Testing independence between networks and nodal attributes via multiscale metrics and multiscale distance correlation

John Smith,[1*] Jane Doe,[1] Joe Scientist[2]

[1]Department of Chemistry, University of Wherever,
An Unknown Address, Wherever, ST 00000, USA
[2]Another Unknown Address, Palookaville, ST 99999, USA

[*]To whom correspondence should be addressed; E-mail: ylee160@jhu.edu.

Network dependence over network space, which refers to the dependence between network topology and its nodal attributes, often exhibits nonlinear dependent properties. Unfortunately, without knowledge on specific neighborhood structures, no statistic has been suggested to test network dependence further than globally linear dependence. This paper introduces a multiscale dependence test statistic called Multiscale Network Test (`MNT`), which borrows the idea of diffusion maps and Multiscale Generalized Correlation (`MGC`). Our methodology for testing network independence to any multivariate nodal attributes can be applied to any exchangeable graph. We prove the consistency of `MGC` in this application and demonstrate better performance than any other model-based, global test statistics.

Outlines

1. Abstract :

   - This paper introduces the novel statistical methods to test independence between network and nodal attributes.

   - KEY WORDS : multivariate independence test, diffusion maps, distance correlation, k-nearest neighbors,

2. INTRODUCTION

   (a) introduce interests in association between network and nodal attributes and the related literatures.

      - explicitly state the problem and its importance, previous approaches' weakness.

      - difference to previous/common approach, what is our specialty.

   (b) Introduce notations we are going to use and introduce a common network model of nodal attributes.

      - briefly explains the difficulties (how to represent network topology) and how we overcome them

      - introduce outlines

3. METHODOLOGY

   (a) introduce a exchangeable graph

      i. definition of exchangeable graph and lemmas under directed graph

      ii. explain difficulties in undirected or more general graph and introduce representation theorem

2

      iii. explains more general version of graph, i.e. graphex

      iv. state corollary that commonly used graphical models which satisfy exchangeability, e.g. SBM, RDPG, Fosdick's latent model

(b) represent a network structures as a one-parameter family or multivariate variables

      i. introduce notations and definition (coifman & ming)

      ii. explains spectral properties of diffusion maps

      iii. Lemmas about properties of diffusion maps under exchangeable graphs

(c) introduce Dcor and MGC to apply testing independence G and X

      i. introduce distance correlation and multiscale version

      ii. discuss the choice of proper metrics for testing network independence and demonstrate diffuion distance is appropriate.

      iii. present the main theorem and a few remarks.

4. SIMULATION PERFORMANCE AND PRACTICAL IMPLEMENTATION

(a) stochastic block model

      i. simplest two block model

      ii. simplest three block model

      iii. sparse version of block model

(b) random-dot product grap

      i. simplest RDPG + "remarks" about consistent estimation on latent variable

      ii. comparison to Fosdick's method.

5. APPLICATION

(black box)

6. CONCLUDING REMARKS

    (a) discuss merits of using MNT

    (b) discuss limitations of using MNT

    (c) discuss possible contributions of MNT and future works

7. APPENDIX

8. REFERENCES

9. SUPPLEMANTARY FIGURES

Contents

1   Introduction

1.1   intro1

Network, a collection of nodes and edges between them, has been a celebrated area of study over a field of psychology, information theory, biology, statistics, economics, etc. The relationship between the way a pair of nodes are connected and the values of their attribute is a common interest in network analysis. There has been a lot of efforts to represent a network as a function of nodal attributes or model an outcome of nodal attribute variables through their underlying network structures. However, it is very obscure to determine which one should be put as a dependent variable. And most of all network often does not have a natural structure. This is why there exist a plethora of works on latent structure of network, which also depends on the characteristics of each node (Hoff et al. (2002) , Austin, Linkletter, and Wu (2013)). In a latent space model, local independence between network and nodal attributes, conditional on a latent variable is often assumed (Lazarsfeld and Henry (1968)), which makes easier to interpret the dependency mechanism. In real network data, however, it is almost impossible to estimate such latent variable without any knowledge on true network generative model and also we cannot guarantee that direction or amount of association between network and nodal attributes keeps consistent across latent variables, i.e. we cannot guarantee linear dependence.

1.2   intro2

Throughout this paper, assume that we are given an unweighted and undirected, connected network G comprised of n nodes, for a fixed $n \in \mathbb{N}$. Suppose that our observation of G is one random sample from true, population distribution of $\mathcal{G}$. Even though we assume that G is an undirected and unweighted network but 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 $A = \{A_{ij} : i, j = 1, .., n\}$, is often introduced to formalize this relational data of network.

Let us introduce a m-variate ($m \in \mathbb{N}$) variable for nodal attributes $\boldsymbol{X} \in \mathbb{R}^m$ which we are interested in. Investigating correlation between $\boldsymbol{G}$ and $\boldsymbol{X}$ and testing whether their distributions are independent or not is the key focus in our study. Even though distribution of graph $\boldsymbol{G}$ is often formalized through modeling adjacent relations $\{a_{ij} : i, j = 1, ..., n\}$, we cannot directly use $\mathbf{A}$ in testing independence to $\mathbf{X}$, mostly due to difficulty of extracting vertex-wise topological information. Rather than regressing a_{ij} on observed attribute values x_i or x_j, [Fosdick and Hoff] proposed a latent variable model to estimate node-specific network factor which provides a one-to-one correspondence between $\boldsymbol{G}$ and $\boldsymbol{X}$ as well as reduces a dimension of network data. In their paper, Fosdick and Hoff also used these factors to test independence between $\boldsymbol{G}$ and $\boldsymbol{X}$ through testing independence between network factors and nodal attributes. However, the performance of this test would not be good enough when the relational data $\boldsymbol{A}$ does not have linear relationship to network factors or has a latent mixture model. Since we never know the structure of networks and the way they are related to other variables, there always exist a limitation on testing based on modeling network.

