij}|X_i, X_j) = 0.5I(|X_i - X_j| = 0) + 0.1I(|X_i - X_j| = 1) + \theta I(|X_i - X_j| = 2)$. Superiority of optimal local scale become evident from $\theta > 0.1$, when distribution of edges have non-linear dependence on X.

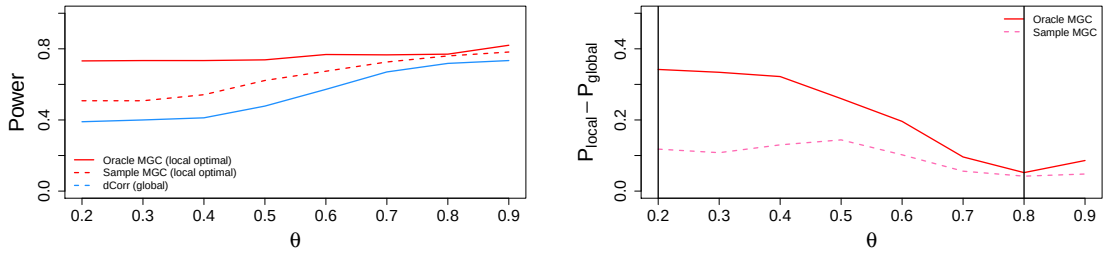

Figure 16: Change of empirical power across θ in $Power(\theta) = E(A_{ij}|X_i, X_j) = 0.2I(|X_i - X_j| = 0) + 0.8I(|X_i - X_j| = 1) + \theta I(|X_i - X_j| = 2)$. Superiority of optimal local scale become evident from $\theta < 0.8$, when distribution of edges have non-linear dependence on X.

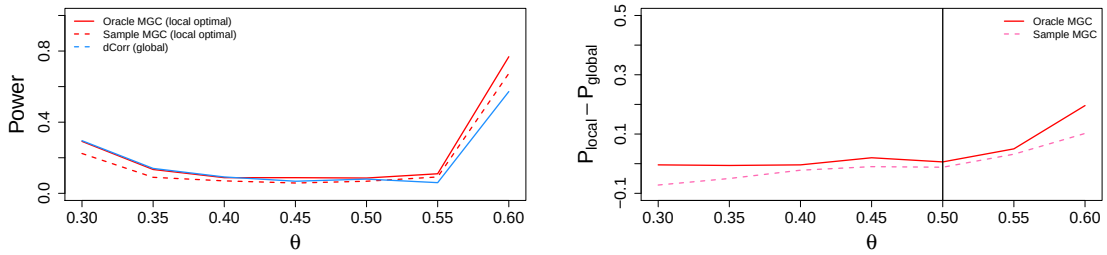

Figure 17: Change of empirical power across θ in $Power(\theta) = E(A_{ij}|X_i, X_j) = 0.5I(|X_i - X_j| = 0) + 0.5I(|X_i - X_j| = 1) + \theta I(|X_i - X_j| = 2)$. You can see that across $\theta(> 0)$, `MGC` and global distance-based tests are very similar since we have linear-dependence for all range or θ.

A.5 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.

function DMAP ($n \times n$ transition matrix P, time points $\{t_1, t_2, \ldots, t_K\}$)
  $\pi :=$ `statdistr`(P) ▷ stationary distribution of P
  $\Pi :=$ `Diag`(π) ▷ Diagonal matrix with diagonal element of π
  $Q := \Pi^{1/2} P \Pi^{-1/2}$
  $\lambda :=$ `eigenvalue`(Q) ▷ a real-valued vector with length of $q(\leq n)$.
  $\Lambda :=$ `Diag`(λ)
  $\Psi :=$ `eigenfunction`(Q) ▷ $n \times q$ real-valued matrix
  $\Phi := \Pi^{-1/2}\Psi$ ▷ $n \times q$ real-valued eigenfunction matrix of P
  for $t_i : i = 1$ **do** K
    `Maps`$[i] := \Phi\Lambda^{t_i}$
  end for
  `Maps` = list(`Maps`[1], `Maps`[2], ..., `Maps`$[K]$)
  return `Maps`
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.

