Biometrika (2019), **106**, 4, *pp.* 857–873
Printed in Great Britain
doi: 10.1093/biomet/asz045
Advance Access publication 30 September 2019

Network dependence testing via diffusion maps and distance-based correlations

BY YOUJIN LEE

Center for Causal Inference, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.
youjin.lee@pennmedicine.upenn.edu

CENCHENG SHEN

Department of Applied Economics and Statistics, University of Delaware, Newark, Delaware 19716, U.S.A.
shenc@udel.edu

CAREY E. PRIEBE

Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, U.S.A.
cep@jhu.edu

AND JOSHUA T. VOGELSTEIN

Department of Biomedical Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, U.S.A.
jovo@jhu.edu

SUMMARY

Deciphering the associations between network connectivity and nodal attributes is one of the core problems in network science. The dependency structure and high dimensionality of networks pose unique challenges to traditional dependency tests in terms of theoretical guarantees and empirical performance. We propose an approach to test network dependence via diffusion maps and distance-based correlations. We prove that the new method yields a consistent test statistic under mild distributional assumptions on the graph structure, and demonstrate that it is able to efficiently identify the most informative graph embedding with respect to the diffusion time. The methodology is illustrated on both simulated and real data.

Some key words: Adjacency spectral embedding; Diffusion distance; Multiscale graph correlation; Normalized graph Laplacian.

1. INTRODUCTION

Network data has seen increased availability and influence in statistics, physics, computer science, biology and social science, and poses many challenges due to its distinct structure. A network or graph is formally defined as an ordered pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of nodes and $\mathcal{E}$ is the set of edges. Let $n = |\mathcal{V}|$. The edge connectivity of a graph can be

compactly represented by the adjacency matrix $\mathcal{A} = \{\mathcal{A}(i,j) : i,j = 1,\ldots,n\}$, where $\mathcal{A}(i,j)$ is the edge weight between node i and node j. For example, for an unweighted and undirected network, $\mathcal{A}(i,j) = \mathcal{A}(j,i) = 1$ if and only if node i and node j are connected by an edge, and zero otherwise. Often, each node has some associated nodal attributes, which we denote as $X_i \in \mathbb{R}^p$ and use $\mathcal{X} = [X_1|\cdots|X_n]$ to represent the collection of attributes.

This paper focuses on independence testing between network connectivity and nodal attributes. Assuming that, for the adjacency matrix $\mathcal{A}$ and attributes $\mathcal{X}$, the connectivity and attribute corresponding to each node are identically and jointly distributed as F_{AX}, the null (1) and alternative hypotheses of interest are:

$$\begin{aligned} H_0 &: F_{AX} = F_A F_X, \\ H_A &: F_{AX} \neq F_A F_X. \end{aligned} \tag{1}$$

There are many network data examples where testing independence can be a crucial first step. For example, determining potential correlation between cultural tastes and relationships over a social network (Lewis et al., 2012), identifying associations between the strength of functional connectivity and brain physiology such as regional cerebral blood flow in the brain network (Liang et al., 2013), and embedding text data and its hyperlink networks jointly into a low-dimensional structure (Shen et al., 2017). Sometimes the correlations among nodes are not proportional to the strength of connectivity between them. For instance, in the signalling network of biological cells, the reaction rate for each cell exhibits a nonlinear dependence on the neighbouring response due to the complex, cooperative biological processes involved (Hernandez-Hernandez et al., 2017). We can observe nonlinear dependence in concentrated propagation among a few focal people in a social network (Nekovee et al., 2007), and in screening informative brain regions for sex and site difference from fMRI image graphs (Wang et al., 2018).

A notable obstacle in network inference is the structure of the edge connectivity. Namely, for an undirected graph, $\mathcal{A}$ is a symmetric binary matrix whose edges are not independent of each other, thus preventing many well-established methods from being directly applicable. One approach is to assume a certain model for the graph structure, then solve the inference question based on the model assumption (Wasserman & Pattison, 1996; Fosdick & Hoff, 2015; kim et al., 2016). Another approach is spectral embedding, which first embeds the $n \times n$ adjacency matrix $\mathcal{A}$ into an $n \times q$ matrix $\mathcal{U}$ by eigendecomposition, then directly works on $\mathcal{U}$ (Rohe et al., 2011; Sussman et al., 2012; Tang et al., 2017). For example, the network dependence test proposed by Fosdick & Hoff (2015) assumes that the adjacency matrix is generated from a multivariate normal distribution of the latent factors, and estimates the latent factor associated with each node from $\mathcal{A}$, followed by applying the standard likelihood ratio test on the normal distribution.

However, model-based approaches are often limited by, and do not perform well beyond, the model assumptions. Moreover, spectral embedding is susceptible to misspecification of the dimension of q. Both of these factors can significantly degrade the later inference performance. Indeed, as a ground truth is unlikely in real networks (Peel et al., 2017), one often desires a method that is effectively nonparametric and robust against algorithm parameter selection (Chen et al., 2016).

We propose a methodology to test network dependency of Equation (1) via diffusion maps and distance-based correlations, which is universally consistent under mild graph distributional assumptions and works well under many popular network models. The proposed method also overcomes parameter selection issues and exhibits superior empirical testing performance.

2. Preliminaries

2.1. *Notation*

We denote a random variable by an upper-case letter such as X with distribution F_X, and a matrix or a set of vectors by an italic letter such as $\mathcal{X}$. For each node $i \in \mathcal{V}$, its attribute is denoted by X_i whose realizations are in $\mathbb{R}^p$, and its edge connectivity vector is denoted by $A_i \in \mathbb{R}^n$, which is a column in the $n \times n$ adjacency matrix $\mathcal{A}$. We assume that $(X_i, A_i) \sim F_{XA}$, i.e., identically distributed attributes and connectivity vectors. Later we introduce a multiscale node-wise representation of the nodes as an $n \times q$ matrix $\mathcal{U}^t = [U_1^t | U_2^t | \cdots | U_n^t]$ for any $t \in \{0\} \bigcup \mathbb{Z}^+$, where q is the embedding dimension and t is the Markov iteration time step. Let $\cdot^*$ denote estimated optimality; $\cdot^t$ denotes either the tth power or time step, which shall be clear in the context; and $\cdot^{\mathrm{T}}$ is the matrix transpose.

2.2. *Diffusion maps*

Because the rows and columns of a symmetric adjacency matrix may be correlated, directly operating on the adjacency matrix breaks theoretical guarantees of existing dependence tests. The diffusion map was introduced as a feature extraction algorithm by Coifman et al. (2005), Coifman & Lafon (2006) and Lafon & Lee (2006), which computes a family of embeddings in Euclidean space by eigendecomposition on a diffusion operator of the given data. Here we introduce a version tailored to adjacency matrices.

To derive the diffusion maps for given observations of size n, the first step is to choose an $n \times n$ kernel matrix $\mathcal{K}$ that represents the similarity within the sample data. The adjacency matrix $\mathcal{A}$ is a natural similarity matrix; for undirected graphs we let $\mathcal{K} = \mathcal{A}$, for directed graphs we let $\mathcal{K} = (\mathcal{A} + \mathcal{A}^{\mathrm{T}})/2$. The next step is to compute the normalized Laplacian matrix by

$$\mathcal{L} = \mathcal{B}^{-1/2} \mathcal{K} \mathcal{B}^{-1/2},$$

where $\mathcal{B}$ is the $n \times n$ degree matrix of $\mathcal{K}$. When $\mathcal{B}(i,i)$ or $\mathcal{B}(j,j)$ is zero, $\mathcal{L}(i,j) = 0$.

The diffusion map $\mathcal{U}^t = \{U_i^t \in \mathbb{R}^q : i = 1, \ldots, n\}$ is then computed by eigendecomposition,

$$U_i^t = \left(\lambda_1^t \phi_{i1}, \quad \lambda_2^t \phi_{i2}, \quad \ldots, \quad \lambda_q^t \phi_{iq}\right)^{\mathrm{T}} \in \mathbb{R}^q \quad (i = 1, \ldots, n), \tag{2}$$

where $\{\lambda_j^t : j = 1, 2, \ldots, q\}$ and $\{\phi_j \in \mathbb{R}^n : (\phi_{1j}, \phi_{2j}, \ldots, \phi_{nj}),\ j = 1, 2, \ldots, q\}$ are the q largest eigenvalues and corresponding eigenvectors of $\mathcal{L}$, respectively, and λ_j^t is the tth power of the jth eigenvalue. The diffusion distance between the ith observation and the jth observation is defined as the weighted ℓ^2 distance of the two points in the observation space, which equals the Euclidean distance in the diffusion coordinate:

$$\mathcal{C}^t(i,j) = \|U_i^t - U_j^t\| \quad (i,j = 1, 2, \ldots, n),$$

where $\|\cdot\|$ is the Euclidean distance.