In this paper without modeling nor estimating networks, we develop multiscale test statistics which are robust to both nonlinearity and high dimensionality. Multiscale is not avoidable because we view relationship or distance over network as dynamic process so we choose the optimal scale where distance in network space and distance in terms of attributes become most correlated each other. Thus basically we define a coordinate over network at each time in process and test independence to attributes $\mathbf{X}$. In the next methodology section 2, we are going to define such multiscale distance and its properties, and test statistic using multiscale distance metrics will be followed. In section 3, we demonstrate best performance of our method compared to the existing under various circumstances through numerical results. Real data example in section 4 show one of the applications among many.

2 Methodology

First step in testing network independence to other variables is to figure out the variable which configurates location of vertices over network space. We hope that our observed network be a realization of such variable so that our test results can be generalized to those from population of given network. Moveover a pair of our observations for each variable we are going to test should be i.i.d realization to guarantee representativeness and avoid redundant intra-dependence.

2.1 Exchangeable Graph

2.1.1 [EG1]

Assuming that we are given one network, equivalently one graph, it is not clear to consider given network as an independent and identically distributed sample of some variable. It is true that if you consider each edge is i.i.d., then observed adjacency matrix is a random sample of a single parameter which is the probability of having edge or not. However, in this case, the resulting network model only depends on number of edges [1] and does not cover undirected or no self-loop networks.

The property of exchangeability is closely related to the use of independent and identically-distributed(i.i.d) random variable. Network or graph **G** is called exchangeable if and only if its adjacency matrix **A** is jointly exchangeable [1].

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, and separately exchangeable if

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

9

for every pair of permutation σ, σ' of n.

2.1.2 [EG2]

However, exchangeability itself cannot guarantee being $i.i.d$. Fortunately, thanks to the celebrated *de Finetti*(A.1)'s representation theorem, it has been proven that there exists a random probability measure η on random variable $\mathbf{Z}$ that a sequence of $Z_1, Z_2, ...$ are *i.i.d* conditional η *if and only if* the sequence is exchangeable [1] [2]. *Aldous-Hoover theorem*(A.2) is the representation theorem of 2-array exchangeable array, which would be useful to explain jointly exchangeable adjacent matrix. Exchangeable graph is commonly called *graphon*. Exchangeable graphon is defined through a random measurable functions follows [3]:

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 $\mathbf{A}$ be an adjacency matrix of graphon. Then for any $i \leq j, i, j = 1, 2, ..., n$:

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

Thanks to *Aldous-Hoover theorem*, we can now better represent exchangeable network through measurable function, but we are still halfway done in case of undirected network where $A_{ij} = A_{ji}$ $(i, j = 1, 2, ..., n)$. Under undirected network where self-loop is still allowed, we can represent $\{A_{ij} : i \leq j\}$ as a function of g conditioning on $\{u_i\}$. In the next section, we will explain the shortcomings of representing network as a graphon and introduce more general exchangeable graph in the next.

$$(A_{ij}) = (A_{\sigma(i)\sigma(j)}) \iff A_{ij} \overset{ind}{\sim} Bern\big(g(u_i, u_j)\big), i \leq j \tag{2}$$

2.1.3 [EG3]

Graphon has been studied widely as a limit of random graphs (ref :). However, despite its advantage on simple representation, is either empty or dense. Thus it fails to represent real network data where sparsity or scale-free distribution if fairly common. In addition to graphon,we introduce a concept of *graphex*, first proposed by Veitch and Roy, 2015[13], which is more generalized version of graphon and also includes sparse exchangeable graphs [2]. A trick Caron and Fox used is representing a network as a point process on $\mathbb{R}_+^2$ based on *Kallengerg Representation Theorem*[5]. Like we are able to represent $\{A_{ij}\}$ through random transformation of *i.i.d* uniform variables, jointly exchangeable point processing network also can be represented via a random function of *i.i.d* unit rate Poisson process and uniform variables. To be specific, undirected graph on a point process on $\mathbb{R}_+^2$ can be thought of

$$\mathbf{A} = \sum_{i,j} A_{ij} \delta_{(\theta_i, \theta_j)} \tag{3}$$

,where $A_{ij} = A_{ji} \in \{0,1\}$ with $\theta_i \in \mathbb{R}$, $i,j = 1,2,....$ We can think Node i is embedded on real line, at $\theta_i \in \mathbb{R}_+$.

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

$$\left(A(V_i \times V_j)\right) \overset{d}{=} \left(A(V_{\sigma(i)} \times V_{\sigma(j)})\right) \tag{4}$$

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

Definition 2.4 (graphex (Veitch and Roy, 2015))**.**

2.1.4 [EG4]

Widely used graphical model, e.g. Stochastic Block Model (SBM), Random Dot Product Graph (RDPG), satisfies this definition of jointly exchangeable graph under some constraints.

