

MEASURING THE GRAPH CONCORDANCE OF LOCALLY DEPENDENT OBSERVATIONS

Kyungchul Song*

Abstract—This paper introduces a simple measure of a concordance pattern among observed outcomes along a network, that is, the pattern in which adjacent outcomes tend to be more strongly correlated than nonadjacent outcomes. The graph concordance measure can be generally used to quantify the empirical relevance of a network in explaining cross-sectional dependence of the outcomes, and as shown in the paper, it can also be used to quantify the extent of homophily under certain conditions. When one observes a single large network, it is nontrivial to make inferences about the concordance pattern. Assuming a dependency graph, this paper develops a permutation-based confidence interval for the graph concordance measure. The confidence interval is valid in finite samples when the outcomes are exchangeable, and under the dependency graph, an assumption together with other regularity conditions, is shown to exhibit asymptotic validity. Monte Carlo simulation results show that the validity of the permutation method is more robust than the asymptotic method to various graph configurations.

I. Introduction

THE role of a network has long received attention in the social sciences. When one studies network data and outcomes on the network (e.g., a friendship network and test scores of the students), a primitive empirical task would be to determine whether a network has any relevance in explaining observed outcomes (i.e., whether it has any explanatory power) and, if so, what aspect of the network is relevant and how to quantify it, taking into account sampling errors. For example, one may ask whether students perform more similarly when they are friends than when they are not. Or one may ask whether racial indicators are more strongly correlated among friends than among nonfriends, that is, whether there is race-based homophily among students. The observation units do not need to be people or firms; they can be geographic districts linked along a transportation network or software products connected by common developers.

This paper introduces a simple parameter that measures the strength of correlation among linked pairs of outcomes relative to that among unlinked pairs: the graph concordance (GC). Roughly speaking, the GC of a random vector laid on a graph is defined to be the difference between the average correlation of linked pairs of outcomes and that of unlinked pairs. A nonzero GC suggests that the graph

induces disparity in correlation between linked pairs of outcomes and unlinked pairs of outcomes.

The GC can be used as an empirical measure of homophily among individuals in a large friendship network. When individuals are more likely to befriend people of the same race, the GC of people's racial backgrounds along the friendship network will be positive. Indeed, this paper shows that under certain conditions, the GC coincides with the inbreeding homophily of Currarini, Jackson, and Pin (2009, based on the measure of Coleman, 1958). The GC can be used for other purposes as well. When pairs of regions with stronger trade links tend to produce more economic outputs, this suggests a conspicuous role of a trade network in shaping regional variations in outputs, which will be reflected in the positive GC of the regional outputs.

Despite the simplicity of GC, developing a formal inference procedure on the GC is a nontrivial task, especially when one observes only a single, large, complex network. This is because GC is a network-specific parameter that involves a large, complex network. This makes a contrast with the environment for testing for stochastic dominance or testing for independence where one makes many repeated observations of a random vector of a small dimension.

To deal with this situation, this paper provides a benchmark inference method under the cross-sectional dependence assumption of a dependency graph. This assumption requires that any two sets of outcomes that have no link between the two sets are independent. While the assumption can be restrictive in certain applications, it is flexible in other dimensions. First, the correlations between linked pairs can be heterogeneous across pairs. Second, the local dependence is different from cluster dependence because the neighborhoods are not disconnected from each other. Therefore, the cross-sectional dependence structure can be very complex when the network is large and complex. Third, one can easily extend this paper's approach to m -dependence, where any two sets of outcomes that are m links away are independent.¹ While it is preferable to adopt parsimonious modeling, some network configurations may allow this extension without obscuring the inference result, such as when there are many components in the large network (e.g., social network studies on households in many villages or students in many schools.)

The main contribution of this paper is to develop a method of permutation-based inference for the GC under the dependency graph assumption. The confidence intervals generated from the method exhibit two-tier validity: when the

¹ The m -dependence in time series can be viewed as a special case of an m -dependency graph where the graph is given as a graph with a single path along a time ordering.

Received for publication September 14, 2016. Revision accepted for publication July 24, 2017. Editor: Bryan S. Graham.

* Vancouver School of Economics, University of British Columbia.