function NETWORKTEST (G, $\mathbf{X}$, $\mathbf{T}$:= (diffusion time points $\{t_1, t_2, \ldots, t_K\}$))
  $A :=$ `get.adjacency`(G) ▷ obtain an adjacency matrix of network G
  $P := A$ / `rowSums`(A)
  $U :=$ `dmap`$(P, \mathbf{T})$ ▷ a list of diffusion maps in each time point
  for $t_i : i = 1$ **do** K
    $C :=$ `dist`$(U[i])$ ▷ distance matrix of diffusion maps at time t_i
    $D :=$ `dist`(X) ▷ distance matrix of nodal attributes
    `MGC`$[i]$ = `MGCPermutationTest`(C, D)
  end for
  `MGC` = list(`MGC`[1], `MGC`[2], ..., `MGC`$[K]$)
  return `MGC`
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 $\{c(v)\}$

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

A.6 Lemmas and Theorems

Theorem A.1 (de Finetti's Theorem)**.** 1. Let $X_1, X_2, ...$ be an infinite sequence of random variables with values in a space of $\mathbf{X}$. The sequence $X_1, X_2, ...$ 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.

2. If the sequence is exchangeable, the empirical distributions

$$\hat{S}_n(.) := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(.), n \in \mathbb{N}$$

converges to η as $n \to \infty$ with probability 1.

Theorem A.2 (Aldous Hoover Theorem)**.** Let $\mathbf{A} = \{A_{ij}\}, 1 \leq i, j \leq \infty$ 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

$$\big(A_{ij}\big) \stackrel{d}{=} \big(f\big(U_i, U_j, U_{ij}\big)\big)\,, \tag{31}$$

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 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

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

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 $i, j = 1, ..., n; i < j$:

$$\begin{aligned} P\big(A_{ij} = a_{ij}\big) &= \int P\big(A_{ij}\big|w_i, w_j\big) Pr(W_i = w_i) Pr(W_j = w_j) dw_i dw_j \\ &= \int_0^1 \int_0^1 g(w_i, w_j)^{a_{ij}} \big(1 - g(w_i, w_j)\big)^{1-a_{ij}} dw_i dw_j \end{aligned} \tag{33}$$

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

□

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 Eq. 12. 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 $\{\lambda_r\}_{r=\{1,2,\ldots,q\}}$ and a set of their corresponding orthonormal eigenvectors $\{\psi_r\}_{r=\{1,2,\ldots,q\}}$, i.e. $Q[i,j] = \sum_{r=1}^{q} \lambda_r \psi_r(i) \psi_r(j)$ $(1 \leq q \leq n)$. Returning to the transition probability between Node i and Node j,

$$\begin{aligned} P[i,j] &= \sqrt{\pi(j)/\pi(i)}\, Q[i,j] \\ &= \sum_{r=1}^{q} \lambda_r \big\{ \psi_r(i)/\sqrt{\pi(i)} \big\} \big\{ \psi_r(j)\sqrt{\pi(j)} \big\} \\ &:= \sum_{r=1}^{q} \lambda_r \phi_r(i) \big\{ \psi_r(j)\sqrt{\pi(j)} \big\} \end{aligned} \tag{34}$$

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:

$$C_t^2[i,j] = \sum_{r=1}^{n} \lambda_r^{2t} \big(\phi_r(i) - \phi_r(j) \big)^2 \tag{35}$$

That is,

$$C_t[i,j] = \| \, \mathbf{U}_t(i) - \mathbf{U}_t(j) \, \| \tag{36}$$

where

$$\mathbf{U}_t(i) = \begin{pmatrix} \lambda_1^t \phi_1(i) & \lambda_2^t \phi_2(i) & \ldots & \lambda_q^t \phi_q(i) \end{pmatrix}^T \in \mathbb{R}^q. \tag{37}$$

□

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 :

$$\mathbf{U}_t(i) = \begin{pmatrix} \lambda_1^t\phi_1(i) & \lambda_2^t\phi_2(i) & \cdots & \lambda_q^t\phi_q(i) \end{pmatrix} \in \mathbb{R}^q. \tag{38}$$