When $t = 0$, the diffusion map is exactly the same as a normalized graph Laplacian embedding in Rohe et al. (2011) up to a linear transformation, and the diffusion maps are weighted graph Laplacian by powered eigenvalues (Lafon & Lee, 2006). The diffusion map at $t = 1$ equals the adjacency spectral embedding up to the degree constant (Sussman et al., 2014). Therefore, the diffusion maps can be viewed as a single-index family of embeddings. The embedding dimension choice q can be selected via the profile likelihood method in Zhu & Ghodsi (2006), which is a standard algorithm in dimension reduction literature. To select the optimal t, we will utilize a smoothing technique to maximize the dependency, as discussed shortly.

2.3. *Distance-based correlations*

The problem of testing general dependencies between two random variables has seen notable progress in recent years. Pearson's correlation (Pearson, 1895) is the most classical approach, and determines the existence of a linear relationship via a correlation coefficient in the range $[-1, 1]$, with 0 indicating no linear association and ± 1 indicating perfect linear association. To better capture dependencies not limited to a linear relationship, a variety of distance-based correlation measures have been suggested, including the distance correlation and energy statistic (Székely et al., 2007; Székely & Rizzo, 2013; Rizzo & Székely, 2016), the kernel-based independence test (Gretton & Gyorfi, 2010), the Heller–Heller–Gorfine test (Heller et al., 2013, 2016) and multiscale graph correlation (Shen et al., 2018; Vogelstein et al., 2019), among others. In particular, distance correlation is a distance-based dependency measure that is consistent against all possible dependencies with finite second moments. The kernel independence test is a kernel variant of distance correlation (Sejdinovic et al., 2013; Shen & Vogelstein, 2018). The multiscale graph correlation inherits the same consistency of distance correlation with better finite-sample testing powers under high-dimensional and nonlinear dependencies, via defining a family of local correlations and efficiently searching for the optimal local scale in testing. Here we briefly introduce distance correlation and multiscale graph correlation.

Given n pairs of sample data that are independently and identically distributed, namely $(\mathcal{U}, \mathcal{X}) = \{(U_i, X_i) \overset{i.i.d.}{\sim} F_{UX} \in \mathbb{R}^q \times \mathbb{R}^p : i = 1, 2, \ldots, n\}$. Denote the pairwise distances within $\{U_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ as $\mathcal{C}(i,j) = \|U_i - U_j\|$ and $\mathcal{D}(i,j) = \|X_i - X_j\|$ for $i, j = 1, 2, \ldots, n$, respectively. The sample distance covariance is denoted as

$$\text{DCOV}_n(\mathcal{U}, \mathcal{X}) = \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{\mathcal{C}}(i,j)\tilde{\mathcal{D}}(i,j),$$

where $\tilde{\mathcal{C}} = \mathcal{H}\mathcal{C}\mathcal{H}$ and $\tilde{\mathcal{D}} = \mathcal{H}\mathcal{D}\mathcal{H}$, and $\mathcal{H} = \mathcal{I}_{n\times n} - \mathcal{J}_{n\times n}/n$ is the centring matrix with $\mathcal{I}_{n\times n}$ being the $n \times n$ identity matrix and $\mathcal{J}_{n\times n}$ being the $n \times n$ matrix of all ones. The distance correlation follows by normalizing distance covariance via Cauchy–Schwarz into the range $[-1, 1]$. Székely et al. (2007) shows that sample distance correlation converges to a population form, which is asymptotically 0 if U and X are independent, resulting in a consistent statistic for testing independence. An unbiased sample version of distance correlation was later proposed to eliminate the sample bias in distance correlation (Székely & Rizzo, 2013, 2014), and this is the default implementation in this paper.

The multiscale graph correlation is an optimal local version of distance correlation that improves its finite-sample testing power. It first derives all local distance covariances as

$$\text{DCOV}_n^{kl}(\mathcal{U}, \mathcal{X}) = \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{\mathcal{C}}^k(i,j)\tilde{\mathcal{D}}^l(i,j) \quad (k = 1, \ldots, \kappa,\ l = 1, \ldots, \gamma),$$

where κ and γ are the number of unique numerical values in $\mathcal{C}$ and $\mathcal{D}$, respectively; $\tilde{\mathcal{C}}^k(i,j) = \tilde{\mathcal{C}}(i,j)\mathbb{I}(R_{ij}^{\mathcal{C}} \leqslant k)$; $\mathbb{I}(\cdot)$ is the indicator function; and $R_{ij}^{\mathcal{C}}$ is a rank function of U_i relative to U_j, i.e., $R_{ij}^{\mathcal{C}} = k$ if U_i is the kth nearest neighbour of U_j, and define equivalently $\tilde{\mathcal{D}}^l(i,j) = \tilde{\mathcal{D}}(i,j)\mathbb{I}(R_{ij}^{\mathcal{D}} \leqslant l)$ for $\{X_i\}$. Then the local distance correlations $\{\text{DCOR}^{kl}\}$ are the normalizations of the local distance covariances into $[-1, 1]$ via Cauchy–Schwarz. Among all possible neighbourhood choices, the multiscale graph correlation equals the maximum local correlation within

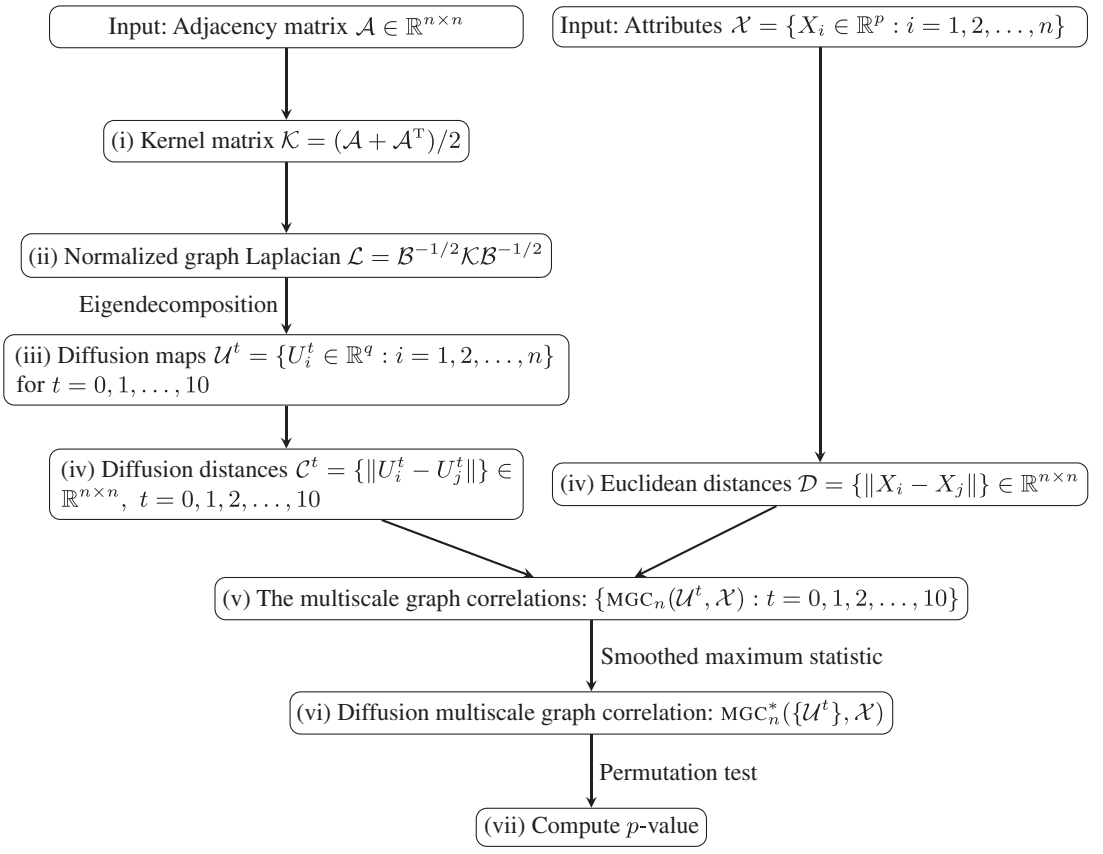

Fig. 1. Flowchart for network dependence testing via diffusion multiscale graph correlation.

the largest connected component of significant local correlations, i.e.,

$$\mathrm{MGC}_n(\mathcal{U}, \mathcal{X}) = \mathrm{DCOR}_n^{(kl)^*}(\mathcal{U}, \mathcal{X}), \text{ where } (kl)^* = \arg\max_{(kl)} \mathcal{S}(\mathrm{DCOR}_n^{kl})$$

for a smoothing operation $\mathcal{S}(\cdot)$ that filters out all insignificant local correlations. The multiscale graph correlation has been shown to have power almost equal to or better than distance correlation throughout a wide variety of common dependencies, while being computationally efficient (Shen et al., 2018).