I thank the editor and three referees for valuable comments and suggestions. I am grateful to Colin Cameron, Nathan Canen, Xiaoqi He, Hiro Kasahara, Brian Krauth, Vadim Marmer, Shu Shen, Aaron Smith, and Jungmo Yoon for valuable conversations and comments. I appreciate comments from the participants of the Seattle-Vancouver Econometrics Conference, Duke University, University of California at Davis, University of California at San Diego, and University of Southern California. I thank Ning Ding for her excellent research assistance. All errors are mine. I acknowledge that this research was supported by Social Sciences and Humanities Research Council of Canada.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00714.

outcomes are exchangeable, the inference procedure is valid in finite samples, and when the outcomes fail to be exchangeable, the inference procedure is shown to be asymptotically valid under a set of regularity conditions.

The finite sample validity under exchangeability is not surprising. The main theoretical contribution of this paper is to establish conditions for the networks so that the permutation inference is asymptotically valid even when the outcomes fail to be exchangeable, which is the case with a nonzero GC. This result is crucial for our purpose because the confidence interval should be valid regardless of whether the GC is 0.

Establishing conditions for the network is crucial here. Given the dependency graph assumption, there is no hope of obtaining a meaningful asymptotic distribution for a test statistic from observing a single large network if the network is too dense. For example, if most nodes are linked to most other nodes, the dependency graph does not produce enough independence. This paper shows that if the maximum number of the neighbors in the network (i.e., the maximum degree of the network) increases no faster than at a certain polynomial rate, as the number of the nodes in the network increases, the asymptotic validity of permutation inference follows.

The main idea of this paper comes from a simple observation, as follows. Since each permutation relabels the observation indices, the cross-sectional dependence structure and heterogeneity of marginal distributions carry over under the relabeling. Hence asymptotic derivation can be performed conditional on a fixed sequence of permutations using the central limit theorem for random variables with dependency graphs (e.g., theorem 2.4 of Penrose, 2003). The main difficulty in applying it to an estimation problem comes from lack of a general method to center the permutation test statistic. However, the permutation test statistic in this paper is approximately centered regardless of whether the GC is 0 because most permutations reduce adjacent vertices to nonadjacent ones. This feature enables us to apply the permutation method to the estimation problem in this paper.

One might wonder whether we can simply use asymptotic critical values instead of permutation-based ones, ignoring the finite sample validity altogether. First, it should be noted that despite the complexity of theoretical arguments showing its asymptotic validity, the actual implementation of the permutation inference is pretty simple, as we simply randomly relabel the cross-sectional units and compute the GC estimator and its variance estimator the same way as the original test statistic. More important, this paper's Monte Carlo simulation study demonstrates that asymptotic critical values are not as stable as the permutation-based critical values, especially as the network becomes denser. Therefore, this paper's adoption of the permutation approach is not merely of theoretical interest.

To demonstrate the usefulness of this paper's proposal, we apply it to the social networks data on Indian villages

used in Banerjee et al. (2013). The empirical application is divided into two parts. First, we estimate the GC of the indicator of whether the household belongs to a minority caste or tribe along the social network to see whether there is homophily along the castes. The main point of this exercise is to quantify the homophily among the observed households and provide its confidence interval. It turns out that the estimated GC is around 0.57 to 0.59 with 95% confidence intervals roughly contained in $[0.50, 0.64]$. Thus, the GC is very high, indicating strong homophily among a minority caste or tribe.

Second, we search for evidence that the social network plays a role in explaining microfinancing decisions. The estimated GC of the household decisions on microfinancing from the same sample as before is around 0.11 to 0.14, and significantly different from 0 at 5%. Therefore, while the GC of the microfinancing decisions is much weaker than the GC of the minority castes, it is significantly positive, indicating the relevance of the social network in explaining microfinancing decisions.

A. Literature Review

Defining and estimating a measure of dependence between variables drew interest early on with the beginning of modern statistics. Among the measures are Pearson's correlation coefficient, Kendall's τ , and Spearman's ρ . (See, e.g., chapter 8 of Cameron & Trivedi, 2013, for various dependence measures.) These measures are typically between two random variables (or their i.i.d. draws), in contrast to the GC in this paper, which measures cross-sectional correlation along a single large network.

Closely related to this paper's GC is Moran's spatial autocorrelation test statistic (the I test statistic) that is popularly used in the literature of spatial modeling (Moran, 1950; Cliff & Ord, 1972). Kelejian and Prucha (2001) established asymptotic theory for Moran's I test statistic in a general setup. Pesaran (2004) developed a general test for cross-sectional dependence in linear panel models with a short time-series dimension. (See Hsiao, Pesaran, & Pick, 2012, for an extension to limited dependent models.) Robinson (2008) proposed a correlation test that can be applied for testing cross-sectional dependence in a spatial model. The main difference that this paper's contribution makes is twofold. First, the GC in this paper is proposed as a population quantity to be estimated, not just as a test statistic. Because it is an estimation problem, one cannot impose the null hypothesis of, say, no spatial autocorrelation, in developing asymptotic validity of inference. Second, the main focus of the paper is on permutation-based inference instead of asymptotic inference.

