Testing independence in networks via family of network metrics

Youjin Lee (ylee160@jhu.edu) *
Department of Biostatistics, Johns Hopkins School of Publich Health

Abstract

Propelled by increasing demand and supply in network data, investigating whether network structures are associated with the attributes of interest has been an important concern in natural or social science. We consider the problem of network dependence, which refers to any types of dependence between network topology and its nodal attributes and propose the method to test network independence. However due to the interdependency in constructing network, standard independence test cannot be directly applied and most of the network model-based tests presume globally persistent dependency patterns. To overcome these challenges, we propose a nonparametric multiscale test statistic which is robust to both high dimensionality and nonlinearity by utilizing a family of network geometries. Our simulation studies demonstrate the outstanding performance of the method under various circumstances.

Keywords: Distance correlation, Network dependence, Diffusion maps, Exchangeable graph

1 Introduction

Statisticians have long considered the problem of revealing the relationship between two data sets. Above all determining the existence any association or any dependence would be the first step in characterizing the relationship. As types of data have diversified or dimension of the data has increased, various forms of multivariate independence tests have been suggested [Taskinen et al., 2005, Heller et al., 2012, Székely et al., 2007]. We consider independence test upon non-traditional but ubiquitous dataset of *network* which is very likely to possess the properties of both high dimensionality and nonlinearity. Network, formally defined as a collection of nodes and edges, has been suffering from a dearth of proper analysis due to its distinct way to be constructed. In this paper we define any kinds of dependency between network topology and nodal attributes as *network dependence* and propose the method to test network independence.

The literature on identifying dependency between network and nodal attributes has primarily focused on their relationship explained only by network model under the boundary of model assumption

*This is a joint work with Cencheng Shen (cshen6@jhu.edu) and Joshua T. Vogelstein (jovo@jhu.edu), Johns Hopkins University.

[Wasserman and Pattison, 1996, Howard et al., 2016]. A fundamental difficulty of model-based independence tests comes from the fact that not all networks exhibit the structures described by known network models. Fosdick and Hoff [2015] overcome this issue by estimating network factors which are believed to embody each node's locations in network space. Their network models, however, still rely on the assumption that all the nodes in network would follow the same pattern of dependence – subject to additive and multiplicative effect.

Different from a random vector, network or equivalently graph involves a particular construction which we should take into account. Throughout this paper, we assume that we are given an unweighted and undirected network without self-loop, comprised of $n(\in \mathbb{N})$ nodes. An adjacency matrix of this given network, denoted by $\mathbf{A} = \{A_{ij} : i, j = 1, .., n\}$, is to formalize the relational data of network, where $A_{ij} = 1$ if node i and node j are adjacent each other and zero otherwise. Let us define a m-variate ($m \in \mathbb{N}$) random variable associated with each node, i.e. nodal attributes, $\mathbf{X} \in \mathbb{R}^m$ which we are interested in. We first have to consider increasing amount of information inherent in network data as the number of nodes increases, which might lead to diverse patterns in dependency as well. In addition, by its definition, an adjacency matrix $\mathbf{A}$ inherits dependency among its columns and rows, so thus it cannot enjoy the traditional setting based on a random vector. To overcome these challenges, we propose applying distance-based statistic called multiscale generalized correlation (`MGC`) [Shen et al., 2016] into testing network independnece along with the network geometries derived from random walk on graph. We are going to elaborate the statistic and demonstrate its validity for a family of graphs in Section 2. In Section 3, simulation results demonstrate the best performance of our method compared to the existing under various circumstances.

2 Methods

2.1 Distance-based Independence Test

Székely et al. [2007] proposed a distance-based statistic having a marvelous closed form called distance correlation (`dCorr`). This distance-based multivariate independence test starts from the assumption that we are given $n \in \mathbb{N}$ pairs of $i.i.d$ random vectors $\big(\mathbf{W}, \mathbf{Y}\big) = \{(\mathbf{w}_i, \mathbf{y}_i) : \mathbf{w}_i \in \mathbb{R}^q; \mathbf{y}_i \in \mathbb{R}^m; i = 1, ..., n\}$. Define distance matrices of $C_{ij} = \| \mathbf{w}_i - \mathbf{w}_j \|$ and $D_{ij} = \| \mathbf{y}_i - \mathbf{y}_j \|$ for $i, j = 1, 2, \ldots, n$, where $\| \cdot \|$ indicates Euclidean distance. Distance correlation (`dCorr`) is defined via distance covariance (`dCov`) $\mathcal{V}_n^2$

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

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

where $\tilde{C}$ and $\tilde{D}$ is doubly-centered C and D by its column mean and row mean respectively. A modified distance covariance (`mCov`) and a modified distance correlation (`mCorr`) for testing high dimensional random vectors were also proposed in Székely and Rizzo [2013]. However, neither `dCorr` nor ever `mCorr` still performs very well in the existence of various nonlinear dependency and under the existence of outliers as well [Shen et al., 2016]. Out of this concern, Shen et al. [2016] proposed Multiscale Generalized Correlation (`MGC`) via adding local scale in a sense of nearest neighbors on correlation coefficients. Multiscale version of distance covariance $\{\mathcal{V}^{*2}_n\}_{kl}$ is defined as following :