3. Main results

3.1. *Testing procedure of diffusion correlation*

Here we develop diffusion multiscale graph correlation, which synthesizes diffusion map embedding, multiscale graph correlation and smoothed maximum statistic to better test network dependency. A flowchart of the testing procedure is illustrated in Fig. 1, and the details of each step are described in Algorithm 1.

The algorithm is flexible in the choice of correlation measures: by following the exact same steps, but replacing the multiscale graph correlation by distance correlation in step (v), one can compute the diffusion distance correlation. Similarly, one can derive the diffusion Heller–Heller–Gorfine method. The motivation of the smoothing step (vi) is the following: suppose there exists

an optimal t for detecting the relationship between edge connectivity and attributes, then the test statistics at adjacent time steps $t-1$ and $t+1$ should also exhibit strong signals. Under independence, a large test statistic at certain t can occur by chance and cause a direct maximum statistic to have a low testing power, while the smoothed maximum statistic effectively filters out any noisy and isolated large test statistic. In practice, it suffices to consider $t \in [0, 1, \ldots, 10]$ or an even smaller upper bound like 3 or 5. When a smoothed maximum statistic does not exist, we set $t = 3$ as the default choice. The permutation test in step (vii) is a common nonparametric procedure used for real data testing in almost all dependency measures, which is valid as long as the observations are exchangeable under the null (Rizzo & Székely, 2016).

Algorithm 1. Testing procedure of diffusion correlation.

Input: Adjacency matrix $\mathcal{A} \in \mathbb{R}^{n \times n}$ and nodal attributes $\mathcal{X} = \{X_i \in \mathbb{R}^p : i = 1, 2, \ldots, n\}$.

(i) Symmetrize $\mathcal{A}$ by $\mathcal{K} = (\mathcal{A} + \mathcal{A}^{\mathrm{T}})/2$.
(ii) Obtain normalized graph Laplacian matrix $\mathcal{L} = \mathcal{B}^{-1/2}\mathcal{K}\mathcal{B}^{-1/2}$.
(iii) Do eigendecomposition to obtain diffusion maps $\mathcal{U}^t = \{U_1^t, U_2^t, \ldots, U_n^t\}$ for $t = 0, 1, 2, \ldots, 10$.
(iv) Derive $n \times n$ Euclidean distance of diffusion map $\mathcal{C}^t$, i.e., diffusion distance, across t, and $n \times n$ Euclidean distance of nodal attributes, $\mathcal{D}$.
(v) Compute the multiscale graph correlations using two distance matrices, $\mathcal{C}^t$ and $\mathcal{D}$, for $t = 0, 1, \ldots, 10$.
(vi) Derive the diffusion multiscale graph correlation: $\mathrm{MGC}_n^*\left(\{\mathcal{U}^t\}, \mathcal{X}\right)$ by estimating t^*.
(vii) Compute p-value using permutation test.

Output: p-value at the estimated optimal step t^*, the estimated optimal time step t^*, dimension choice of q via profile likelihood method, multiscale local correlation maps $\{\mathrm{DCOR}_n^{kl}(\mathcal{U}^t, \mathcal{X})\}$, the optimal neighbourhood choice (k^*, l^*).

3.2. *Theoretical properties under the exchangeable graph*

To derive consistency of our methodology, the following mild assumptions are required on the distribution of the graph and the nodal attributes.

Condition 1. Graph $\mathcal{G}$ is an induced subgraph of an infinitely exchangeable graph. Namely, the adjacency matrix $\mathcal{A}$ satisfies

$$\mathcal{A}(i,j) \overset{d}{=} \mathcal{A}\{\sigma(i), \sigma(j)\} \tag{3}$$

for any $i, j = 1, 2, \ldots, n$ and any permutation σ of size $n \in \mathbb{N}$. The notation $\overset{d}{=}$ stands for equality in distribution.

Condition 2. Each nodal attribute X_i is generated independently and identically from F_X, and has finite second moment.

Condition 3. The matrix $\mathcal{A}$ is constrained to a domain Ω where the diffusion map embedding from $\mathcal{A} \in \Omega$ to $\mathcal{U}^t$ is injective for some t.

Condition 1 states that $\mathcal{G}$ is a collection of independently sampled nodes and their induced subgraph (Orbanz & Roy, 2015; Orbanz, 2017; Tang et al., 2017), which is a distributional assumption satisfied by many popular statistical network models. Based on Condition 1, the diffusion map $\mathcal{U}^t$ at each t can furnish an exchangeable and asymptotic conditional independent and identically distributed embedding for the set of nodes $\mathcal{V}(\mathcal{G})$.

THEOREM 1. *Assume $\mathcal{G}$ satisfies Condition 1. Then at each fixed t, the embedded diffusion maps $\mathcal{U}^t = \{U_i^t,\ i = 1, 2, \ldots, n\}$ by (2) are exchangeable. As a result, there exists an underlying variable θ^t distributed as the limiting empirical distribution of $\mathcal{U}^t$, such that $U_i^t \mid \theta^t$ are asymptotically independently and identically distributed for $i = 1, 2, \ldots, n$ as $n \to \infty$.*

Due to Condition 1, the permutation test is applicable to any $\mathcal{U}^t$ from an exchangeable sequence. Condition 2 is merely a regularity condition, and the distribution of U_i^t automatically satisfies the same finite-moment assumption, as shown in the Supplementary Material in the proof of Theorem 2. We then have consistency between the diffusion map at each t and the nodal attribute.

THEOREM 2. *Assume the graph $\mathcal{G}$ and the nodal attributes satisfy Conditions 1 and 2. Then as $n \to \infty$, the multiscale graph correlation between the diffusion map $\mathcal{U}^t$ at any fixed t and the nodal attributes $\mathcal{X}$ satisfies:* $\text{MGC}_n(\mathcal{U}^t, \mathcal{X}) \to c \geqslant 0$, *with equality if and only if $F_{U^tX} = F_{U^t}F_X$.*

The testing consistency naturally extends to the diffusion correlation, which further holds between edge connectivity and nodal attributes if Condition 3 is satisfied.

THEOREM 3. *Under the same assumption as Theorem 2, it holds that* $\text{MGC}_n^*(\{\mathcal{U}^t\}, \mathcal{X}) \to c \geqslant 0$, *with equality if and only if $F_{U^tX} = F_{U^t}F_X$ for all $t \in [0, 10]$. Therefore, the diffusion multiscale graph correlation is a valid and consistent statistic for testing independence between the diffusion maps $\{\mathcal{U}^t\}$ and nodal attributes $\mathcal{X}$.*

If Condition 3 holds, then $\text{MGC}_n^*(\{\mathcal{U}^t\}, \mathcal{X})$ *is also valid and consistent for testing independence between the adjacency matrix and nodal attributes, i.e., it converges to* 0 *if and only if the nodal attribute X is independent of the node connectivity A.*

COROLLARY 1. *Theorem 3 still holds when any of the following changes are applied to the testing procedure described in § 3.1:*

(a) The multiscale graph correlation in step (ii) is replaced by distance correlation or the Heller–Heller–Gorfine statistic;

(b) When $\mathcal{A}$ is restricted to be symmetric, binary and of finite rank $q < n$.