2.2 Multiscale Distance Metrics

2.2.1 [MDM1]

There have been a lot of efforts to represent the network in terms of a summarizing network factor[4] or some meaningful coefficients, e.g. centrality[7][8]. However, there has been no vertex-wise variable which provides a configuration of vertex over network space without losing any information. Coifman & Lafon[10] demonstrated that diffusion maps provide a meaningful multiscale geometries of data while keeping information on every local relation. Diffusion map is constructed via Markov chain on graph. Here an adjacency matrix $\boldsymbol{A}$ acts as a kernel, representing a similarity between each node in $\boldsymbol{G}$; thus we do not have to estimate anything to obtain multidimensional representation of network topology through diffusion maps.

Let $(\boldsymbol{G}, \mathcal{A}, \mu)$ be a measure space. Throughout all of the arguments, assume that we have a countable vertex set with size of $n \in \mathbb{N}$. The vertex set of network $\boldsymbol{G}$ is the data set of vertices and edges and $\mathcal{A}$ is a set of a pair of nodes $\{(i, j) : v_i, v_i \in V(\boldsymbol{G})\}$. A measure of μ which represents a distribution of the vertices on $\boldsymbol{G}$, is equivalent to an adjacency matrix $\boldsymbol{A}$. A transition matrix $\mathbf{P} = \{P[i, j] : i, j = 1, ..., n\}$ in Markov chain on $\boldsymbol{G}$, which represents the probability that flow or signal goes from Node i to Node j, is defined as below:

$$P[i, j] = A_{ij} / \sum_{j=1} A_{ij} \tag{5}$$

A transition matrix $\mathbf{P}$ is a new kernel of a Markov chain of which element $P[i, j]$ represents the probability of travel from Node i to Node j in one time step. On the other hand, corresponding probability in t steps is given by the t th ($t \in \mathbb{N}$) power of P. How to derive diffusion distance over a directed network or weighted network is provided in [Tang & Trosset (2010)]. Other than a transition matrix, we need a stationary probability $\boldsymbol{\pi} = \{\pi(1), \pi(2), ..., \pi(n)\}$ of which $\pi(i)$ represents the probability that the chain stays in Node i regardless of the starting

state. In our setting, $\pi(i)$ is proportional to the degree of Node i, i.e. $\pi(i) = \sum\limits_{j=1}^{n} A_{ij} \Big/ \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} A_{ij}$ ($i = 1, 2, ..., n$).

For each time point $t \in \mathbb{N}$, we can define a diffusion distance C_t given by :

$$C_t^2[i,j] = \sum_{w=1}^{n} \left(P^t[i,w] - P^t[j,w]\right)^2 \frac{1}{\pi(w)} = \sum_{w=1}^{n} \left(\frac{P^t[i,w]}{\sqrt{\pi(w)}} - \frac{P^t[j,w]}{\sqrt{\pi(w)}}\right)^2 \qquad (6)$$
$$= \parallel P^t[i,\cdot] - P^t[j,\cdot] \parallel_{L^2(\boldsymbol{G}, d\mu/\pi)}^2$$

As diffusion time t increases, distance matrix C_t is more likely to take into account distance between two nodes far away in terms of the length of the path. Key idea behind such diffusion distance at fixed time t is that it measures the chance that we are likely to stay between `Node i` and `Node j` at t step on our journey of all other possible paths. The higher chance is, the smaller distance between two is. This distance well reflects the connectivity between two nodes. Connectivity between two nodes is higher if we need to eliminate more number of vertices to disconnect these two. Unlike an adjacency relation or geodesic distance, a connectivity between two nodes depends on their relationship to other vertices in a given network so it is more robust to the unexpected edges. Often a set of nodes with higher connectivity have a higher propensity of having edges within this set and they are likely to form a cluster. Thus diffusion distance is very robust measure and also very sensitive to the clustering structure of network.

2.2.2 [MDM2]

Diffusion distance of $\boldsymbol{G}$ defined as above can be represented via a spectral decomposition of its transition matrix P. That is, we can derive diffusion distance using its eigenvectors and eigenvalues. The spectral analysis on diffusion distance or diffusion maps have been studied for its usefulness for nonlinear dimensionality reduction ([Coifman & Lafon (2006)], [Lafon & Lee (2006)]). Keeping mind that diffusion distance at time t, C_t is a functional L^2 distance, weighted by $1/\pi$. If we transform the way to represent $C_t[i,j]$ slightly, we are able to obtain

an orthonormal basis of $L^2(\mathbf{G}, d\mu/\pi)$ via eigenvalues and eigenvectors. Since an adjacency matrix A does not guarantee a symmetric of P, define a symmetric kernel $\boldsymbol{Q} = \boldsymbol{\Pi}^{1/2} \boldsymbol{P} \boldsymbol{\Pi}^{-1/2}$, where $\boldsymbol{\Pi}$ is a $n \times n$ diagonal matrix of which ith diagonal element is $\pi(i)$. Under compactness of P, $\boldsymbol{Q}$ has a discrete set of real nonzero eigenvalues $\{\lambda_r\}_{r=\{1,2,...,q\}}$ and a set of their corresponding orthonormal eigenvectors $\{\psi_r\}_{r=\{1,2,...,q\}}$, i.e. $Q[i,j] = \sum_{r=1}^{q} \lambda_r \psi_r(i) \psi_r(j)$ $(1 \leq q \leq n)$. Since $P[i,j] = \sqrt{\pi(j)/\pi(i)} Q[i,j]$, $P[i,j] = \sum_{r=1}^{q} \lambda_r \{\psi_r(i)/\sqrt{\pi(i)}\}\{\psi_r(j)\sqrt{\pi(j)}\} := \sum_{r=1}^{q} \lambda_r \phi_r(i)\{\psi_r(j)\sqrt{\pi(j)}\}$, where $\phi_r(i) := \psi_r(i)/\sqrt{\pi(i)}$. Then from $\sum_{r=1}^{q} \psi_r(j)\sqrt{\pi(j)} = 1$ for all $j \in \{1, 2, ..., n\}$, we can represent the diffusion distance as:

$$C_t^2[i,j] = \sum_{r=1}^{n} \lambda_r^{2t} \big(\phi_r(i) - \phi_r(j)\big)^2$$
$$= \| P^t[i,\cdot] - P^t[j,\cdot] \|_{L^2(\boldsymbol{G}, d\mu/\pi)}^2 \tag{7}$$