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. $(A_{ij}) \overset{d}{=} \left(A_{\sigma(i)\sigma(j)}\right)$. 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 $i \neq j; i, j = 1, 2, \ldots, n$:

$$\left(P_{ij}\right) = \left(\frac{A_{ij}}{1 - A_{ij} + \sum_{j=1}^{n} A_{ij}}\right) \overset{d}{=} \left(\frac{A_{\sigma(i)\sigma(j)}}{1 - A_{\sigma(i)\sigma(j)} + \sum_{\sigma(j)=1}^{n} A_{\sigma(i)\sigma(j)}}\right) = \left(P_{\sigma(i)\sigma(j)}\right) \tag{39}$$

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. □

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 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

$$(A_{ij}) \overset{d}{=} (A_{v_i,v_j}) \overset{d}{=} Bern(g(\vartheta_i, \vartheta_j)), \quad i,j = 1,\ldots,n; i < j, \tag{40}$$

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 size graphs can 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 $i,j = 1,...,n; i < j$:

$$\begin{aligned} P\big(A_{ij} = a_{ij} \big| V_i, V_j\big) &= \int P\big(A_{ij} \big| v_i, v_j\big) f_\vartheta(\vartheta_i) f_\vartheta(\vartheta_j) d\vartheta_i d\vartheta_j \\ &= \int_0^\nu \int_0^\nu g(\vartheta_i,\vartheta_j)^{a_{ij}} \big(1 - g(\vartheta_i,\vartheta_j)\big)^{1-a_{ij}} \cdot dPois_1(\vartheta_i) \cdot dPois_1(\vartheta_j) d\vartheta_i d\vartheta_j. \end{aligned} \tag{41}$$

□

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.

Proof of corollary 2.3.1. [Triangle inequality of diffusion distance] Let $x, y, z \in V(G)$.

$$\begin{aligned} D_t^2(x,z) &= \sum_{w \in V(G)} \big(P^t(x,w) - P^t(z,w)\big)^2 \frac{1}{\pi(w)} \\ &= \sum_{w \in V(G)} \big(P^t(x,w) - P^t(y,w) + P^t(y,w) - P^t(z,w)\big)^2 \frac{1}{\pi(w)} \\ &= \sum_{w \in V(G)} \big(P^t(x,w) - P^t(y,w)\big)^2 \frac{1}{\pi(w)} + \sum_{w \in V(G)} \big(P^t(y,w) - P^t(z,w)\big)^2 \frac{1}{\pi(w)} \\ &\quad + 2 \sum_{w \in V(G)} \big(P^t(x,w) - P^t(y,w)\big)\big(P^t(y,w) - P^t(z,w)\big) \frac{1}{\pi(w)} \\ &= D_t^2(x,y) + D_t^2(y,z) + 2 \sum_{w \in V(G)} \big(P^t(x,w) - P^t(y,w)\big)\big(P^t(y,w) - P^t(z,w)\big) \frac{1}{\pi(w)} \end{aligned} \tag{42}$$

Thus it suffices to show that

$$\sum_{w\in V(G)} \big(P^t(x,w) - P^t(y,w)\big)\big(P^t(y,w) - P^t(z,w)\big)\frac{1}{\pi(w)} \leq D_t(x,y)\cdot D_t(y,z). \tag{43}$$

Let $a_w = \big(P^t(x,w) - P^t(y,w)\big)\sqrt{1/\pi(w)}$ and $b_w = \big(P^t(y,w) - P^t(z,w)\big)\sqrt{1/\pi(w)}$. Then the above inequality is equivalent to :

$$\sum_{w\in V(G)} a_w \cdot b_w \leq \sqrt{\sum_{w\in V(G)} a_w^2 \cdot \sum_{w\in V(G)} b_w^2}. \tag{44}$$

which is true by Cauchy-Schwarz inequality. □

Proof of Lemma 2.4 *Convergence of empirical characteristic function 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 $\{(\mathbf{x}_j, \mathbf{y}_j), j = 1, 2, ..., n\}$, e.g., it holds without assuming exchangeability, nor identically distributed, nor finite second moments. □

Proof of Lemma 2.5 *Empirical characteristic function of exchangeable variables.* It suffices to prove the first argument 16 since the second argument 17 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 $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ to be independently identically distributed as $(\mathbf{x}, \mathbf{y})$ with finite second moments; here we have exchangeable $\{(\mathbf{x}_i, \mathbf{y}_i)\}$ instead.