Namely, point (a) suggests that under diffusion maps, another consistent dependency measure can also be used to produce a valid and consistent diffusion correlation, which enables us to compare a number of diffusion correlations in the simulations. Point (b) offers an example of a random matrix $\mathcal{A}$ where the diffusion map is guaranteed injective within the domain.

3.3. *Consistency under the random dot product graph*

In this section we illustrate the theoretical results via the random dot product graph model, which is widely used in network modelling. It assumes that each node has a latent position $W_i \overset{i.i.d.}{\sim} F_W$ for $i = 1, 2, \ldots, n$, and the edge probability $P\{\mathcal{A}(i, j) = 1 \mid W_i, W_j\}$ is determined by the dot product of the latent positions, i.e.,

$$\mathcal{A}(i,j) \mid W_i, W_j \overset{i.i.d.}{\sim} \text{Ber}\big(\langle W_i, W_j\rangle\big),\ i,j = 1, 2, \ldots, n \text{ and } i < j,$$

under the restriction that all the W_i are nonnegative vectors and the dot product must be normalized within $[0, 1]$.

A random dot product graph is an exchangeable graph model that satisfies Condition 1. In addition, the random dot product graph fully specifies all exchangeable graph models that are unweighted and symmetric, and whose probability-generating matrix $\mathcal{P}(i,j) = \langle W_i, W_j \rangle$ is positive semidefinite.

PROPOSITION 1 (Sussman et al. (2014)). *An exchangeable random graph has a finite rank q and positive semidefinite link matrix $\mathcal{P}$ if and only if the random graph is distributed according to a random dot product graph with independent and identically distributed latent vectors $\{W_i \in \mathbb{R}^q,\ i = 1, \ldots, n\}$.*

Indeed, many other popular network modellings are special cases of a random dot product graph, including the stochastic block model (Airoldi et al., 2008; Hanneke & Xing, 2009; Rohe et al., 2011; Xin et al., 2017), its degree-corrected version (Karrer & Newman, 2011) and the latent factor model from Fosdick & Hoff (2015).

PROPOSITION 2 (Rohe et al. (2011)). *Let $\mathcal{L}$ be the normalized graph Laplacian for an adjacency matrix $\mathcal{A}$ generated by a random dot product graph with latent positions which construct the matrix of $\mathcal{W} = [W_1|W_2|\cdots|W_n] \in \mathbb{R}^{q\times n}$. Let $\mathcal{U}^{t=1} = [U_1^{t=1}|U_2^{t=1}|\cdots|U_n^{t=1}] \in \mathbb{R}^{q\times n}$. Then there exists a fixed diagonal matrix $\mathcal{M}$ and an orthonormal rotational matrix $\mathcal{Q} \in \mathbb{R}^{q\times q}$ such that $\|\mathcal{U}^{t=1} - \mathcal{Q}\mathcal{M}\mathcal{W}\| \to 0$ almost surely.*

Therefore, under a random dot product graph, the diffusion map $\mathcal{U}^{t=1}$ asymptotically equals the latent position $\mathcal{W}$ up to a linear transformation. As the latent position under a random dot product graph can be asymptotically recovered by diffusion maps, diffusion correlation is consistent against testing general dependency between $\mathcal{A}$ and $\mathcal{X}$ under a random dot product graph.

COROLLARY 2. *Under an induced subgraph from an exchangeable graph with positive semidefinite link function, the diffusion multiscale graph correlation is consistent for testing independence between edge connectivity and nodal attributes.*

3.4. *Discussion of the conditions*

Here we discuss the robustness of the methodology with respect to Conditions 1–3. These conditions are essential to guarantee a consistent and valid testing framework in general, which is not just limited to network dependence testing.

Condition 1 is a crucial condition for the permutation test to be valid. When it is violated and neither set of data can be assumed exchangeable, all aforementioned test statistics may no longer be valid because the permutation test fails to control the Type 1 error level, as demonstrated in Guillot & Rousset (2013). In certain special cases like testing independence between two stationary time series, the block permutation technique can be used to yield a valid test (Lacal & Tjøstheim, 2018), which can be readily used here, but is not guaranteed valid for general nonexchangeable data. Condition 2 is a regularity condition required for a distance-based correlation measure to be well behaved, without which the distance variance can explode to infinity and cause the correlation measure to be ill behaved.

In comparison, the diffusion correlation methodology is still valid without Condition 3. However, the second part of Theorem 3 will no longer hold, and the methodology is no longer universally consistent. Namely, certain signals of dependency may be lost during the diffusion map embedding procedure. As a result, the diffusion correlation could be asymptotically 0 for

some dependencies, and thus no longer able to detect all possible dependencies between the edge connectivity and nodal attributes. In the Supplementary Material we illustrate the performance of the test statistics under the violation of the positive semidefinite link function, and show the relative robustness of distance-based tests compared to model-based tests when Condition 3 is violated.

4. Numerical studies

4.1. *Stochastic block model*

We compare diffusion multiscale graph correlation, diffusion distance correlation, the diffusion Heller–Heller–Gorfine method, the Fosdick–Hoff likelihood ratio test and direct embedding-based tests. We use adjacency spectral embedding and the latent factors to embed the adjacency matrix first, followed by either the multiscale graph correlation, the distance correlation or the Heller–Heller–Gorfine method. For each simulation we generate a sample graph and the corresponding attributes, compute the test statistic of each method, carry out the permutation test with $r = 500$ random permutations and reject the null if the resulting p-value is less than $\alpha = 0.05$. The testing power of each method equals the percentage of correct rejections out of $m = 100$ replicates, and a higher power implies a better method against the given dependency structure.

The first simulation samples graphs from the stochastic block model. It assumes that each of the n nodes in $\mathcal{G}$ must belong to one of $K \in \mathbb{N}$ blocks, and determines the edge probability based on the block membership of the connecting nodes: for $i = 1, \ldots, n$, assume there exists a latent variable of $Z_i \overset{i.i.d.}{\sim} \mathrm{Mu}(\pi_1, \pi_2, \ldots, \pi_K)$ denoting the block membership of each node, and denote the edge probability between any two nodes of class k and l as $b_{kl} \in \{0, 1\}$. Then the upper-triangular entries of $\mathcal{A}$ are independently and identically distributed when conditioning on $\mathcal{Z} = \{Z_i : i = 1, 2, \ldots, n\}$:

$$\mathcal{A}(i,j) \mid Z_i, Z_j \overset{i.i.d.}{\sim} \mathrm{Ber}\left\{ \sum_{k,l=1}^{K} b_{kl} \mathbb{I}(Z_i = k, Z_j = l) \right\} \quad (i < j,\ i,j = 1, 2, \ldots, n),$$

where $\mathbb{I}(\cdot)$ is the indicator function. The sample data is generated with $n = 100$ by using the following parameters:

$$\begin{aligned} &Z_i \overset{i.i.d.}{\sim} \mathrm{Mu}(1/3, 1/3, 1/3), \\ &\mathcal{A}(i,j) \mid Z_i, Z_j \sim \mathrm{Ber}\left\{0.5\mathbb{I}(|Z_i - Z_j| = 0) + 0.2\mathbb{I}(|Z_i - Z_j| = 1) + 0.4\mathbb{I}(|Z_i - Z_j| = 2)\right\}, \\ &X_i \mid Z_i \sim \mathrm{Mu}[\{1 + \mathbb{I}(Z_i = 1)\}/4,\ \{1 + \mathbb{I}(Z_i = 2)\}/4,\ \{1 + \mathbb{I}(Z_i = 3)\}/4], \end{aligned} \tag{4}$$

where $\mathcal{X}$ is a randomly polluted block assignment. For each i, $X_i = Z_i$ with probability 0.5, and equally likely to take other values in Ω, i.e., the true block membership is observed half of the time. For the adjacency matrix, the within-block edge probability is always 0.5, while the between-block edge probability is 0.2 when the block labels differ by 1, and 0.4 when the block labels differ by 2. As the edge probability between a node of block 1 and a node of block 3 is higher than the edge probability between block 1 and block 2, this three-block stochastic block model generates a noisy and nonlinear dependency structure between $\mathcal{A}$ and $\mathcal{X}$, and we now verify how successful the methods are in detecting this.