Permutation tests were among the earliest methods of nonparametric testing in statistics and have been applied in wide areas of statistics, biometrics, and econometrics. They have been mostly used for testing equality of the means or the distributions from different subpopulations and testing

independence between two random variables. Daniels (1944) focused on the permutation-based test on sample correlation measures between two sets of observations conditional on the samples. Friedman and Rafsky (1979) considered a two-sample problem based on minimum spanning trees. Romano (1989) investigated bootstrap and permutation tests and considered, among several others, an example of testing independence. Delgado (1996) developed a permutation test of serial dependence. A textbook treatment of the method and references are found in Lehmann and Romano (2005).

Asymptotic robustness of permutation-based tests when exchangeability fails has already drawn attention in the literature. Neuhaus (1993) considered a two-sample problem with random censoring and found that one can construct a permutation test that controls the directional error through an asymptotic pivotal test. Janssen (1997) extended the approach to other two-sample testing problems. This approach is substantially generalized by Chung and Romano (2013), who proposed a permutation test in two-sample problems where the parameter of interest takes a general form. More recently, Canay, Romano, and Shaikh (2017) investigated asymptotic validity of randomized tests under approximate symmetry conditions, where the symmetry conditions are imposed not on the finite sample distribution but on the asymptotic distribution of the statistic.

B. Organization of the Paper

The next section introduces GC and discusses examples. Section III presents the paper's main proposal: the permutation inference on the GC. The section establishes the asymptotic validity of the proposed inference method. Section IV presents and discusses Monte Carlo simulation studies, and section V illustrates the usefulness of the proposal through an empirical application on the Indian village data. Section VI concludes. The appendix provides heuristics behind the theoretical results. The full proof of the main theorem and auxiliary results are found in the supplemental note to this paper.

II. Graph Concordance of Cross-Sectional Observations

A. Graph Concordance

Suppose that we observe random vector $Y = (Y_i)_{i \in N_n}$, $Y_i \in \mathbf{R}$, with $N_n \equiv \{1, \dots, n\}$, and a graph (or a network) $G_n = (N_n, E_n)$, where N_n indicates the set of vertices (or nodes) and E_n the set of pairs (i, j) , $i, j \in N_n$, each pair (i, j) representing an edge (or a link) between i and j . From here on, we write ij interchangeably with (i, j) . We assume that the graph is undirected, so that $ij \in E_n$ if and only if $ji \in E_n$. When $ij \in N_n$, we say that i and j are adjacent. We do not allow a loop— $ii \notin E_n$ for any $i \in N_n$. Let \tilde{N}_n be the set of all the edges possible from N_n : $\tilde{N}_n = \{(i, j) \in N_n^2 : i \neq j\}$. For each $i \in N_n$, define

$$N_n(i) = \{j \in N_n : ij \in E_n\}$$

to be the (open) neighborhood of vertex i and any vertex in the neighborhood a neighbor of i . We let $d_n(i) = |N_n(i)|$ and call it the degree of vertex i —the number of the neighbors of i . We also introduce the closed neighborhood of vertex i which is defined to be

$$\bar{N}_n(i) = N_n(i) \cup \{i\}.$$

Throughout the paper, we assume that the graph G_n is not a complete graph, that is, E_n is a proper subset of \tilde{N}_n . This means that there are pairs of vertices that are not adjacent. We also assume that $\bar{N}_n(i)$ is a proper subset of N_n for all $i \in N_n$, which means that for each vertex i , there exists a vertex that is not its neighbor.