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

where $r(C_{ij})$ $(r(D_{ij}))$ denotes a rank of $\mathbf{w}_i$ $(\mathbf{y}_i)$ relative to $\mathbf{w}_j$ $(\mathbf{y}_j)$, i.e. $r(C_{ij}) = k$ means w_i is $w_j^{'}$s k-nearest neighbor. Based on this set of statistics, `MGC` finds the *best statistic* which exhibits the largest correlation between the two data sets without inducing multiple testing problems. It has already been shown that this local scaled statistic performs no worse than `dCorr` and that it actually results in improved sensitivity to nonlinear dependence than `dCorr`.

Now back to the network independence test, we are ready to derive the Euclidean distance of $\mathbf{X}$, e.g. D; while how to construct the distance matrix corresponding to network structure is still questionable. We are required $i.i.d$ node-specific coordinates of which Euclidean distance as C in statistics 2 reflects a network-based distance between nodes. We can first conjecture directly using an Euclidean distance of an adjacency matrix $\mathbf{A}$. The Euclidean distance from each column of adjacency matrix, however, may not satisfy the requirement for consistency since columns of $\mathbf{A}$ are often dependent each other.

2.2 Family of Network Independence Test Statistics

In order to satisfy the requirement of being $i.i.d$, we are going to restrict applicable network to a certain family of graphs and suggest valid graph geometries that furnish us useful metrics. A graph $\mathbf{G}$ is called exchangeable if and only if its adjacency matrix $\mathbf{A}$ is jointly exchangeable [Orbanz and Roy, 2015], i.e. for every permutation σ of n, $(A_{ij}) \stackrel{d}{=} (A_{\sigma(i)\sigma(j)})$. Even though exchangeability itself cannot guarantee being *i.i.d*, thanks to the *de Finetti*'s representation theorem and *Aldous-Hoover theorem*, exchangeable graph, often called *graphon* [Lovász and Szegedy, 2006], can be defined through conditionally *i.i.d* edge distribution [Chan et al., 2013]. We then take advantage of this *i.i.d* expression which draws *i.i.d* network geometries called *diffusion maps*.

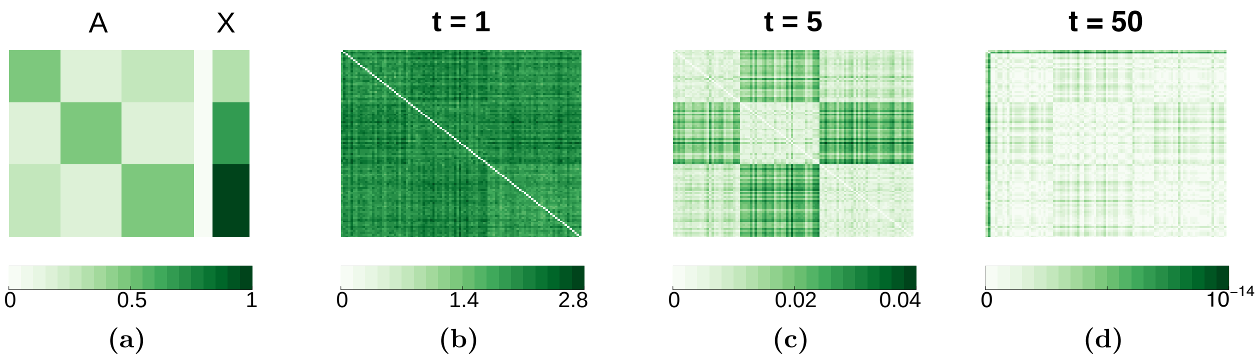

Figure 1: Figure (a) shows data generating probability of an adjacency matrix **A** and nodal attributes **X**. Diffusion matrix, as a proposed network metric, provides one-parameter family of network-based distances where as time goes by the pattern shown in the distance matrix changes, and at time point $t = 5$, distance matrix (c) illustrates most clear block structures and at the same time it exhibits most dependence to distance matrix of **X**.

Coifman and Lafon [2006] proposed diffusion maps as a meaningful multiscale geometries which is defined by eigenvectors of Markov matrix constructed over a connected network. In the process of constructing such Markov matrix, we basically run random walks by iterating transition matrix and diffusion maps accordingly locates each node's position at each iteration time. Distance between each pair of nodes, defined as a *diffusion distance*, then can be derived from an Euclidean distance of such diffusion maps. Each of diffusion distance, i.e. C_t, can also be represented using a discrete set of real nonzero eigenvalues $\{\lambda_r\}$ and eigenvectors $\{\phi_r\}$ of a transition matrix [Coifman and Lafon, 2006, Lafon and Lee, 2006].