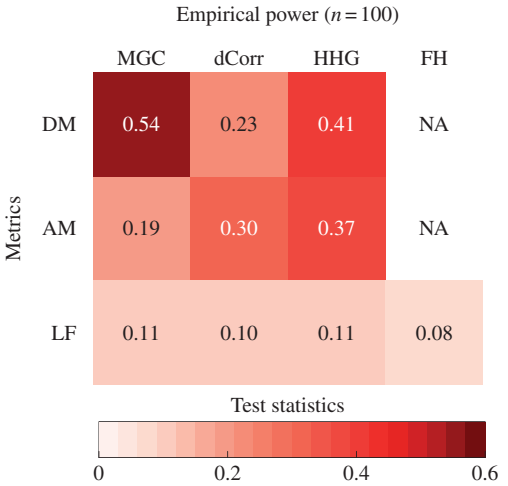

Fig. 2. The testing powers for the three-block stochastic block model in (4). The y-axis lists the embedding choices: diffusion map, DM, adjacency spectral embedding, AM, and latent factor embedding, LF. The x-axis corresponds to the correlation measure in use: the multiscale graph correlation, MGC, distance correlation, dCorr, Heller–Heller–Gorfine, HHG, and the Fosdick–Hoff method, FH.

Figure 2 shows that diffusion multiscale graph correlation has the best testing powers among all the methods, because multiscale graph correlation captures high-dimensional nonlinear dependencies better than distance correlation and the Heller–Heller–Gorfine method. The top three entries in the first row represent the diffusion correlation methods proposed in this paper, which outperform other embedding choices, with diffusion multiscale graph correlation having the best power.

4.2. *Stochastic block model with linear and nonlinear dependencies*

To further understand and demonstrate the advantage of the diffusion approach, here we use the same three-block stochastic block model and its block membership $\{Z_i : i = 1, 2, \ldots, n = 100\}$ as in § 4.1, except that the edge probability is now controlled by $\beta \in (0, 1)$ for all $i, j = 1, \ldots, n$:

$$\mathcal{A}(i,j) \mid Z_i, Z_j \sim \text{Ber}\big\{0.5\mathbb{I}(|Z_i - Z_j| = 0) + 0.2\mathbb{I}(|Z_i - Z_j| = 1) + \beta\mathbb{I}(|Z_i - Z_j| = 2)\big\}. \quad (5)$$

The noisy block membership $\mathcal{X}$ is generated in the same way as before. When $\beta = 0.2$, the three-block stochastic block model is the same as a two-block stochastic block model, where within-block edge probability equals 0.5 while the between-block edge probability is always 0.2, i.e., it represents a linear association between the adjacency matrix and the block membership. When $\beta < 0.2$, the dependency is still monotonic. When $\beta > 0.2$ and gets further away, the relationship becomes strongly nonlinear. Figure 3(a) plots the power against β for all diffusion maps-based methods. When β shifts from less than 0.2 to higher than 0.2, it represents a structural change in the relationship from monotone to nonmonotone. The figure demonstrates that the main approach using the multiscale graph correlation is the most powerful method against varying dependency structure.

4.3. *Degree-corrected stochastic block model*

We now compare different embeddings under the degree-corrected stochastic block model, which is an extension of the stochastic block model obtained by introducing an additional random variable c_i to control the degree of each node, and which better reflects many real-world networks.

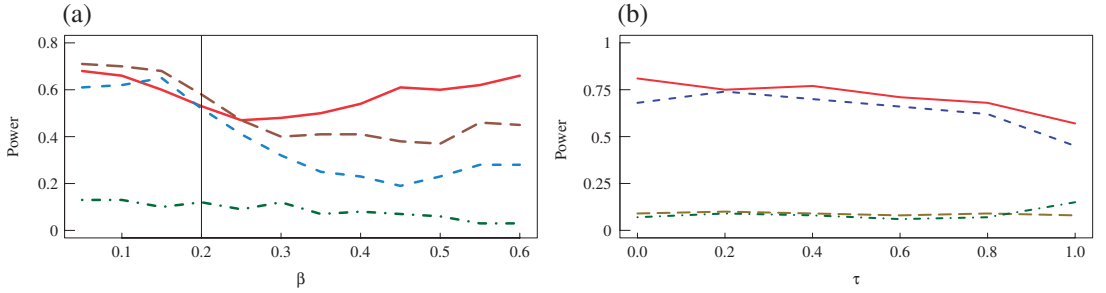

Fig. 3. (a) The power curve with respect to increasing β under the three-block stochastic block model (5). The methods utilizing diffusion maps are: diffusion MGC (solid); diffusion DCOR (long dashes), diffusion HHG (short dashes) and FH test (dot-dash). (b) The power curve with respect to increasing τ under the degree-corrected stochastic block model (6). Diffusion MGC (solid), adjacency spectral embedding followed by MGC (short dashes), latent factor embedding followed by MGC (long dashes) and FH (dot-dash).

We set $n = 200$ with two blocks, select the binary block membership Z_i uniformly in $\Omega = \{0, 1\}$, and generate the edge probability by

$$\mathcal{A}(i,j) \mid Z_i, Z_j, C_i, C_j \sim \text{Ber}\{0.2C_iC_j \cdot \mathbb{I}(|Z_i - Z_j| = 0) + 0.05C_iC_j \cdot \mathbb{I}(|Z_i - Z_j| = 1)\}, \quad (6)$$

where $C_i \overset{i.i.d.}{\sim} \text{Un}(1-\tau, 1+\tau)$ for $i = 1, \ldots, n$, and $\tau \in [0, 1]$ is a parameter to control the amount of variability in the edge degree: as τ increases, the model becomes more complex as the variability of the edge probability becomes larger; when $\tau = 0$, the above model reduces to a two-block stochastic block model without any variability induced by $\{C_i : i = 1, 2, \ldots, n\}$. We again generate the nodal attributes $\mathcal{X}$ as a noisy version of the true block membership via a Bernoulli distribution, i.e., for each i, $X_i = Z_i$ with probability 0.6, and equals the wrong label with probability 0.4. Figure 3(b) compares different embedding choices using multiscale graph correlation.

4.4. *Random dot product graph simulations*

Next we present a variety of random dot product graph simulations by generating the latent variables via the 20 relationships in Shen et al. (2018) with different levels of noise, consisting of various linear, monotonic and nonmonotonic relationships. The details of the simulation schemes are given in the Supplementary Material, and a general outline for a data-generating process is:

$$\begin{aligned} \begin{pmatrix} \tilde{W}_i & \tilde{X}_i \end{pmatrix} &\overset{i.i.d.}{\sim} F_{\tilde{W}\tilde{X}} && (i = 1, 2, \ldots, n), \\ \mathcal{A}(i,j) \mid W_i, W_j &\sim \text{Ber}\left(\langle W_i, W_j \rangle\right) && (i < j = 1, 2, \ldots, n), \end{aligned} \quad (7)$$

where

$$W_i = \frac{\tilde{W}_i - \min(\{\tilde{W}_j : j = 1, 2, \ldots, n\})}{\max(\{\tilde{W}_j : j = 1, 2, \ldots, n\}) - \min(\{\tilde{W}_j : j = 1, 2, \ldots, n\})} \quad \text{for } i = 1, 2, \ldots, n,$$

so that all the latent variables range from 0 to 1. We apply the same scaling from $\tilde{X}_i$ to X_i for visual consistency.