We introduce a parameter that measures the tendency of linked pairs of outcomes, say, (Y_i, Y_j) 's with $ij \in E_n$, exhibiting a stronger correlation than unlinked pairs of outcomes, say, (Y_i, Y_j) 's with $ij \in \tilde{N}_n \setminus E_n$. First, define

$$\bar{Y}_i = \frac{1}{d_n(i)} \sum_{j \in N_n(i)} Y_j,$$

$$\bar{Y}_i^c = \frac{1}{|N_n \setminus \bar{N}_n(i)|} \sum_{j \in N_n \setminus \bar{N}_n(i)} Y_j,$$

if $d_n(i) \geq 1$, and $\bar{Y}_i = 0$, otherwise. Thus \bar{Y}_i is the average of Y_j 's in the neighborhood of i . We let

$$\gamma = \frac{1}{nv_n^2} \sum_{i \in N_n} \text{Cov}(Y_i, \bar{Y}_i), \quad (1)$$

$$\gamma^c = \frac{1}{mv_n^2} \sum_{i \in N_n} \text{Cov}(Y_i, \bar{Y}_i^c),$$

where $v_n^2 = \frac{1}{n} \sum_{i \in N_n} \text{Var}(Y_i)$. The quantity γ measures the average correlation of the outcomes Y_i and its neighbors Y_j among linked outcomes and γ^c measures that among unlinked outcomes. We define the graph concordance (GC) of Y along G_n as

$$C(G_n) = \gamma - \gamma^c.$$

When GC is strongly positive, this suggests a conspicuous role played by the graph in shaping the joint dependence pattern of Y . Suppose that

$$\sum_{i \in N_n} \text{Var}(\bar{Y}_i) \leq \sum_{i \in N_n} \text{Var}(Y_i), \text{ and}$$

$$\sum_{i \in N_n} \text{Var}(\bar{Y}_i^c) \leq \sum_{i \in N_n} \text{Var}(Y_i).$$

Then GC takes values in $[-2, 2]$ in general but takes values in $[-1, 1]$ when the correlation between Y_i and Y_j , $i, j \in N_n$ is nonnegative. For example, in the extreme case where G_n consists of several complete subgraphs disconnected from each other and Y_i and Y_j are perfectly correlated whenever $ij \in E_n$ and uncorrelated whenever $ij \in \tilde{N}_n \setminus E_n$, the GC

becomes 1. We say that Y is graph concordant if $C(G_n) > 0$ and graph discordant if $C(G_n) < 0$.

Throughout this paper, we regard the graph G_n as fixed and pursue conditional inference given G_n . This does not mean that the graph G_n is required to be exogenous for the validity of inference. The conditional inference merely means that the coverage probability of the proposed confidence set is expressed in terms of the conditional probability given G_n . When one views the graph G_n as a random graph generated by a certain data-generating process, the conditional validity of the inference means that it is conditionally valid for each realization of the graph. Therefore, the inference is unconditionally valid as well.

B. Graph Concordance and Inbreeding Homophily

Currarini et al. (2009) considered a measure of homophily (due to Coleman, 1958), where the extent of homophily is measured by what they call the inbreeding homophily. Here we show that under certain conditions, graph concordance and inbreeding homophily coincide. To see this, suppose that each individual $i \in N_n$ assumes a type $t \in T$, where T is a finite set so that i is associated with a random variable $D_i \in T$. The inbreeding homophily for type t is defined as²

$$IH(G_n) = \frac{H - w}{1 - w},$$

where $H = s_n / (s_n + d_n)$, and s_n denotes the average of the probabilities $P\{D_i = t, D_j = t\}$ over $i, j \in N_n$ such that $ij \in E_n$ and d_n denotes the average of the probabilities $P\{D_i = t, D_j \neq t\}$ over $i, j \in N_n$ such that $ij \in E_n$, and w the average of $P\{D_i = t\}$ over $i \in N_n$. (see Currarini et al., 2009, pp. 1007–1008 for motivation.) Letting $Y_i = 1\{D_i = t\}$, we can rewrite H and w as

$$H = \frac{\sum_{i \in N_n} \mathbf{E}[Y_i \bar{Y}_i] d_n(i)}{\sum_{i \in N_n} \mathbf{E}[Y_i] d_n(i)}, \text{ and } w = \frac{1}{n} \sum_{i \in N_n} \mathbf{E}[Y_i].$$

Now assume the following three conditions:

- a. $\mathbf{E}[Y_i]$'s are the same across $i \in N_n$.
- b. $d_n(i)$'s are the same across $i \in N_n$.
- c. Y_i and Y_j are independent whenever i and j are not linked.

The equal degree condition, b can be removed if we modify the definition of inbreeding homophily as follows:

$$IH'(G_n) = \frac{H' - w}{1 - w},$$

² The original definition of the inbreeding homophily measure in Currarini et al. (2009) involves only the samples (i.e., without the expectation). For inference, we modify their measure as a population quantity here.

with the modified definition of H :

$$H' = \frac{\sum_{i \in N_n} \mathbf{E}[Y_i \bar{Y}_i]}{\sum_{i \in N_n} \mathbf{E}[Y_i]},$$

which can be viewed as the inbreeding homophily with H weighted by the inverse of degree $d_n(i)$. This modified version attenuates the influence of nodes with many links. Condition c becomes plausible when G_n is a subgraph of a much larger sparse graph and Y_i 's are locally dependent, satisfying the law of the large numbers. (See the remarks after theorem 1 in section IIIA for details.)