$$C_t^2[i,j] := \parallel \mathbf{U}_t(i) - \mathbf{U}_t(j) \parallel \quad i,j = 1,2,\ldots,n, \tag{3}$$

where $\mathbf{U}_t(i) = \begin{pmatrix} \lambda_1^t \phi_1(i) & \lambda_2^t \phi_2(i) & \ldots & \lambda_q^t \phi_q(i) \end{pmatrix}^T \in \mathbb{R}^q$ is a diffusion map at time t. As diffusion time t increases, distance matrix C_t is more likely to take into account distance between two nodes which are relatively difficult to reach each other. Figure 1 shows three exemplary distance matrices among whole one-parameter family of distance $\{C_t : t \in \mathbb{N}\}$. Compared to adjacent relation or geodesic distance which are two extremes, diffusion distance well reflects the connectivity since it takes into account every possible path between the two nodes. We prove that each diffusion map provides *i.i.d* multivariate coordinates for each node under jointly exchangeable graph. Lemma 2.1 below provides us with *i.i.d* one-parameter family of $\{\mathbf{U}_t\}_{t\in\mathbb{N}}$.

Lemma 2.1 (Exchangeability and *i.i.d* of diffusion maps $\mathbf{U}_t$)**.** Assume that a connected, undirected and unweighted graph $\mathbf{G}$ is an exchangeable random graph. Then its transition probability so thus diffusion maps at fixed time t is also exchangeable, conditioned on underlying distribution of graph. Furthermore, by *de Finetti's Theorem*, such diffusion maps at t, $\mathbf{U}_t(i) = \begin{pmatrix} \lambda_1^t\phi_1(i) & \lambda_2^t\phi_2(i) & \dots & \lambda_q^t\phi_q(i) \end{pmatrix}^T$, are conditionally *i.i.d* given its underlying distribution.

If a set of diffusion distances are applicable as a proper Euclidean distance in a distance-based test statistic 1, we are able to dispense with obstacles in testing network independence. First of all in a general setting, assume that we have a finite sample of infinitely exchangeable sequence $(\mathbf{W}, \mathbf{Y}) = \{(\mathbf{w}_i, \mathbf{y}_i); i = 1, 2, \dots, n\}$, which is identically distributed as $(\mathbf{w}, \mathbf{y})$ with finite second moment.

Theorem 2.2. Suppose that we are given n pairs of exchangeable observations $(\mathbf{W}, \mathbf{Y}) = \{(\mathbf{w}_i, \mathbf{y}_i); i = 1, 2, \dots, n\}$ having finite second moment. Assume $\mathbf{w}_i \overset{i.i.d}{\sim} f_{\mathbf{w}}$ and $\mathbf{y}_i \overset{i.i.d}{\sim} f_{\mathbf{y}}$ given underlying distribution for $i = 1, 2, \dots, n$. Then

$$\mathcal{V}_n^2(\mathbf{W}, \mathbf{Y}) \quad \longrightarrow 0 \quad \text{ as } n \to \infty \tag{4}$$

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

Note that if $\{\mathbf{w}_i : i = 1, 2, \dots, n\}$ are *i.i.d*, they are also exchangeable. Thus estimated latent network factors, which are assumed *i.i.d* by Fosdick and Hoff [2015] can also be applied to Theorem 2.2. We already have shown that even under undirected network, diffusion maps remain exchangeable at each diffusion time point t.

Theorem 2.3. For n-pair of diffusion map and *i.i.d* nodal attributes $\{(\mathbf{u}_t(i), \mathbf{x}_i) : i = 1, 2, \dots, n\}$ in a jointly exchangeable graph $\mathbf{G}$, we have $\mathbf{u}_t(i) \overset{i.i.d}{\sim} f_{U(t)}$ and $\mathbf{x}_i \overset{i.i.d}{\sim} f_{\mathbf{X}}$ conditioned on underlying distribution of graph. Then `MGC` is consistent in testing network independence with null of $H_0 : f_{\mathbf{U}^{(t)} \cdot \mathbf{X}} = f_{\mathbf{U}^{(\mathbf{t})}} \cdot f_{\mathbf{X}}$. In particular, the consistency also holds for using the estimated latent network factors [Fosdick and Hoff, 2015] or an adjacency matrix of directed network as a distance matrix.

2.3 Measure for Node Contribution

On the other hand, some nodes often exert more reliance on their attributes than the others. Here we suggest the measure of node's contribution to detecting dependence as a byproduct of `MGC` statistic. Let (k^*, l^*) be the optimal neighborhood choice in distance matrix (C, D) respectively. Denote the contribution of node $v \in V(G)$ to the testing statistic by $c(\cdot) : v \to \mathbb{R}$

$$c(v) \propto \sum_{j=1}^{n} \tilde{C}_{jv} \tilde{D}_{jv} I\big(r(C_{jv}) \leq k^*\big) I\big(r(D_{jv}) \leq l^*\big), \tag{5}$$

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

3 Simulation Study

In our simulation studies, we make a comparison between empirical testing power between `MGC`, `mCorr`, Heller-Heller-Gorfine (`HHG`) [Heller et al., 2012], and likelihood ratio test of Fosdick and Hoff (`FH`) [Fosdick and Hoff, 2015]. We use type I error of $\alpha = 0.05$ and obtain p-values from each simulated network via permutation test. All the simulation models can be illustrated by joint distribution of adjacent matrix $\mathbf{A}$, nodal attributes $\mathbf{X}$, and latent variable $\mathbf{Z}$ that explains dependence structure between $\mathbf{A}$ and $\mathbf{X}$. You can find the detailed simulation models in the Appendix 6.2. We introduced a popular network model of Stochastic Block Model (SBM) as our main simulated networks.