That is,

$$C_t[i,j] = \| \boldsymbol{U}_t(i) - \boldsymbol{U}_t(j) \| \tag{8}$$

, where

$$\boldsymbol{U}_t(i) = \begin{pmatrix} \lambda_1^t \phi_1(i) \\ \lambda_2^t \phi_2(i) \\ \vdots \\ \lambda_q^t \phi_q(i) \end{pmatrix} \in \mathbb{R}^q. \tag{9}$$

Now we call the family of q-variate$(q \leq n)$ diffusion maps $\{U_t\}_{t \in \mathbb{N}}$, of which Euclidean distance is diffusion distance.

2.2.3 [MDM3]

- *[MDM3] properties of diffusion maps under exchangeable graphs*

Due to the inter-correlated construction of U, 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 samples without any restrictions on networks. Independence within sample of one variable is important in testing independence with other variable. (why? example)

To talk about independence of U, we need more general concept of exchangeable graph. Remind that under jointly exchangeable graph, $(A_{ij}) \overset{d}{=} (A_{\sigma(i)\sigma(j)})$ for every permutation σ of n. The main advantage from such representation of $\mathbf{A}$ is that it becomes now possible to represent given network $\boldsymbol{G}$ as a one-parameter family of $q(< n)$-coordinates called diffusion coordinate, of which metric well reflects how corresponding vertices are connected each other at each diffusion time point. A set of diffusion maps for each vertex have some desirable properties which will be helpful in testing. The followings are concerning about a few conditions for earning them.

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 $\{A_{ij} : i = 1, 2, ..., n, i \leq j\}$ are i.i.d. conditioning on random link function $g : [0, 1]^2 \rightarrow [0, 1]$. Thus for fixed row (column) of $\mathbf{A}$, $\{A_{i1}, A_{i2}, ..., A_{in}\}, i \in \{1, 2, ..., n\}$ are conditionally i.i.d. on random link function g.

Lemma 2.2 (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 $\{A_{ij} : i = 1, 2, ..., n, i \leq j\}$ need cautions

From the above Lemma 2.1, 2.2, we can also prove exchangeability and conditional i.i.d. of diffusion maps at each time point.

Lemma 2.3 (Exchangeability and iid of U)**.** Assume that a connected, undirected and un-weighted graph $\mathbf{G}$ is a graphon or graphex, i.e. any exchangeable random graph from an infinite

graph. Then its transition probability P_{ij} so thus diffusion maps at fixed time t also exchange-able conditional on link function of graph. Furthermore, by [de Finetti]'s TheoremA.1, we can say that such diffusion maps at t are conditionally i.i.d given random probability measure η on U_t and random link function g.

Lemma 2.3 above provides us i.i.d. one-parameter family of $\{\mathbf{U}_t\}_{t \in \mathbb{N}}$ conditional on a random probability measure of $\mathbf{U}_t$ and a random link function of g.

2.3 Multiscale Generalized Correlation

2.3.1 [MGC1]

Relationship between network and nodal attributes often exhibits local or nonlinear properties. Moreover, dimension of spectrum of network(q) increases as a sample size increases. Unfortunately, widely used correlation measures often fail to capture nonlinear associations or multivariate associations. Szekely et al. (2007)[14] extended pairwise constructed generalized correlation coefficient and developed a novel statistics called distance correlation (dCov) 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 random samples $\{(x_i, y_i) : x_i \in \mathbb{R}^q, y_i \in \mathbb{R}^m, i = 1, ..., n\}$. Define $C_{ij} = \| x_i - x_j \|$ and $D_{ij} = \| y_i - y_j \|$ for $i, j = 1, ..., n$, where $\| \cdot \|$ denotes Euclidean distance defined on any vectors. Distance correlation (dCor) is defined via distance covariance (dCov) $\mathcal{V}_n^2$ of $\boldsymbol{X}$ and $\boldsymbol{Y}$, which is the following:

$$\mathcal{V}_n^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{C}_{ij} \tilde{D}_{ij} \tag{10}$$

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

$$\mathcal{R}_n^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\mathcal{V}_n^2(\boldsymbol{X}, \boldsymbol{Y})}{\sqrt{\mathcal{V}_n^2(\boldsymbol{X}, \boldsymbol{X})\mathcal{V}_n^2(\boldsymbol{Y}, \boldsymbol{Y})}} \tag{11}$$

On the other hand, 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 Szekely et al. (2013)[15]. However, dCov and even MCorr still perform not very well in existence of various nonlinear dependency and under existence of outliers ([16]). Out of this concern, Cencheng at al (2016) [16] developed Multiscale Generalized Correlation (MGC) by adding

local scale on correlation coefficients. Multiscale version of distance covariance $\{\mathcal{V}^{*2}_n\}_{kl}$ is defined as following :

$$\mathcal{V}^{*2}_n(\boldsymbol{X}, \boldsymbol{Y})_{kl} = \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{C}_{ij} \tilde{D}_{ij} I\big(r(C_{ij}) \leq k\big) I\big(r(D_{ij})) \leq l\big), \quad k, l = 1, 2, ..., n \qquad (12)$$