Then we can write

$$IH(G_n) = \frac{\sum_{i \in N_n} \text{Cov}(Y_i, \bar{Y}_i)}{\sum_{i \in N_n} \text{Var}(Y_i)} = C(G_n). \tag{2}$$

Therefore, under conditions a to c, $IH(G_n) = C(G_n)$, and under conditions a and c, $IH'(G_n) = C(G_n)$. (When condition c is violated, the same statement holds if we replace $C(G_n)$ by γ defined in (equation 1).) Thus the graph concordance $C(G_n)$ can be viewed as a measure of homophily as inbreeding homophily in this case.

C. Residual Graph Concordance

One may consider graph concordance of Y_i after taking into account variations of observable node characteristics, say, X_i . We consider the following linear model,

$$Y_i = X_i' \beta_0 + u_i,$$

where we assume that each X_i is a discrete random vector of a small dimension having a finite support \mathcal{X} and $\mathbf{E}[u_i | (X_j)_{j \in N_n}] = 0$. (We do not necessarily view the linear model as a causal model.) For each $x \in \mathcal{X}$, we define

$$N_{n,x} = \{i \in N_n : X_i = x\},$$

and $N_{n,x}(i) = \{j \in N_{n,x} : ij \in E_n\}$, and let

$$\bar{u}_{x,i} = \frac{1}{|N_{n,x}(i)|} \sum_{j \in N_{n,x}(i)} u_j,$$

$$\bar{u}_{x,i}^c = \frac{1}{|N_{n,x} \setminus \bar{N}_{n,x}(i)|} \sum_{j \in N_{n,x} \setminus \bar{N}_{n,x}(i)} u_j,$$

where $\bar{N}_{n,x}(i) = N_{n,x}(i) \cup \{i\}$. Then we may consider the conditional graph concordance of u_i 's as follows:

$$C_x(G_n) = \gamma_x - \gamma_x^c,$$

where

$$\gamma_x = \frac{1}{m_{n,x}^2} \sum_{i \in N_n : X_i = x} \text{Cov}(u_i, \bar{u}_{x,i} | X), \tag{3}$$

$$\gamma_x^c = \frac{1}{m_{n,x}^2} \sum_{i \in N_n : X_i = x} \text{Cov}(u_i, \bar{u}_{x,i}^c | X),$$

and $v_{n,x}^2 = \frac{1}{n} \sum_{i \in N_n: X_i=x} \text{Var}(u_i|X)$, $X = [X_1, \dots, X_n]'$, and $\text{Cov}(\cdot, \cdot|X)$ and $\text{Var}(\cdot|X)$ denote conditional covariance and conditional variance given X . Let us call $C_x(G_n)$ the residual graph concordance of Y_i with respect to G_n at $X_i = x$.

D. Examples

Homophily on social characteristics. Many studies in sociology and economics are interested in the social phenomenon of homophily, which refers to the tendency of individuals to associate with others who are similar to themselves in terms of social characteristics such as race, income, or religion. (See Currarini et al., 2009 for an economic model of friendship formation.) The strength of homophily is a matter of empirical question and can be measured by GC, where the observed vector Y is taken to be race, income, or religion and the graph is the social network among people. Of course, one needs to take care in interpreting the results, because the cross-sectional dependence captured by the GC can arise due to other sources.

Linear network interference on treatment outcomes. One may use GC to measure the network interference on treatment outcomes.³ Let D_i denote the treatment state for individual $i \in N_n$, where individuals are on a friendship network $G'_n = (N_n, E'_n)$. The outcome Y_i for individual i is specified as