3.1 Stochastic Block Model

We present the SBM with $K = 3$ blocks (model 15) where block affiliation for each node is correlated with its attributes X. Figure 2 illustrates superior performance of `MGC` as a test statistic combined with diffusion maps (`DF`) as a network metric. To simply represent model 15, we have

$$E(A_{ij}|X_i, X_j) = 0.5I(|X_i - X_j| = 0) + 0.2I(|X_i - X_j| = 1) + 0.3I(|X_i - X_j| = 2). \tag{6}$$

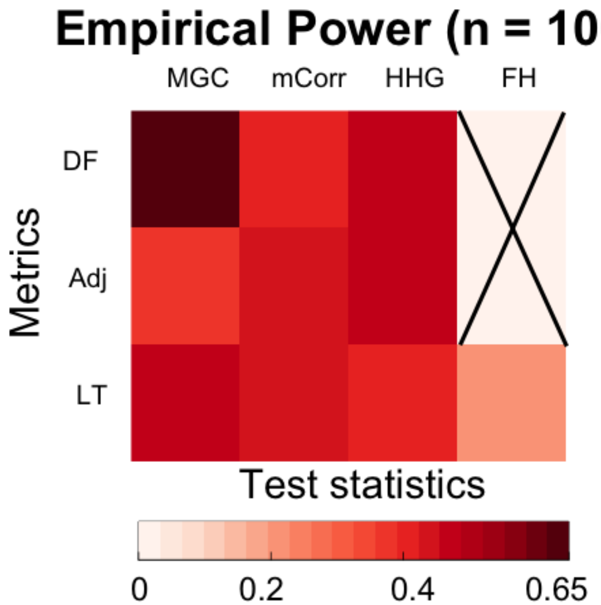

Figure 2: This power heatmap illustrates the superior power of multiscale generalized correlation (`MGC`) under diffusion distance matrix (`DF`) in three SBM (model 15), compared to under adjacency matrix distance (`Adj`) or latent factor distance (`LT`). This demonstrates one exemplary network where `MGC` statistic along with a family of diffusion distances catches non monotonic correlations efficiently than the other statistics and metrics.

When $X_i = X_j$, these two nodes are most likely to have an edge but when X_i and X_j differ by one, they are even less likely to have an edge, with probability of 0.2, than the most different pairs of nodes. This actually describes nonlinear dependence where `MGC` is believed to work better than the distance correlation. To scrutinize our conjecture on better performance of local optimal scaled `MGC` over global scale of `mCorr`, we control the amount of *nonlinear dependency* through changing the value of $\theta \in (0, 1)$ in the three block model 7.

$$E(A_{ij}|X_i, X_j) = 0.5I(|X_i - X_j| = 0) + 0.2I(|X_i - X_j| = 1) + \theta I(|X_i - X_j| = 2). \tag{7}$$

When $\theta > 0.2$, linear dependency of edge distribution in $\mathbf{A}$ upon nodal attribute of X is lost. If you see Figure 3, power of `mCorr` starts to drop from $\theta = 0.2$ while that of `MGC` almost stays clam, which implies `MGC` is significantly more sensitive to nonlinear dependency compared to `mCorr`.

On the other hand, the SBM connotes that all nodes within the same block have the same expected degree. Thus, this block model is limited by homogeneous distribution within block and provides a poor fit to networks with highly varying node degrees within block or community. Instead the Degree-Corrected Stochastic Block model (DCSBM) proposed by Karrer and Newman [2011] add another random variable associated with each node to vary the node degrees. In the model 8, we controlled the amount of such variability by τ; the larger the value τ is, the more variability degree or edge distribution has. In Figure 4, power based on Euclidean distance of $\mathbf{A}$ or that of estimated network factors (locations) becomes less

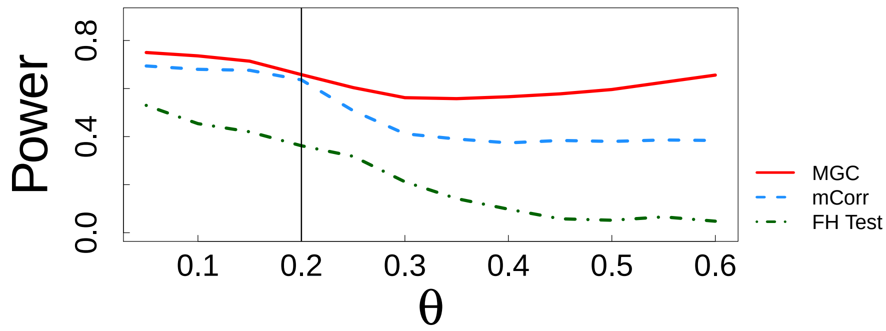

Figure 3: X-axis of θ controls the existence/amount of nonlinear dependency and in this particular case nonlinearity exists when $\theta > 0.2$ and gets larger as it increases. You can see the discrepancy in power between global and local scale tests also gets larger accordingly, mostly due to decreasing power of `mCorr` or `FH` test but relatively stable power of `MGC` under nonlinear dependency.