Thus the latent positions and nodal attributes are correlated via a joint distribution of $F_{\tilde{W}\tilde{X}}$, with our choices including joint distributions with linear, quadratic and circular dependencies amongst

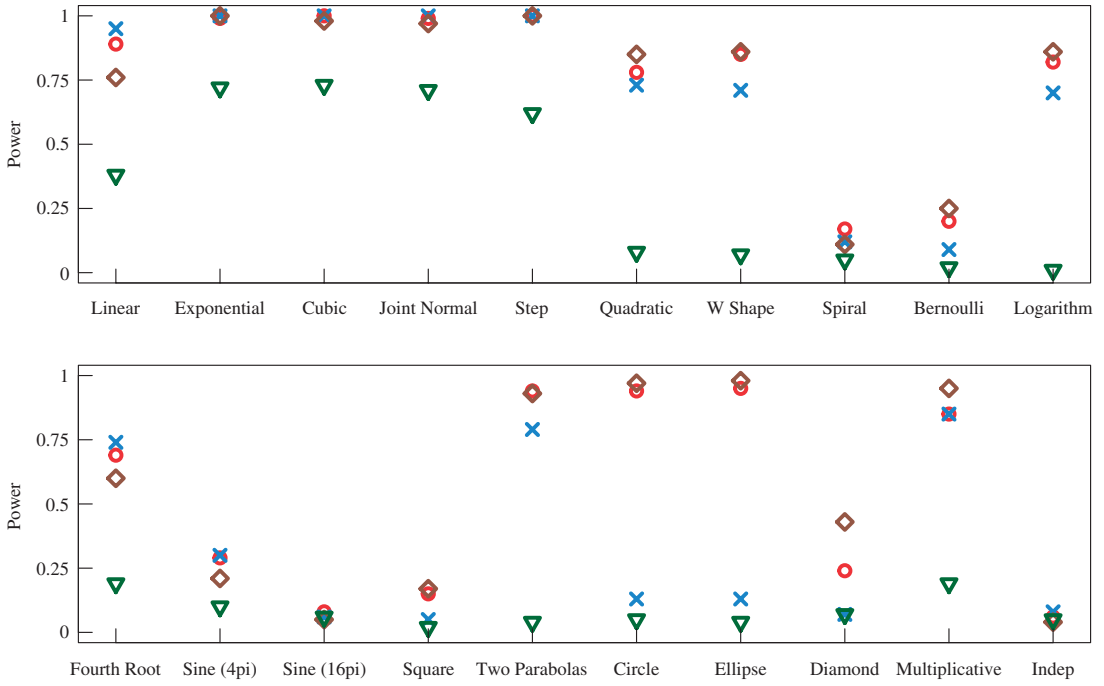

Fig. 4. Power comparison for 20 different random dot product graphs with $n = 50$ nodes per $m = 100$ independent replicates. The methods are diffusion multiscale graph correlation (circle), diffusion distance correlation (cross), the Heller–Heller–Gorfine method (diamond) and the Fosdick–Hoff test (triangle).

others. Figure 4 shows the empirical power obtained from $m = 100$ independent replicates when the number of nodes is $n = 50$, for which all the diffusion-map-based methods work fairly well. It shows that when latent positions W_i and nodal attributes X_i are dependent via a close-to-linear relationship, see the upper panel, diffusion multiscale graph correlation, diffusion distance correlation and the Heller–Heller–Gorfine method achieve similar power while the Fosdick–Hoff test is slightly worse due to its model-based nature. When nonlinearity between W_i and X_i becomes evident, circle or ellipse in the lower panel, multiscale graph correlation and the Heller–Heller–Gorfine method are the two best-performing correlation measures, which is somewhat consistent with the empirical results in Shen et al. (2018) for non-network data. The last scenario is an independent relationship and all tests achieve a power of approximately 0.05, implying that they are all valid tests; there are also a few dependencies of very low power due to the complexity of the relationship, e.g., sine, spiral, square, etc., but their powers all converge to 1 as the sample size n increases.

5. GRAPH EMBEDDING USING DIFFUSION MULTISCALE GRAPH CORRELATION

We now demonstrate that in deriving the diffusion correlation we preserve dependency structure between $\mathcal{A}$ and $\mathcal{X}$ without crossvalidation or overfitting by virtue of effectively estimating the parameters t and q. As a reminder, the dimension choice q is selected by the second elbow of the absolute eigenvalue scree plot via the profile likelihood method from Zhu & Ghodsi (2006). The choice of t^* is based on a smoothed maximum statistic. Viewed another way, diffusion multiscale graph correlation selects the optimal diffusion map that maximizes the multiscale graph correlation. Thus any testing advantage comes down to whether it is

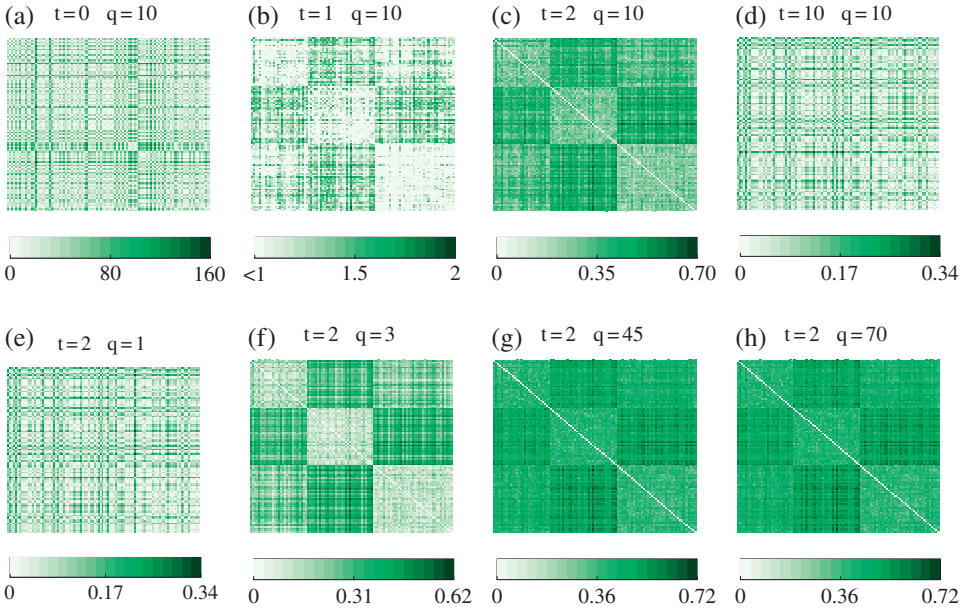

Fig. 5. Generate a three-block adjacency matrix $\mathcal{A}$ by (4) at $n = 100$, and compute the diffusion distances at each combination of (t, q). A visualization of the adjacency matrix is provided in Fig. 6(a).

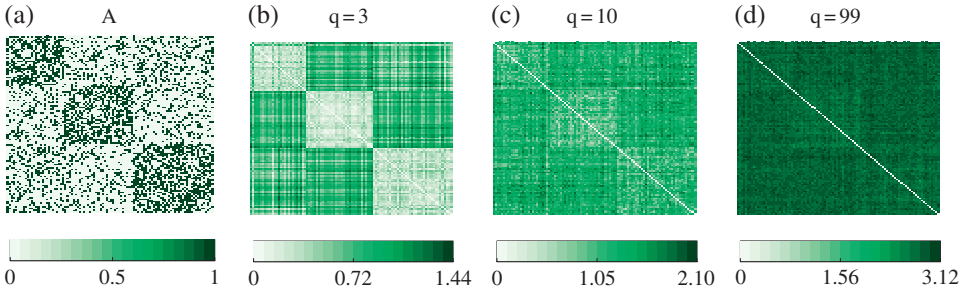

Fig. 6. Panel (a) shows the adjacency matrix of the three-block adjacency matrix $\mathcal{A}$ generated by (4). Panels (b)–(d) show the Euclidean distance matrix of adjacency spectral embedding at increasing q, using the same adjacency matrix as panel (a).

able to optimize the embedding without overfitting, and we investigate how well our procedure is able to preserve the dependency compared to adjacency spectral embedding.

Figure 5 presents the diffusion distances at different t and q for the three-block stochastic block model in (4). Although the resulting embedding is sensitive to both t and q in Fig. 5(a)–(d), at optimal $t^* = 2$ it is robust against q; e.g., Fig. 5(e)–(h) show that for a wide range of q the block structure is preserved in the resulting diffusion maps including the second elbow, so the diffusion correlation-based embedding preserves the dependency structure well.

On the other hand, Fig. 6 shows the Euclidean distance of the adjacency spectral embedding (Sussman et al., 2012) applied to the same adjacency matrix. For adjacency spectral embedding, the correct dimensional choice equals the number of blocks, i.e., the distance matrix at $q = 3$ shows a clear block structure in Fig. 6(b). However, a slight misspecification of q can cause the embedding to have a more obscure block structure, and the elbow method often fails to find the correct q for adjacency spectral embedding.