,where $r(C_{ij})$ $(r(D_{ij}))$ is a rank $\mathbf{x}_i$ $(\mathbf{y}_i)$ relative to $\mathbf{x}_j$ $(\mathbf{y}_j)$. That is c

Local scale of test statistics

(brief explanation on neighborhood choice)

2.3.2 [MGC2]

Returning to the problem of network setting, the fundamental problem is in measuring all types of dependence between $\boldsymbol{G}$ and $\boldsymbol{X}$ via `MGC`, we are required a *i.i.d.* vertex-wise coordinates of which Euclidean distance measures a distance between them. You might first propose directly using a column of an adjacency matrix so that we have a n-pair of observations $\big\{\big(\boldsymbol{A}_{i\cdot}, \boldsymbol{X}_i\big) : \boldsymbol{A}_{i\cdot} = (A_{i1}, ..., A_{in}), \boldsymbol{X}_i \in \mathbb{R}^m, i = 1, ..., n\big\}$. In an undirected graph, $\{\mathbf{A_{i\cdot}}\}$ cannot be independent. Even if it is in e.g. directed graph, Euclidean distance between $\{\boldsymbol{A}_{i\cdot} : i = 1, ..., n\}$ is not a proper metric over network space. Let us introduce a simple example. Let a given network $\boldsymbol{G}$ having 8 nodes be an unweighted, directed network and possibly having self-loop. Let $\boldsymbol{A}$ be its 8×8 binary adjacency matrix. Assume `Node 1`, `Node 4` and `Node 8` have the following row entries:

$$\begin{aligned}
\boldsymbol{A}_{1\cdot} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \\
\boldsymbol{A}_{4\cdot} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad (13) \\
\boldsymbol{A}_{8\cdot} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

,which results $\| \boldsymbol{A}_{1\cdot} - \boldsymbol{A}_{4\cdot} \|^2 = 4$, $\| \boldsymbol{A}_{1\cdot} - \boldsymbol{A}_{8\cdot} \|^2 = 7$, and $\| \boldsymbol{A}_{4\cdot} - \boldsymbol{A}_{8\cdot} \|^2 = 3$. Accordingly, $\| \boldsymbol{A}_{4\cdot} - \boldsymbol{A}_{8\cdot} \| < \| \boldsymbol{A}_{1\cdot} - \boldsymbol{A}_{4\cdot} \|$. However, you can easily see that this does not make sense because

`Node 4` and `Node 8` are connected each other only through `Node 1`. Therefore instead of using an adjacency matrix directly, we are considering embedding a vertex $v \in V(\boldsymbol{G})$ into its diffusion map of $\boldsymbol{U}$ and apply Euclidean distance metric, which is exactly a diffusion distance. As explained before, its Euclidean distance takes into account all possible paths between every pair of node 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 in [Appendix]).

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

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

Thanks to these properties of diffusion maps, we have better interpretation of its Euclidean distance so that we will use it to the distance matrix in `MGC`.

2.3.3 [MGC3]

We have discussed one-parameter family *i.i.d.* representation of network structures, called diffusion maps, and also discussed when given *i.i.d.* pair of observations of two variables, how distance correlation and its multiscale version test independence between these two. Now it is time to combine these two to test independence between network structures and nodal attributes efficiently and effectively.

Theorem 2.4 (MGC of testing independence)**.** Assume that a connected, undirected and unweighted graph $\mathbf{G}$ is an exchangeable graph. Assume that we are given n-pair of observations $\{(u_i^{(t)}, x_i) : i = 1, 2, ..., n, t \in \mathbb{N}\}$. Then $u_t(i) \overset{i.i.d.}{\sim} f_{U_t}(\eta, g)$, $t \in \mathbb{N}$ and $x_i \overset{i.i.d.}{\sim} f_X$, where $f_{U_t}(\eta, g)$ are conditional distribution function given a link function g and a random probability

measure η_t of U_t. Then MGC applied to these pair of data is theoretically consistent against all dependent alternatives in testing :

$$H_0 : f_{U_t \cdot X} = f_{U_t} \cdot f_X$$

Since U provides a configuration of vertices in G, the above hypothesis implies testing independence between the configuration of vertices in network space and in attribute space, but as a function of a link function g and a random function (variable) of η of $\mathbf{U}$.

Remark 1. Roughly speaking, we can say that diffusion maps are i.i.d. function of a link function g and a random function of $\mathbf{U}$, η. Thus testing independence between conditional U and X can be considered as testing independence between $f(g, \eta)$ and X. A link function g concerns the distribution of edges and a random function η concerns nature distribution of diffusion maps. Our testing basically examines whether how edges are constructed and how diffusions(propagation) process are correlated to nodal attributes.

3 Simulation Study

In simulation studies presented in this paper, we make a comparison of estimated testing power across various multivariate independence test statistics: `MGC`, `dCov`, Heller-Heller-Gorfine (`HHG`)[11], 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 a additive model of latent factors, which is mostly targeted by `FH`.