sensitive as τ increases. Compared to these two, diffusion maps are more robust to such variability.

$$E(A_{ij}|\mathbf{X}, \mathbf{V}) = 0.2V_iV_j \cdot I(|X_i - X_j| = 0) + 0.05V_iV_j \cdot I(|X_i - X_j| = 1), \quad \text{where } V_i \overset{i.i.d}{\sim} Uniform(1-\tau, 1+\tau). \tag{8}$$

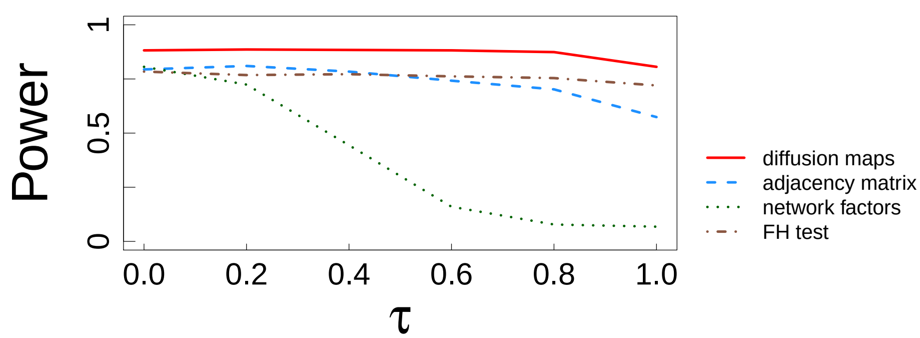

Figure 4: In degree-corrected SBM where the variability in degree distribution increases as τ increases, testing power of diffusion maps are more likely to be robust against increasing variability compared to other network metrics, e.g. adjacency matrix or latent positions. `FH` test statistics allowing different dimensions of network factors perform consistently well but still have less power than `MGC`.

3.2 Node Contribution Test

To examine the effectiveness of node contribution measure in testing dependency as presented in the statistic 5, we deliberately simulate the network and its nodal attributes as half of the nodes are independent while the other half are dependent on network (model 16). As an ad hoc test of node contribution, we rank the nodes in terms of decreasing order of $c(v)$ and count the ratio of dependent samples's ranks within the number of dependent nodes. If it works perfectly, all dependent nodes would take higher rank than every independent node so thus the rate equals to one. We call this rate as

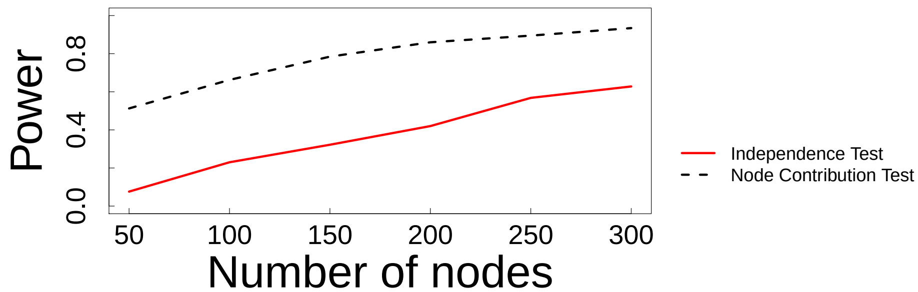

Figure 5: This plot describes that both power of `MGC` and the rate of correctly-ranked node contribution increase as the number of nodes increases when only half of the nodes for each simulation actually are set to be dependent on network, which validates the use of node contribution measure in independence test.

inclusion rate:

$$\text{inclustion rate}(c(v)) = \sum_{v \in V(\mathbf{G})} \{rank_{c(v)}(v) \leq m\}/m, \tag{9}$$