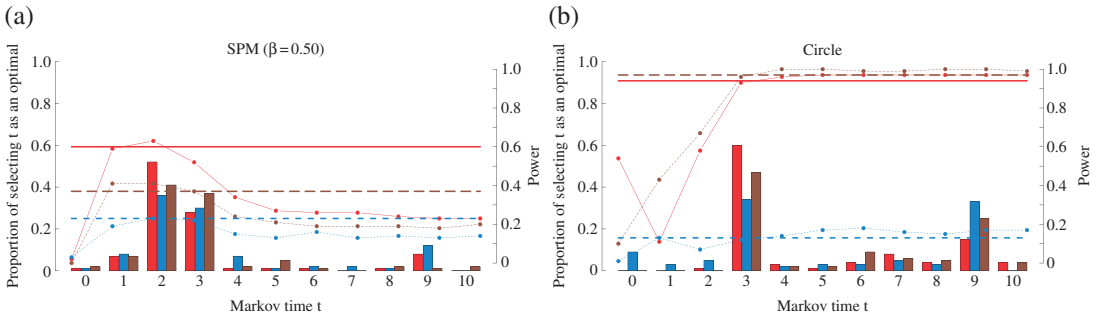

Fig. 7. Testing power comparison between diffusion MGC and MGC on each diffusion map. Using $m = 100$ replicates, the solid red line plots the power of $\text{MGC}_n^*(\{\mathcal{U}^t\}, \mathcal{X})$; the dashed line plots the power of $\text{MGC}_n(\mathcal{U}^t, \mathcal{X})$ for $t \in \{0, 1, 2, \ldots, 10\}$; the bar plot shows the proportion that diffusion MGC (solid) selects each $t \in \{0, 1, 2, \ldots, 10\}$ as the optimal t^*. Diffusion HHG (small dashes) and diffusion DCOR (dashes) are also added in different colours.

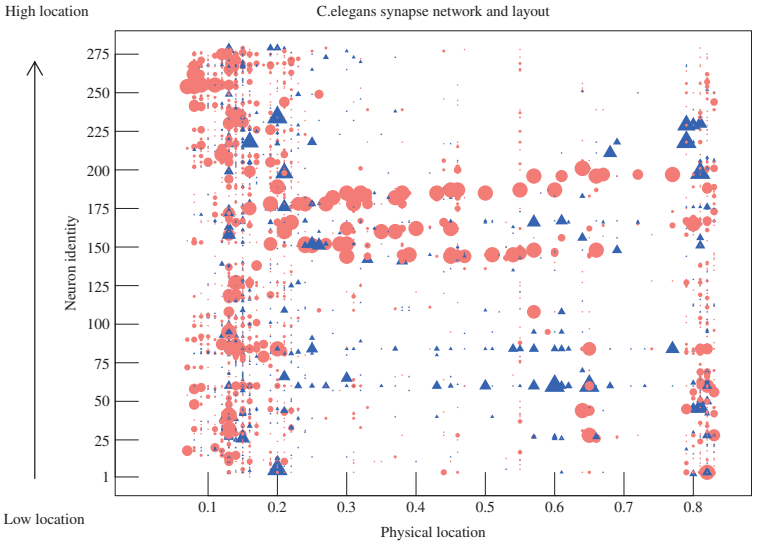

Fig. 8. Each dot at (x, y) represents the existence of synapses from the neuron at y to the neuron at x. The y-axis represents each neuron's index assigned from low location to high location, and the x-axis represents 68 different locations where neurons are positioned. The colour of each dot represents synapse type, either chemical (red circle) or electrical (blue triangle), and the size of each dot is proportional to the number of synapses, but is capped at 10.

Next we compare the testing performance of the diffusion correlation-based embedding $\mathcal{U}^{t^*}$ versus all other diffusion maps $\mathcal{U}^t$. Figure 7 shows the proportion of time each t is chosen as optimal, and the testing power for each t and also t^*. Figure 7(a) illustrates that under the stochastic block model dependency structure in (5) with $\beta = 0.50$, diffusion multiscale graph correlation is mostly likely to choose $t^* = 2$ as the optimal time step, and the testing power is almost equivalent to the best power among all $t \in \{0, 1, 2, \ldots, 10\}$. The same phenomena hold for other diffusion correlations, and Fig. 7(b) illustrates the results via the random dot product graph simulation example by (7).

6. REAL DATA APPLICATION

As a real data example, we apply the methodology to the neuronal network of hermaphrodite *Caenorhabditis elegans*, which is composed of 279 nonpharyngeal neurons connected to each

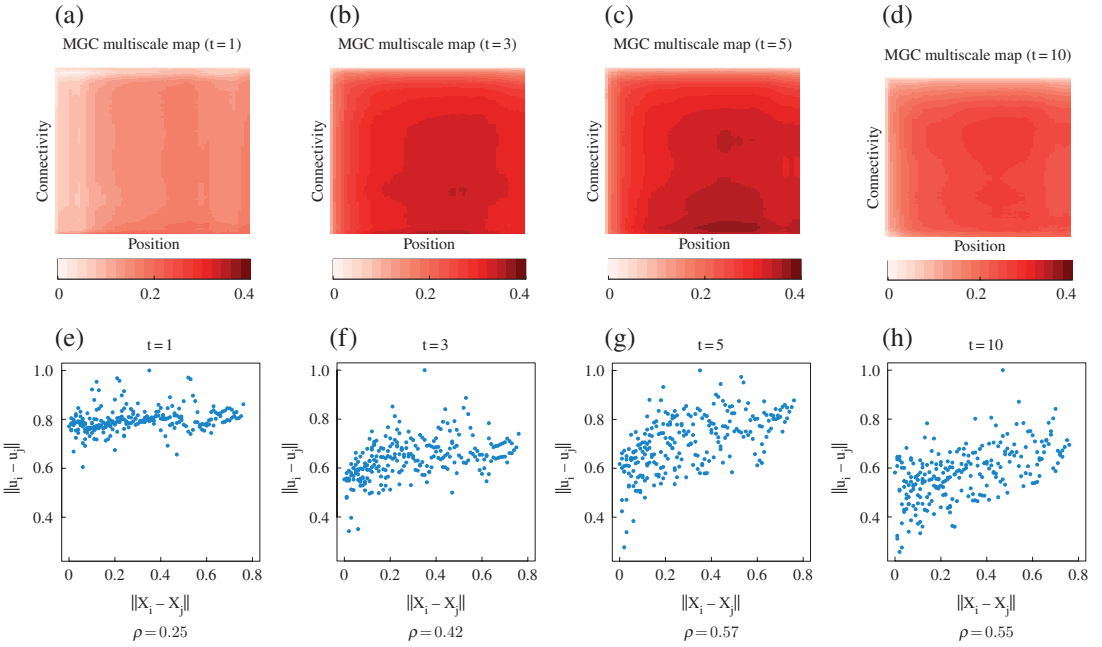

Fig. 9. Local correlation maps at different diffusion times. Panel (c) presents the correlation map at the optimal time $t^* = 5$ identified by diffusion multiscale graph correlation. Panels (e)–(h) show the standardized Euclidean pairwise distance between $\{U_i^t\}$ and $\{X_i\}$ for $t = 1, 3, 5, 10$.

other through chemical and electrical synapses (Varshney et al., 2011). Each node represents an individual neuron, and each edge weight indicates the number of synapses between them. Among a few known attributes, including types of neurotransmitter and role of neurons, we use the one-dimensional, continuous position of each neuron as the nodal attribute $\mathcal{X}$. Figure 8 shows that neurons at low location and high location are connected to other neurons distributed throughout the region, while those at relatively central locations are connected to the neurons only within a narrower area. The independence test between synapse connectivity and each neuron's position can be connected to a growing number of studies on the relationship between physical arrangement and functional connectivity in *C. elegans* (Cherniak et al., 2004; Chen et al., 2006; Kaiser & Hilgetag, 2006; Alexander-Bloch et al., 2012). We binarize and symmetrize both chemical and electrical synapses, add them together to represent the overall synapse connectivity of *C. elegans*, then apply diffusion multiscale graph correlation, diffusion distance correlation, the Heller–Heller–Gorfine method and the Fosdick–Hoff method to test the independence between connectivity through synapses and the neuron's position. All methods result in similar significant p-values less than 0.002.