3.1 [sim1] Stochastic Block Model

- *[sim1] present simplest examples of stochastic block model (graphon/graphex)*

We mentioned in the Introduction that latent network model is very common followed by the assumption of local independence. Stochastic Block Model (SBM) is one of the most popular and also useful network generative model, especially as a tool for community detection ([Karrerl and Newman]). The SBM, in the simplest setting, assumes that each of n vertices in graph $\boldsymbol{G}$ belongs to one of $K \in \mathbb{N}(\leq n)$ blocks or groups. Block affiliation is important because the probability of having edges between a pair of vertices depends on which blocks they are in.

Assume that the latent variable $Z_1, Z_2, ..., Z_n \overset{i.i.d.}{\sim} Multinomial\big(\pi_1, \pi_2, ..., \pi_K\big)$. Then the uppwer triangular entries of $\mathbf{A}$ are independent with $A_{ij} \overset{i.i.d.}{\sim} Bern\big(\sum_{k,l=1}^{K} p_{kl}I\big(Z_i = k, Z_j = l\big)\big), \forall i < j$.

Let $W_1, W_2, ..., W_n \overset{i.i.d.}{\sim} Unif[0,1]$. Then such $\mathbf{A}$ can be represented with respect to $\{W_i\}$ as follows:

$$A_{ij} \overset{i.i.d.}{\sim} Bern\big(g(W_i, W_j)\big), \forall i < j \tag{15}$$

, where $g\big(W_i, W_j\big) = \sum_{k,l=1}^{K} p_{kl}I\big(W_i \in [\sum_{j=1}^{k-1}\pi_j, \sum_{j=1}^{k}\pi_j], W_j \in [\sum_{j=1}^{l-1}\pi_j, \sum_{j=1}^{l}\pi_j]\big)$

21

3.1.1 [sim1.1] simplest two block model

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

$$Z_i \sim \left\{ \begin{array}{ll} Bern(0.6) & X_i = 0 \\ Bern(0.4) & X_i = 1 \end{array} \right. \tag{16}$$

$$A_{z_i,z_j} \sim Bern \left[\begin{array}{cc} 0.4 & 0.1 \\ 0.1 & 0.4 \end{array} \right]$$

figure which compares power of MGC/dCov/hhg/FH

3.1.2 [sim1.2] simplest three block model

$$X_i \overset{i.i.d}{\sim} Multi(1/3, 1/3, 1/3), i = 1, 2, ..., n$$

$$Z_i \sim \left\{ \begin{array}{ll} Multi(1/2, 1/4, 1/4) & X_i = 1 \\ Multi(1/4, 1/2, 1/4) & X_i = 2 \\ Multi(1/4, 1/4, 1/2) & X_i = 3 \end{array} \right. \tag{17}$$

$$A_{z_i,z_j} \sim Bern \left[\begin{array}{ccc} 0.5 & 0.2 & 0.3 \\ 0.2 & 0.5 & 0.2 \\ 0.3 & 0.2 & 0.5 \end{array} \right]$$

comparison to patter of power of MGC/dCov/hhg/FH from "simple" three block graph -

figures in appendix

explain briefly why local optimal performs better in this case

figure which compares power of MGC/dCov/hhg/FH

3.2 [sim]

3.2.1 [sim2.1] sparse version of block model

figure which compares degree distribution of SBM vs. sparse graph

simulation study explained in Veitch and Roy, 2015 Sample a (latent) unit rate Poisson

process Π on $\theta \times \vartheta \in [0, \nu] \times [0, c]$.

$$N_\nu \sim Poi(c\nu)$$

$$\{\theta_i\}\big|N_\nu \overset{iid}{\sim} Uniform[0, \nu]$$

$$\{\vartheta_i\}\big|N_\nu \overset{iid}{\sim} Uniform[0, 1] \tag{18}$$

$$\{X_i\}|N_\nu \overset{ind}{\sim} Normal(\vartheta_i, 0.1)$$

$$(\theta_i, \theta_j)\big|W, \vartheta_i, \vartheta_j \overset{ind}{\sim} Bernoulli(W(\vartheta_i, \vartheta_j))$$

3.2.2 [sim2.2] comparison to Fosdick's method

Fosdick and Hoff proposed a approach of jointly modelling network and its attributes, where networks represent additional structure via sender-specific(or row-specific) and receiver-specific(or column-specific) latent factors.

As a connection to RDPG, this model can also represented as a product of combined latent factor $\tilde{u}$ and $\tilde{v}$.

(factor additive model)

$$u_i \overset{i.i.d}{\sim} Uniform[0, 1]$$

$$A_{ij} \overset{i.i.d}{\sim} Bern\big((1 - u_i)^2 \times (1 - u_j)^2\big) \tag{19}$$

Additionally some noise could be introduced as well due to artificially added edges to construct a connected graph.

4   Real Data Examples

4.1   c.elegans

5   Discussions

5.1   [dis1]

Throughout this study, we demonstrate that multiscale network test statistic to test network independence performs well in diverse settings, being supported by thorough theory on distance correlation and diffusion maps. Testing independence is often the very first step in investigating relationship between network topology and nodal variables 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 due to both neighborhood choice and time spent in diffusion processes provides us a hint on a latent dependence structure as well.

5.2   [dis2]

- choosing the optimal time t, optimal choice of neighborhood (k, l) might be burdensome.

- real network data cannot satisfy the condition of exchangeability

- we often do not have any idea about a latent function of g or η.

5.3   [dis3]

- We can use it to study latent structure of network (graph) by exploring testing p-value maps across diffusion times.

- Other than testing network independence to nodal attributes, we can use MNT to test independence between a pair of networks.

- This novel, nonparametric testing would help to understand complex data, i.e. network, which could be entangled with variables that we are interested in.