$$Y_i = \beta_0 + \beta_1 \bar{D}_i + \beta_2 D_i + \eta_i, \quad (4)$$

where (D_i, η_i) 's are i.i.d. across i 's, $\bar{D}_i = \frac{1}{d'_n(i)} \sum_{j \in N'_n(i)} D_j$, and $N'_n(i)$ denotes the neighborhood of i and $d'_n(i)$ the degree of i in G'_n , and η_i and D_i are allowed to be correlated. Of frequent interest is whether network interference is present, that is, whether the outcome Y_i depends on \bar{D}_i . (See Aronow & Samii, 2015, and Athey, Eckles, & Imbens, 2015, and Leung, 2016, for network interference issues in program evaluations and for references in the related literature.) Suppose further that D_i 's are not precisely observed. Then if $\beta_1 > 0$, the network interference will be revealed through the cross-sectional positive correlation among Y_i 's where Y_i and Y_j are positively correlated whenever $N'_n(i) \cap N'_n(j) \neq \emptyset$. Once one defines a graph $G_n = (N_n, E_n)$, where $ij \in E_n$ if and only if $N'_n(i) \cap N'_n(j) \neq \emptyset$, this cross-sectional positive correlation is captured by the positive GC of Y_i along G_n .

Knowledge spillover and collaboration networks. A given set of projects or research papers can form a graph, with an edge formed by a common developer or an author. Alternatively, a given set of developers or researchers can form a graph, where an edge is formed between two vertices if they work on the same project or the same research. (See Jackson & Wolinsky, 1996, for a network formation model of coauthor relations and Fershtman & Gandal, 2011, for

an empirical analysis of knowledge spillover along a graph of software developers.) One might be interested in testing whether projects that are linked through more common developers tend to exhibit higher outcomes than those that are not. Let N_n be the set of the projects and Y_i the outcome of project i , and let a graph $G_n = (N_n, E_n)$ be such that $ij \in E_n$ if and only if projects i and j have more common developers than a benchmark number. If knowledge spillover induces cross-sectional dependence of Y_i 's along the graph G_n , this cross-sectional dependence can be explored through the GC of Y_i 's along the graph G_n .

III. Permutation Inference on Graph Concordance

A. Point Estimators and Confidence Intervals

We introduce an estimated version of $C(G_n)$. First, define

$$\hat{e}_i = \frac{Y_i - \bar{Y}}{\hat{v}},$$

where $\bar{Y} = \frac{1}{n} \sum_{j \in N_n} Y_j$ and $\hat{v}^2 = \frac{1}{n} \sum_{i \in N_n} (Y_i - \bar{Y})^2$. Let $\hat{a}_i = \frac{1}{d_n(i)} \sum_{j \in N_n(i)} \hat{e}_j$ if $d_n(i) \geq 1$ and $\hat{a}_i = 0$ otherwise, and let

$$\hat{a}_i^c = \frac{1}{|N_n \setminus \bar{N}_n(i)|} \sum_{j \in N_n \setminus \bar{N}_n(i)} \hat{e}_j.$$

Thus \hat{a}_i is the difference between the average outcome over i 's neighbors and the overall average outcome after normalized by \hat{v} . Then the point estimator of the GC is defined as

$$\hat{C}(G_n) = \hat{y} - \hat{y}^c,$$

where

$$\hat{y} = \frac{1}{n} \sum_{i \in N_n} \hat{e}_i \hat{a}_i \text{ and } \hat{y}^c = \frac{1}{n} \sum_{i \in N_n} \hat{e}_i \hat{a}_i^c.$$

As for the confidence intervals, we use the permutation distribution of the test statistic given as

$$T = \frac{\sqrt{n}\{\hat{C}(G_n) - C(G_n)\}}{\hat{\sigma}_+}, \quad (5)$$

where $\hat{\sigma}_+$ denotes the scale normalizer, which we explain now. A proper scale normalizer should accommodate cross-sectional dependence along the edges. Let us introduce the population versions of \hat{e}_i 's, \hat{a}_i 's, and \hat{a}_i^c 's:

$$e_i = \frac{Y_i - \mathbf{E}Y_i}{v_n}$$

and

$$a_i = \frac{1}{d_n(i)} \sum_{j \in N_n(i)} e_j \text{ and } a_i^c = \frac{1}{|N_n \setminus \bar{N}_n(i)|} \sum_{j \in N_n \setminus \bar{N}_n(i)} e_j.$$

³ This example was inspired by suggestions by one of the referees.