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

4 Real Data Examples

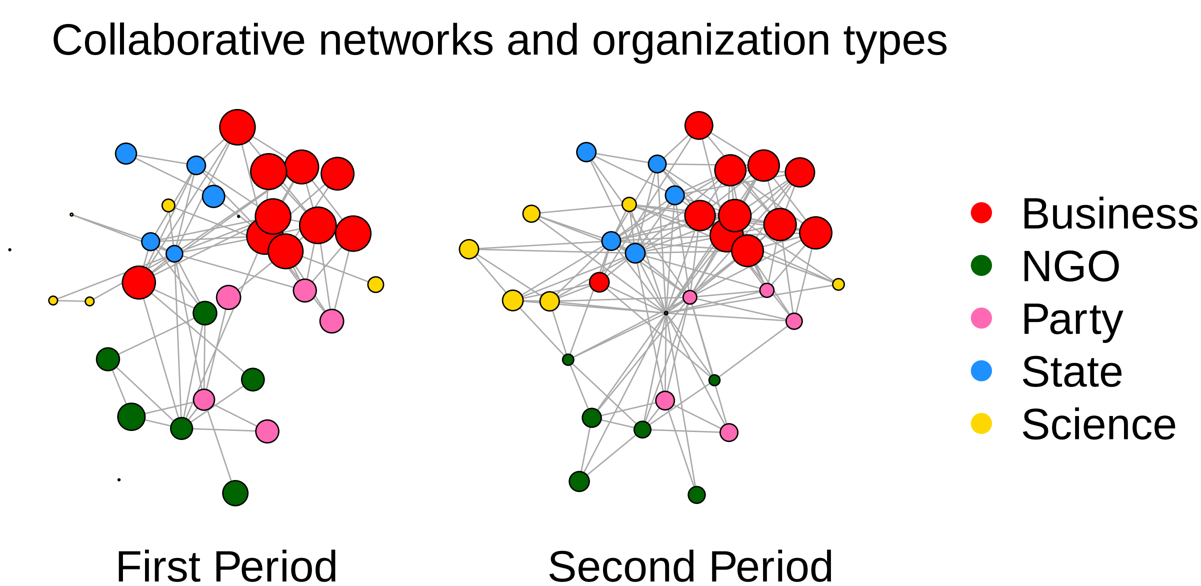

Figure 6: Both panels depict the collaborative networks during the two time periods having significant network dependency in types of organizations. Using `MGC` statistics, we are not only able to test network independence but also calculate each node's amount of contribution to detecting dependence, which is proportional to node size here. You can tell that the tendency to collaborate within the same type is strongest among the business group while scientist relatively collaborates less with any others, especially in the first period.

In the field of political science, who exerts more powerful impacts than the others over political network and which factors impact on the power differentials are one of the interests [Ingold and Leifeld, 2014]. Minhas et al. [2016] made an inference from political networks [Cranmer et al., 2016] via the additive and multiplicative effects (AME). The AME model estimates the latent factors and uses them to test independence with the nodal attributes. Among diverse attributes that Cranmer et al. [2016] provided, we focus on the types of organizations and how 34 political organizations having different types are participating policy network. We changed a given directed network into undirected network and use a dissimilarity matrix for distance matrix of the attributes, i.e., $\| \mathbf{X}_i - \mathbf{X}_j \| = 0$ if and only if node i and node j are from the same type and one otherwise. Two collaboration networks comprised of the same set of nodes across two time periods are provided [Ingold and Leifeld, 2014]. Figure 6 and Figure 7 illustrates these two networks and shows each node's reliance on its organization type when collaborating. During the two periods, the network independence test statistics of `MGC` and `dCorr` using diffusion distance matrices result in significant p-values across diffusion times from $t = 1$ to $t = 10$. The conclusion from the `FH` test is also the same.

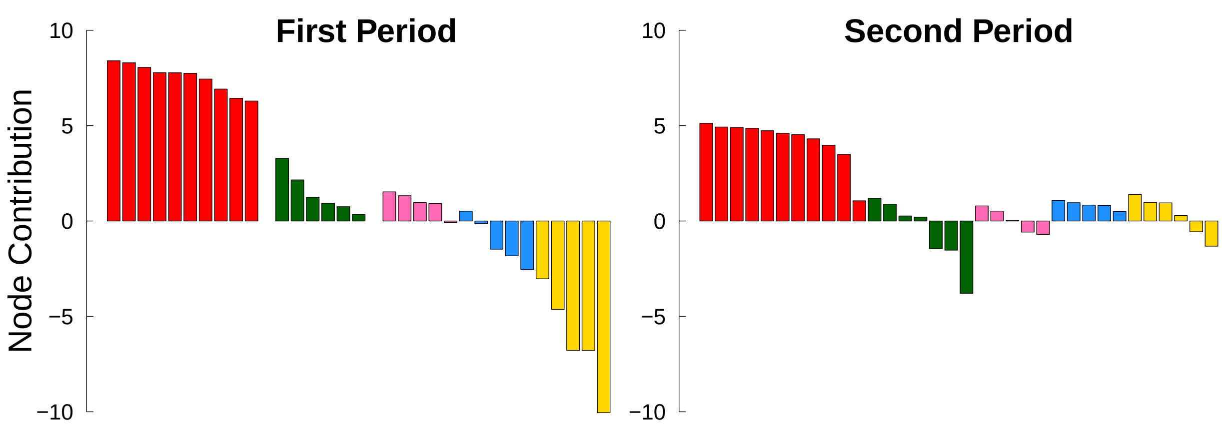

Figure 7: In the first period, we have two extreme cases among the business group and science group, which reflects our observations in Figure 6. Generally organizations cooperate more actively between different types in the second period but still their collaboration network is highly dependent on their organization types especially for business group.

5 Conclusion

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

However obtaining a full family of statistics are computationally infeasible. Also we did not suggest any theoretically supported tools to select one metrics among them so thus we have one single statistic. As an ad hoc, we selected an *optimal* diffusion time t with highest power from $t = 1$ to $t = 10$ for our simulation since we could observe a stabilized empirical power within this period. Developing the adaptive method to find this optimal t where dependence is maximized would be a natural next step. Despite these shortcomings, we expect that we could also enjoy the properties of `MGC` and a family of diffusion distances in solving diverse problems which require to utilize local relationship of the data sets. For instance, we might be able to implement independence testing between two networks of same size by using diffusion distance of each network to investigate whether a pair of networks are topologically or structurally independent. This kind of work would shed light on revealing any relationship between the data sets which are not necessarily a random vector.