Followied by *de Finetti's Theorem* A.1, if and only if $\{x_i\}$ are (infinitely) exchangeble, there exist an underlying distribution $f_{\mathbf{x}}$ of $\mathbf{x}$ such that $\mathbf{x}_i \overset{i.i.d}{\sim} f_{\mathbf{x}}$. 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 $(\mathbf{x}_i, \mathbf{y}_i) \overset{i.i.d}{\sim} f_{\mathbf{x},\mathbf{y}}$. 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.,

$$\int_{D(\delta)} \|g^n_{\mathbf{x},\mathbf{y}}(t,s) - g^n_{\mathbf{x}}(t)g^n_{\mathbf{y}}(s)\|^2 dw \overset{n\to\infty}{\longrightarrow} \int_{D(\delta)} \|g_{\mathbf{x},\mathbf{y}}(t,s) - g_{\mathbf{x}}(t)g_{\mathbf{y}}(s)\|^2 dw, \tag{45}$$

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

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 $\mathcal{V}^2_n(\mathbf{X},\mathbf{Y}) \overset{n\to\infty}{\longrightarrow} 0$ if and only if underlying distribution of $\{\mathbf{x}_i\}$, $\mathbf{x}$ 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 (see in [Cencheng]), `MGC` statistic constructed by `dCorr` or `mCorr` is also consistent in testing dependence. □

Proof of Theorem 2.7 *Consistency of* `MGC` *applied to exchangeable variables.* Suppose that we have undirected, connected network $\mathbf{G}$ with a family of diffusion maps $\{\mathbf{u}_t\}$ and with nodal attributes $\{\mathbf{x}\}$. 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 $\mathbf{u} \overset{i.i.d}{\sim} f_{\mathbf{u}^{(t)}}$ for $t = 1, 2, \ldots$; and we have $\mathbf{x}_i \overset{i.i.d}{\sim} f_{\mathbf{X}}$. Under the assumption of finite second moment of $\mathbf{u}^{(t)}$ and $\mathbf{x}$, `MGC` statistics constructed by $\{(\mathbf{u}_{ti}, \mathbf{x}_i) : i = 1, 2, \ldots, n\}$ 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 conditioning 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. $A_{ij} \overset{i.i.d}{\sim} f_A$ for all $i, j = 1, 2, \ldots, n$; otherwise, each column is dependent on one another. □

References

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144), 1236498.

Barabasi, A.-L. and Z. N. Oltvai (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics 5*(2), 101–113.

Bartlett, T. E., S. C. Olhede, and A. Zaikin (2014). A dna methylation network interaction measure, and detection of network oncomarkers. *PloS one 9*(1), e84573.

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.

Ellison, N. B., C. Steinfield, and C. Lampe (2007). The benefits of facebook friends: social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication 12*(4), 1143–1168.

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.

Gross, R. and A. Acquisti (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pp. 71–80. ACM.

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.

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.

Mantzaris, A. V., D. S. Bassett, N. F. Wymbs, E. Estrada, M. A. Porter, P. J. Mucha, S. T. Grafton, and D. J. Higham (2013). Dynamic network centrality summarizes learning in the human brain. *Journal of Complex Networks 1*(1), 83–92.

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.

Palla, K., D. Knowles, and Z. Ghahramani (2012). An infinite latent attribute model for network data. *arXiv preprint arXiv:1206.6416*.

Pinquart, M. and S. Sörensen (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. *Psychology and aging 15*(2), 187.

Pujol, A., R. Mosca, J. Farrés, and P. Aloy (2010). Unveiling the role of network and systems biology in drug discovery. *Trends in pharmacological sciences 31*(3), 115–123.

Raftery, A. E., X. Niu, P. D. Hoff, and K. Y. Yeung (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics 21*(4), 901–919.

Sporns, O., C. J. Honey, and R. Kötter (2007). Identification and classification of hubs in brain networks. *PloS one 2*(10), e1049.

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.

Tang, M. and M. Trosset (2010). Graph metrics and dimension reduction. *Indiana University, Indianapolis, IN*.

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.

SUPPLEMENTARY MATERIAL

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