A   appendix

A.1   [apd1]

- *[apd1] additional plots*

$$X_i \stackrel{i.i.d}{\sim} Multi(1/3, 1/3, 1/3), i = 1, 2, ..., n$$

$$Z_i \sim \begin{cases} Multi(1/2, 1/4, 1/4) & X_i = 1 \\ Multi(1/4, 1/2, 1/4) & X_i = 2 \\ Multi(1/4, 1/4, 1/2) & X_i = 3 \end{cases}$$

$$A_{z_i, z_j} \sim Bern \begin{bmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix}$$

$$(20)$$

A.2   [apd2]

- *[apd2] (proof of) 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 $\mathbf{X}$. 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

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

$$(21)$$

26

where $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij})_{i,j>i \in mathbbN}$ 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] TheoremA.2, a random array (A_{ij}) is jointly exchangeable if and only if it can be represented as follows :

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

$$(A_{ij}) \overset{d}{=} Bern(g(W_i, W_j))$$

, 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 $i \leq j$, $i, j = 1, ..., n$:

$$
\begin{aligned}
P(A_{ij} = a_{ij}) &= \int P(A_{ij}|w_i, w_j) Pr(W_i = w_i) Pr(W_j = w_j) dw_i dw_j \\
&= \int_0^1 \int_0^1 g(w_i, w_j)^{a_{ij}} (1 - g(w_i, w_j))^{1-a_{ij}} dw_i dw_j
\end{aligned}
\tag{22}
$$

Then within each row, adjacent elements are independent and also identically distributed ,i.e. for fixed row index to $i \in \{1, 2, ..., n\}$,

$$P(A_{i1} = a_{i1}, A_{i2} = a_{i2}, ..., A_{in} = a_{in}) = \prod_{j=1}^n P(A_{ij} = a_{ij})$$

$\square$

Proof of Lemma 2.2. Based on Kallenberg and Exchangeable Graph (KEG) frameworks, introduced in [Veitch and Roy], a random array (A_{ij}) is jointly exchangeable if and only if it can be represented as follows : there is a random function $g : [0,1]^2 \rightarrow [0,1]$ such that

$$(A_{ij}) \overset{d}{=} (A_{v_i,v_j}) \overset{d}{=} Bern(g(\vartheta_i, \vartheta_j)) \tag{23}$$

, where $v_i \overset{i.i.d.}{\sim} Poisson(1), \vartheta_i \overset{i.i.d.}{\sim} Poisson(1), v_i \leq \nu, i = 1, 2, ..., n$, for some pre-specified $\nu > 0$ so that finite size graphs can include vertices only if they participate in at least one edges.

Thus if $\mathbf{A}$ is an adjacency matrix of an undirected, exchangeable network, for any $i \leq j$, $i, j = 1, ..., n$:

$$
\begin{aligned}
P(A_{ij} = a_{ij}) &= \int P(A_{ij}|v_i, v_j) Pr(V_i = v_i) Pr(V_j = v_j) Pr(\vartheta_i = \vartheta_i) Pr(\vartheta_j = \vartheta_j) dv_i dv_j d\vartheta_i d\vartheta_j \\
&= \int_0^\tau \int_0^\tau \int_0^\infty \int_0^\infty g(\vartheta_i, \vartheta j)^{a_{ij}} \left(1 - g(\vartheta_i, \vartheta_j)\right)^{1-a_{ij}} \\
&\quad \times dPois_1(x_1) \times dPois_1(x_2) \times dPois_1(x_3) \times dPoi_1(x_4) dx_1 dx_2 dx_3 dx_4.
\end{aligned}
\tag{24}
$$

$\square$

Proof of Lemma 2.3. We have shown that for fixed time t, diffusion distance is defined as an Euclidean distance of diffusion maps. Diffusion map is represented as follows :

$$
\boldsymbol{U}_t(i) = \begin{pmatrix} \lambda_1^t \phi_1(i) & \lambda_2^t \phi_2(i) & \cdots & \lambda_q^t \phi_q(i) \end{pmatrix} \in \mathbb{R}^q.
\tag{25}
$$

,where $\Phi = \Pi^{-1/2} \Psi$ and $\mathbf{Q} = \boldsymbol{\Psi \Lambda \Psi}^T = \boldsymbol{\Pi^{1/2} P \Pi^{-1/2}}$. Thus $\mathbf{P \Pi^{-1/2} \Psi = \Pi^{-1/2} \Psi \Lambda}$.

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 $\phi_r = \left(\frac{\psi_r(1)}{\sqrt{\pi(1)}} \quad \frac{\psi_r(2)}{\sqrt{\pi(2)}} \quad \cdots \quad \frac{\psi_r(n)}{\sqrt{\pi(n)}} \right)$. Therefore to guarantee exchangeability (or i.i.d.) of $\mathbf{U}_t$, it suffices to show exchangeability (or i.i.d.) of $\mathbf{P}$.

Assume joint exchangeability of $\mathbf{G}$, i.e. $(A_{ij}) \overset{d}{=} (A_{\sigma(i)\sigma(j)})$. Since A_{ij} is binary, $\frac{A_{ij}}{\sum\limits_{ij} A_{ij}} = \frac{A_{ij}}{1 + \sum\limits_{l \neq j} A_{il}}$. Moreover, A_{ij} and $(1 + \sum\limits_{l \neq j} A_{il})$ are independent given its link function g, and $A_{\sigma(i)\sigma(j)}$ and $(1 + \sum\limits_{l \neq j} A_{\sigma(i)\sigma(l)})$ are independent also given g.