References

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

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

Skyler J Cranmer, Philip Leifeld, Scott D McClurg, and Meredith Rolfe. Navigating the range of statistical tools for inferential network analysis. *American Journal of Political Science*, 2016.

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

Ruth Heller, Yair Heller, and Malka Gorfine. A consistent multivariate test of association based on ranks of distances. *Biometrika*, page ass070, 2012.

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

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

Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

Stephane Lafon and Ann B Lee. 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, 2006.

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

Shahryar Minhas, Peter D Hoff, and Michael D Ward. Inferential approaches for network analyses: Amen for latent factor models. *arXiv preprint arXiv:1611.00460*, 2016.

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

Cencheng Shen, Carey E Priebe, Mauro Maggioni, and Joshua T Vogelstein. Discovering relationships across disparate data modalities. *arXiv preprint arXiv:1609.05148*, 2016.

Gábor J Székely and Maria L Rizzo. The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis*, 117:193–213, 2013.

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

Sara Taskinen, Hannu Oja, and Ronald H Randles. Multivariate nonparametric tests of independence. *Journal of the American Statistical Association*, 100(471):916–925, 2005.

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

6 Appendix

6.1 Lemmas and Theorems

Proof of Lemma 2.1. Diffusion map at time t is represented as follows :

$$\mathbf{U}_t(i) = \begin{pmatrix} \lambda_1^t \phi_1(i) & \lambda_2^t \phi_2(i) & \cdots & \lambda_q^t \phi_q(i) \end{pmatrix} \in \mathbb{R}^q. \tag{10}$$

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

Assume joint exchangeability of $\mathbf{G}$, i.e. $(A_{ij}) \overset{d}{=} \big(A_{\sigma(i)\sigma(j)}\big)$. Since A_{ij} is binary, $A_{ij}/\sum\limits_j A_{ij} = A_{ij}/(1 + \sum\limits_{l \neq j} A_{il})$. 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 for $i \neq j; i, j = 1, 2, \ldots, n$:

$$\big(P_{ij}\big) = \left(\frac{A_{ij}}{1 - A_{ij} + \sum\limits_{j=1}^{n} A_{ij}} \right) \overset{d}{=} \left(\frac{A_{\sigma(i)\sigma(j)}}{1 - A_{\sigma(i)\sigma(j)} + \sum\limits_{\sigma(j)=1}^{n} A_{\sigma(i)\sigma(j)}} \right) = \big(P_{\sigma(i)\sigma(j)}\big) \tag{11}$$

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

Proof of Theorem 2.2 *Consistency of `dCorr` applied to exchangeable variables.* For exchangeable sequence of $(\mathbf{W}, \mathbf{Y}) = \{(\mathbf{w}_i, \mathbf{y}_i); i = 1, 2, \ldots, n\}$ which is identically distributed as $(\mathbf{w}, \mathbf{y})$ with finite second moment, we have

$$\mathcal{V}_n^2(\mathbf{W}, \mathbf{Y}) \longrightarrow \mathcal{V}^2(\mathbf{w}, \mathbf{y}) \qquad \text{as } n \to \infty \tag{12}$$

where $\mathcal{V}^2(\mathbf{w},\mathbf{y}) := \|g_{\mathbf{w},\mathbf{y}}(t,s) - g_{\mathbf{w}}(t)g_{\mathbf{y}}(s)\|^2$, and $g_.$ is a characteristic function, e.g., $g_{\mathbf{w},\mathbf{y}}(t,s) = E\{\exp\{i\langle t,\mathbf{w}\rangle + i\langle s,\mathbf{y}\rangle\}\}$. This follows exactly the same as *Theorem 1* in Székely et al. [2007]. Note that this Lemma always holds without any assumption on $\{(\mathbf{w}_i,\mathbf{y}_i), i = 1,2,...,n\}$.

Followed by *de Finetti's Theorem*, if and only if $\{\mathbf{w}_i\}$ are (infinitely) exchangeble, there exists an underlying distribution $f_{\mathbf{w}}$ of $\mathbf{w}$ such that $\mathbf{w}_i \overset{i.i.d}{\sim} f_{\mathbf{w}}$. By the same logic there exists a random, we have an underlying distribution $f_{\mathbf{y}}$ where $\mathbf{y}_i \overset{i.i.d}{\sim} f_{\mathbf{y}}$. Let $(\mathbf{w}_i,\mathbf{y}_i) \overset{i.i.d}{\sim} f_{\mathbf{w},\mathbf{y}}$. Then under the assumption of finite second moment of the underlying distributions and measurable, conditioned random functions, we have a strong large number for V-statistics followed by Székely et al. [2007], i.e.,

$$\int_{D(\delta)} \|g^n_{\mathbf{w},\mathbf{y}}(t,s) - g^n_{\mathbf{w}}(t)g^n_{\mathbf{y}}(s)\|^2 dw \quad \overset{n\to\infty}{\longrightarrow} \int_{D(\delta)} \|g_{\mathbf{w},\mathbf{y}}(t,s) - g_{\mathbf{w}}(t)g_{\mathbf{y}}(s)\|^2 dw, \tag{13}$$