We can show that under the stated conditions in theorem 1, below,

$$\sqrt{n} \{ \hat{C}(G_n) - C(G_n) \} = \frac{1}{\sqrt{n}} \sum_{i \in N_n} (q_i - \mathbf{E}q_i) + o_P(1),$$

where $q_i = e_i \{ a_i - e_i \gamma \}$. Note that q_i is essentially a modified version of $e_i a_i$, where the modification (i.e., subtraction by $e_i^2 \gamma$) is needed for accommodating the first-stage estimation error from \hat{v} . As for q_i , we later assume that for each $d = 1, \dots, d_{mx,n}$, $\mathbf{E}q_i$'s are the same for all $i \in N_n$ having the same degree, d . Define

$$S_n(i) = \{ j \in N_n : d_n(j) = d_n(i) \}, \text{ and } s_n(i) = |S_n(i)|,$$

that is, $S_n(i)$ is the set of the vertices that have the same degree as vertex i . We construct

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i_1, i_2 \in N_n: \bar{N}_n(i_1) \sim \bar{N}_n(i_2)} (\hat{q}_{i_1} - \bar{q}_{i_1})(\hat{q}_{i_2} - \bar{q}_{i_2}),$$

where $\bar{N}_n(i_1) \sim \bar{N}_n(i_2)$ means that a vertex from $\bar{N}_n(i_1)$ is adjacent to a vertex from $\bar{N}_n(i_2)$ and

$$\hat{q}_i = \hat{e}_i \{ \hat{a}_i - \hat{e}_i \hat{\gamma} \} \text{ and } \bar{q}_i = \frac{1}{s_n(i)} \sum_{j \in S_n(i)} \hat{e}_j \{ \hat{a}_j - \hat{e}_j \hat{\gamma} \}. \quad (6)$$

The location normalization by \bar{q}_i is later justified by the assumption that $\mathbf{E}q_i$ is the same for all i 's having the same degree. (See assumption 2ii and discussions below.) Define

$$\hat{\sigma}_+^2 = \begin{cases} \hat{\sigma}^2, & \text{if } \hat{\sigma}^2 > 0 \\ \hat{\sigma}_1^2, & \text{if } \hat{\sigma}^2 \leq 0 \end{cases}, \quad (7)$$

where $\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i \in N_n} (\hat{q}_i - \bar{q}_i)^2$. The introduction of $\hat{\sigma}_+^2$ ensures positivity of the scale normalizer $\hat{\sigma}_+^2$ in finite samples.

For critical values, we construct a permutation test statistic as follows. Define Π_n to be the set of permutations on N_n . The permutation test statistic T_π is obtained by replacing \hat{e}_i 's in T by $\hat{e}_{\pi(i)}$'s. More specifically, for each $\pi \in \Pi_n$, we let

$$\hat{a}_{i,\pi} = \frac{1}{d_n(i)} \sum_{j \in N_n(i)} \hat{e}_{\pi(j)} \text{ and } \hat{a}_{i,\pi}^c = \frac{1}{|N_n \setminus \bar{N}_n(i)|} \sum_{j \in N_n \setminus \bar{N}_n(i)} \hat{e}_{\pi(j)}$$

and construct

$$\hat{C}_\pi(G_n) = \frac{1}{n} \sum_{i \in N_n} \hat{e}_{\pi(i)} \hat{a}_{i,\pi} - \frac{1}{n} \sum_{i \in N_n} \hat{e}_{\pi(i)} \hat{a}_{i,\pi}^c.$$

As for the permutation version of the scale normalizer, we define

$$\hat{q}_{i,\pi} = \hat{e}_{\pi(i)} \{ \hat{a}_{i,\pi} - \hat{e}_{\pi(i)} \hat{\gamma}_\pi \} \text{ and } \bar{q}_{i,\pi} = \frac{1}{s_n(i)} \sum_{j \in S_n(i)} \hat{e}_{\pi(j)} \{ \hat{a}_{j,\pi} - \hat{e}_{\pi(j)} \hat{\gamma}_\pi \},$$

where $\hat{\gamma}_\pi = \frac{1}{n} \sum_{i \in N_n} \hat{e}_{\pi(i)} \hat{a}_{i,\pi}$. Then we construct $\hat{\sigma}_\pi^2$ and $\hat{\sigma}_{+,\pi}^2$ just as $\hat{\sigma}^2$ and $\hat{\sigma}_+^2$ except that $\hat{q}_{i,\pi}$'s and $\bar{q}_{i,\pi}$'s replace \hat{q}_i 's and \bar{q}_i 's. Define the permutation test statistic as

$$T_\pi = \frac{\sqrt{n} \hat{C}_\pi(G_n)}{\hat{\sigma}_{+,\pi}}, \quad (8)$$

and let

$$c_\alpha = \inf \left\{ c \in \mathbf{R} : \frac{1}{|\Pi_n|} \sum_{\pi \in \Pi_n} 1 \{ |T_\pi| \leq c \} > 1 - \alpha \right\}.$$

Hence T_π is computed just as T except that \hat{e}_i 's are replaced by $\hat{e}_{\pi(i)}$'s. Then the permutation-based two-sided confidence interval for the GC is given as

$$C_\alpha(G_n) = [\hat{C}(G_n) - c_\alpha \hat{\sigma}_+ / \sqrt{n}, \hat{C}(G_n) + c_\alpha \hat{\sigma}_+ / \sqrt{n}], \quad (9)$$

where $\hat{\sigma}_+^2$ is as defined in equation (7).