Then the following joint exchangeability of transition probability holds:

$$
(P_{ij}) = \left(\frac{A_{ij}}{1 - A_{ij} + \sum\limits_{j=1}^n A_{ij}} \right) \overset{d}{=} \left(\frac{A_{\sigma(i)\sigma(j)}}{1 - A_{\sigma(i)\sigma(j)} + \sum\limits_{\sigma(j)=1}^n A_{\sigma(i)\sigma(j)}} \right) = (P_{\sigma(i)\sigma(j)})
\tag{26}
$$

Thus, transition probability is exchangeable. This results exchangeable eigenfunctions $\{\Phi(1), \Phi(2), , ..., \Phi(n)\}$ where $\Phi(i) := \begin{pmatrix} \phi_1(i) & \phi_2(i) & \cdots & \phi_q(i) \end{pmatrix}^T$. 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 $\mathbf{U}(t) = \{U_1^{(t)}, U_2^{(t)}, ..., U_n^{(t)}\}$ are conditionally independent on a random probability measure η. □

Proof of Theorem 2.4. □

Proof of corollary 2.3.1. [Triangle inequality of diffusion distance] □

Let $x, y, z \in V(G)$.

$$
\begin{aligned}
D_t^2(x, z) &= \sum_{w \in V(G)} \left(P^t(x, w) - P^t(z, w)\right)^2 \frac{1}{\pi(w)} \\
&= \sum_{w \in V(G)} \left(P^t(x, w) - P^t(y, w) + P^t(y, w) - P^t(z, w)\right)^2 \frac{1}{\pi(w)} \\
&= \sum_{w \in V(G)} \left(P^t(x, w) - P^t(y, w)\right)^2 \frac{1}{\pi(w)} + \sum_{w \in V(G)} \left(P^t(y, w) - P^t(z, w)\right)^2 \frac{1}{\pi(w)} \\
&\quad + 2 \sum_{w \in V(G)} \left(P^t(x, w) - P^t(y, w)\right)\left(P^t(y, w) - P^t(z, w)\right) \frac{1}{\pi(w)} \\
&= D_t^2(x, y) + D_t^2(y, z) + 2 \sum_{w \in V(G)} \left(P^t(x, w) - P^t(y, w)\right)\left(P^t(y, w) - P^t(z, w)\right) \frac{1}{\pi(w)}
\end{aligned}
$$
(27)

Thus it suffices to show that

$$
\sum_{w \in V(G)} \left(P^t(x, w) - P^t(y, w)\right)\left(P^t(y, w) - P^t(z, w)\right) \frac{1}{\pi(w)} \leq D_t(x, y) \cdot D_t(y, z).
$$
(28)

Let $a_w = \left(P^t(x, w) - P^t(y, w)\right)\sqrt{1/\pi(w)}$ and $b_w = \left(P^t(y, w) - P^t(z, w)\right)\sqrt{1/\pi(w)}$. Then

the above inequality is equivalent to :

$$\sum_{w \in V(G)} a_w \cdot b_w \le \sqrt{\sum_{w \in V(G)} a_w^2 \cdot \sum_{w \in V(G)} b_w^2}. \tag{29}$$

,which is true by Cauchy-Schwarz inequality.

References and Notes

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

[2] Caron, F., & Fox, E. B. (2014). Sparse graphs using exchangeable random measures. arXiv preprint arXiv:1401.1137.

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

[4] Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. Journal of the american Statistical association, 97(460), 1090-1098.

[5] Kallenberg, O. (1990). Exchangeable random measures in the plane. Journal of Theoretical Probability, 3(1), 81-136.

[6] Austin, A., Linkletter, C., & Wu, Z. (2013). Covariate-defined latent space random effects model. Social Networks, 35(3), 338-346.

[7] Mantzaris, A. V., Bassett, D. S., Wymbs, N. F., Estrada, E., Porter, M. A., Mucha, P. J., ... & Higham, D. J. (2013). Dynamic network centrality summarizes learning in the human brain. Journal of Complex Networks, 1(1), 83-92.

[8] Sporns, O., Honey, C. J., & Ktter, R. (2007). Identification and classification of hubs in brain networks. PloS one, 2(10), e1049.

[9] Lazarsfeld, P. F, & Henry, N. W. (1968). Latent structure analysis. New York: Houghton, Mifflin.

[10] Coifman, R. R., & Lafon, S. (2006). Diffusion maps. Applied and computational harmonic analysis, 21(1), 5-30.

[11] Heller, R., Heller, Y., & Gorfine, M. (2012). A consistent multivariate test of association based on ranks of distances. Biometrika, ass070.

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

[13] Veitch, V., & Roy, D. M. (2015). The class of random graphs arising from exchangeable random measures. arXiv preprint arXiv:1512.03099.

[14] Szkely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. The Annals of Statistics, 35(6), 2769-2794.

[15] Szkely, G. J., & Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension. Journal of Multivariate Analysis, 117, 193-213.

[16] Cencheng at al, Revealing the structure of dependency between multimodal datasets via multiscale generalized correlation.

[17] Lyons, R. (2013). Distance covariance in metric spaces. The Annals of Probability, 41(5), 3284-3305.

[18] Lee, J. R. (2006). Distance scales, embeddings, and metrics of negative type. University of California, Berkeley.

[19] Tang, Minh, and Michael Trosset. "Graph metrics and dimension reduction." Indiana University, Indianapolis, IN (2010).

[20] Lafon, S., & Lee, A. B. (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.

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