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

$$\mathcal{V}^2_n(\mathbf{W},\mathbf{Y}) \quad \to 0 \quad \text{ as } n \to \infty \tag{14}$$

if and only if $g_{\mathbf{w},\mathbf{y}}(t,s) = g_{\mathbf{w}}(t)g_{\mathbf{y}}(s)$, i.e., $\mathbf{w}$ is independent of $\mathbf{y}$. Therefore, the `dCorr` or `mCorr` converges to 0 if and only if underlying distributions are independent; and its testing power converges to 1 under any joint distribution of finite moments. Since the multiscale generalized correlation based on any consistent global correlation is also consistent [Shen et al., 2016], `MGC` statistic constructed by `dCorr` or `mCorr` is also consistent in testing dependence. □

***Proof of Theorem* 2.3** *Consistency of* `MGC` *applied to exchangeable variables.* Under the exchangeability and finite second moment assumptions of underlying distribution, $\mathcal{V}^2_n(\mathbf{W},\mathbf{Y}) \xrightarrow{n\to\infty} 0$ if and only if underlying distribution of $\{\mathbf{w}_i\}$, $f_{\mathbf{w}}$ is independent from underlying distribution of $\{\mathbf{y}_i\}$, $f_{\mathbf{y}}$. Now suppose that we have undirected, connected network $\mathbf{G}$ with a family of diffusion maps $\{\mathbf{u}_t\}$ and with nodal attributes $\{\mathbf{x}\}$. We have shown in the Lemma 2.1 that $\{\mathbf{u}_t\}$ are exchangeable for each $t \in \mathbb{N}$. Thus there exists an underlying distribution of $\mathbf{u}_t$ such that $\mathbf{u}_t(i) \overset{i.i.d}{\sim} f_{\mathbf{u}^{(t)}}$ for each of $t = 1,2,\ldots$; and we have $\mathbf{x}_i \overset{i.i.d}{\sim} f_{\mathbf{X}}$. Under the assumption of finite second moment of $\mathbf{u}^{(t)}$ and $\mathbf{x}$, `MGC` statistics constructed by $\{(\mathbf{u}_t(i),\mathbf{x}_i) : i = 1,2,\ldots,n\}$ yield a consistent testing which determines the independence between underlying distributions of $\mathbf{u}^{(t)}$ and $\mathbf{x}$. From the same setting of network $\mathbf{G}$, we have estimated $i.i.d$ node-specific network factors $\{\mathbf{F}_i\}$ so that n-pair of $i.i.d$ $\{(\mathbf{F}_i,\mathbf{x}_i)\}$ can be applied to `MGC` or other distance-based tests without assuming conditioning underlying distribution. In case of using adjacency matrix directly into test, we must assume that the adjacency matrix comes from connected directed network, i.e. $A_{ij} \overset{i.i.d}{\sim} f_A$ for all $i, j = 1, 2, \ldots, n$; otherwise, each column is dependent on one another. $\square$

6.2 Simulation Data

- **Three Block SBM** ($n = 100$)

$$\begin{aligned}
&X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Multi(1/3, 1/3, 1/3), i = 1, \ldots, n \\
&Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Multi(0.5, 0.25, 0.25)I(x = 1) + Multi(0.25, 0.5, 0.25)I(x = 2) \\
&\qquad + Multi(0.25, 0.25, 0.5)I(x = 3), \quad i = 1, \ldots, n \\
&A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \overset{d}{=} Bern(0.5)I(|z_i - z_j| = 0) + Bern(0.2)I(|z_i - z_j| = 1) \\
&\qquad + Bern(0.3)I(|z_i - z_j| = 2), \quad i < j, i, j = 1, \ldots n
\end{aligned} \tag{15}$$

- **Node Contribution**

$$\begin{aligned}
&X_i \overset{i.i.d}{\sim} f_X(x) \overset{d}{=} Bern(0.5) \quad i = 1, \ldots, n/2, \ldots, n \\
&Z_i|X_i \overset{i.i.d}{\sim} f_{Z|X}(z|x) \overset{d}{=} Bern(0.6)I(x = 0) + Bern(0.4)I(x = 1), \quad i = 1, \ldots, n/2, \ldots, n \\
&A_{ij}|Z_i, Z_j \overset{i.i.d}{\sim} f_{A|Z}(a_{ij}|z_i, z_j) \\
&\quad \overset{d}{=} \begin{cases} Bern(0.4)I(|z_i - z_j| = 0) + Bern(0.1)I(|z_i - z_j| > 0) & i = 1, \ldots, n/2 \\ Bern(0.25) & i = 1 + n/2, \ldots, n \end{cases}
\end{aligned} \tag{16}$$