Figure 9(a)–(d) presents the local distance correlation map $\text{DCOR}^{kl}(\mathcal{U}, \mathcal{X})$ across diffusion times. These plots show that the optimal local correlation is detected at nonglobal neighbourhood choice, i.e., $l^* \neq 68$ (the global maximum), which implies a nonlinear dependence between connectivity and position and an optimal $t^* = 5$. Figure 9(e)–(h) illustrate the relationship between Euclidean distance in diffusion maps and nodal attributes at different diffusion times at $t = 1, 3, 5, 10$, which is again the most significant at $t = 5$.

7. Discussion

There are several potential follow-ups that would further advance this work. One example is more theoretical investigation into the smoothed maximum statistic and dimension selection

of t'. Assuming t' is the true optimal diffusion time, it will be helpful to either identify a more systematic and reliable way to estimate t', or quantify the variability in the estimated optimal t^* by the smoothed maximum statistic. This would hopefully reduce the computational burden instead of going over all possible diffusion times, e.g., $t = 0, 1, 2, \ldots, 10$. Moreover, although we briefly discussed one example in § 5, the impact of the dimensional choice of q on the embedding quality is still obscure. Finally, since one can apply a diffusion map to any data and one can think of any affinity or kernel matrix as a graph, this method is actually applicable to more general testing scenarios beyond networks.

ACKNOWLEDGEMENT

The authors thank Dr. Minh Tang, Dr. Daniel Sussman, the editor and reviewers for their insightful suggestions to improve the paper. This work was supported by the National Science Foundation and the Defense Advanced Research Projects Agency. Vogelstein is also affiliated with the Institute for Computational Medicine and the Kavli Neuroscience Discovery Institute.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes proofs of the theoretical results and simulation settings of random dot product graphs. The `R` code and accompanying data are publicly available online at `http://neurodata.io/tools/mgc` and `https://github.com/neurodata/mgc`.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. & XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9**, 1981–2014.

ALEXANDER-BLOCH, A. F., VERTES, P. E., STIDD, R., LALONDE, F., CLASEN, L., RAPOPORT, J., GIEDD, J., BULLMORE, E. T. & GOGTAY, N. (2012). The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cereb. Cortex* **23**, 127–38.

CHEN, B. L., HALL, D. H. & CHKLOVSKII, D. B. (2006). Wiring optimization can relate neuronal structure and function. *Proc. Nat. Acad. Sci.* **103**, 4723–8.

CHEN, L., SHEN, C., VOGELSTEIN, J. T. & PRIEBE, C. E. (2016). Robust vertex classification. *IEEE Trans. Pat. Anal. Mach. Intel.* **38**, 578–90.

CHERNIAK, C., MOKHTARZADA, Z., RODRIGUEZ-ESTEBAN, R. & CHANGIZI, K. (2004). Global optimization of cerebral cortex layout. *Proc. Nat. Acad. Sci.* **101**, 1081–6.

COIFMAN, R. R. & LAFON, S. (2006). Diffusion maps. *Appl. Comp. Harmon. Anal.* **21**, 5–30.

COIFMAN, R. R., LAFON, S., LEE, A. B., MAGGIONI, M., NADLER, B., WARNER, F. & ZUCKER, S. W. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Nat. Acad. Sci.* **102**, 7426–31.

FOSDICK, B. K. & HOFF, P. D. (2015). Testing and modeling dependencies between a network and nodal attributes. *J. Am. Statist. Assoc.* **110**, 1047–56.

GRETTON, A. & GYORFI, L. (2010). Consistent nonparametric tests of independence. *J. Mach. Learn. Res.* **11**, 1391–423.

GUILLOT, G. & ROUSSET, F. (2013). Dismantling the mantel tests. *Meth. Ecol. Evol.* **4**, 336–44.

HANNEKE, S. & XING, E. P. (2009). Network completion and survey sampling. *Proc. Mach. Learn. Res.* **5**, 209–15.

HELLER, R., HELLER, Y. & GORFINE, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100**, 503–10.

HELLER, R., HELLER, Y., KAUFMAN, S., BRILL, B. & GORFINE, M. (2016). Consistent distribution-free k-sample and independence tests for univariate random variables. *J. Mach. Learn. Res.* **17**, 1–54.

HERNANDEZ-HERNANDEZ, G., MYERS, J., ALVAREZ-LACALLE, E. & SHIFERAW, Y. (2017). Nonlinear signaling on biological networks: The role of stochasticity and spectral clustering. *Phys. Rev. E* **95**, 032313.

KAISER, M. & HILGETAG, C. C. (2006). Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.* **2**, e95.

KARRER, B. & NEWMAN, M. E. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107.

Kim, J. Y., Howard, M., Cox Pahnke, E., & Boeker, W. (2016). Understanding network formation in strategy research: Exponential random graph models. *Strat. Manage. J.* **37**, 22–44.

Lacal, V. & Tjøstheim, D. (2018). Estimating and testing nonlinear local dependence between two time series. *J. Bus. Econ. Statist.*, DOI: 10.1080/07350015.2017.1407777.

Lafon, S. & Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Trans. Pat. Anal. Mach. Intel.* **28**, 1393–403.

Lewis, K., Gonzalez, M. & Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proc. Nat. Acad. Sci.* **109**, 68–72.

Liang, X., Zou, Q., He, Y. & Yang, Y. (2013). Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. *Proc. Nat. Acad. Sci.* **110**, 1929–34.

Nekovee, M., Moreno, Y., Bianconi, G. & Marsili, M. (2007). Theory of rumour spreading in complex social networks. *Physica A: Statist. Mech. Appl.* **374**, 457–70.

Orbanz, P. (2017). Subsampling large graphs and invariance in networks. *arXiv:*1710.04217.

Orbanz, P. & Roy, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pat. Anal. Mach. Intel.* **37**, 437–61.

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **58**, 240–42.

Peel, L., Larremore, D. B. & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Sci. Adv.* **3**, e1602548.

Rizzo, M. & Székely, G. (2016). Energy distance. *Wiley Interdisc. Rev.: Comput. Statist.* **8**, 27–38.

Rohe, K., Chatterjee, S. & Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39**, 1878–915.

Sejdinovic, D., Sriperumbudur, B., Gretton, A. & Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **5**, 2263–91.

Shen, C., Priebe, C. E. & Vogelstein, J. T. (2018). From distance correlation to multiscale graph correlation. *J. Amer. Statist. Assoc.*, DOI: 10.1080/01621459.2018.1543125.

Shen, C. & Vogelstein, J. T. (2018). The exact equivalent of distance and kernel methods for hypothesis testing. *arXiv:*1806.05514v3.

Shen, C., Vogelstein, J. T. & Priebe, C. (2017). Manifold matching using shortest-path distance and joint neighborhood selection. *Pat. Recog. Lett.* **92**, 41–8.

Sussman, D., Tang, M., Fishkind, D. & Priebe, C. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107**, 1119–28.

Sussman, D. L., Tang, M. & Priebe, C. E. (2014). Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE Trans. Pat. Anal. Mach. Intel.* **36**, 48–57.

Székely, G. & Rizzo, M. (2013). The distance correlation *t*-test of independence in high dimension. *J. Multivar. Anal.* **117**, 193–213.

Székely, G. & Rizzo, M. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42**, 2382–412.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–94.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V. & Priebe, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random dot product graphs. *Bernoulli* **23**, 1599–630.

Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. (2011). Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001066.

Vogelstein, J. T., Wang, Q., Bridgeford, E., Priebe, C. E., Maggioni, M. & Shen, C. (2019). Discovering and deciphering relationships across disparate data modalities. *eLife* **8**, e41690.

Wang, S., Shen, C., Badea, A., Priebe, C. E. & Vogelstein, J. T. (2018). Signal subgraph estimation via iterative vertex screening. *arXiv:*1801.07683.

Wasserman, S. & Pattison, P. (1996). Logit models and logistic regressions for social networks I. An introduction to Markov graphs and p^*. *Psychometrika* **61**, 401–25.

Xin, L., Zhu, M., & Chipman, H. (2017). A continuous-time stochastic block model for basketball networks. *Ann. Appl. Statist.* **11**, 553–97.

Zhu, M. & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Comp. Statist. Data Anal.* **51**, 918–30.

[*Received on* 29 *January* 2018. *Editorial decision on* 14 *February* 2019]