It is important that we do not use $\hat{a}_{\pi(i)}$ in constructing a permutation-based critical value, because we need to create perturbations of the graph structure through permutations. Its motivation is easier to see in the case of the problem of testing for the null of zero graph concordance. In order for the test to work, the resulting graph structure should be totally different from the original graph after a permutation, so that under the null of no graph concordance, both the original test statistic and the permutation test statistic behave similarly, but under the alternative hypothesis, where the graph governs the cross-sectional dependence structure, the permutation test statistic should behave differently from the original test statistic, thereby giving power to the test. If the graph structure remains almost invariant after the permutations, the permutation test will not have much power. That is why we use the same neighborhood structure but permute only the sample index of \hat{e}_i . We cannot use $\hat{a}_{\pi(i)}$ because then the neighbor structure will also be permuted in tandem with the node indices, leaving the graph structure unchanged and giving no nontrivial power to the test.

B. Inference on Residual Graph Concordance

Let us see how the previous approach applies to the residual GC. Unlike the GC for Y_i 's, the residual GC involves u_i 's that are not observed. To deal with this, we first take

$$\hat{u}_i = Y_i - X_i' \hat{\beta},$$

where $\hat{\beta}$ is the usual least squares estimator of β , that is, $\hat{\beta} = (X'X)^{-1}X'y$, with X being the $n \times d$ matrix whose j th row is given by X_j' and y the n -dimensional vector whose j th entry is given by Y_j . Hence, we have for each $i \in N_{n,x}$,

$$\hat{u}_i = u_i - x'(X'X)^{-1}X'u, \quad (10)$$

where $u = [u_1, \dots, u_n]'$. Let us take

$$\hat{C}_x(G_n) = \hat{\gamma}_x - \hat{\gamma}_x^c$$

and

$$\hat{\gamma}_x = \frac{1}{|N_{n,x}|} \sum_{i \in N_{n,x}} \hat{e}_{x,i} \hat{a}_{x,i} \text{ and } \hat{\gamma}_x^c = \frac{1}{|N_{n,x}|} \sum_{i \in N_{n,x}} \hat{e}_{x,i} \hat{a}_{x,i}^c$$

with the following definitions:

$$\hat{e}_{x,i} = \frac{\hat{u}_{x,i} - \bar{u}_x}{\hat{v}_x}, \quad \hat{a}_{x,i} = \frac{1}{d_{n,x}(i)} \sum_{j \in N_{n,x}(i)} \hat{e}_{x,j}, \text{ and}$$

$$\hat{a}_{x,i}^c = \frac{1}{|N_{n,x} \setminus \bar{N}_{n,x}(i)|} \sum_{j \in N_{n,x} \setminus \bar{N}_{n,x}(i)} \hat{e}_{x,j},$$

and \bar{u}_x denotes the average of $\hat{u}_{x,i}$'s over $i \in N_{n,x}$ and \hat{v}_x the average of $(\hat{u}_i - \bar{u}_x)$'s over $i \in N_{n,x}$. Thus we consider the following test statistic:

$$T_x = \frac{\sqrt{n}(\hat{C}_x(G_n) - C_x(G_n))}{\hat{\sigma}_{x,+}}$$

where $\hat{\sigma}_{x,+}$ is constructed in the same way as we constructed $\hat{\sigma}_+$ except that we use \hat{u}_i in place of Y_i 's, $N_{n,x}$ in place of N_n , and $N_{n,x}(i)$ in place of $N_n(i)$.

Permutation-based inference can proceed similarly as before. Let $\Pi_{n,x}$ be the collection of permutations on $N_{n,x}$. The permutation test statistic $T_{x,\pi}$ is obtained for each $\pi \in \Pi_{n,x}$ by replacing $\hat{e}_{x,i}$'s in T by $\hat{e}_{x,\pi(i)}$'s.

Due to equation (10), we can see that $\hat{C}_x(G_n)$ and $\hat{e}_{x,i}$ remain the same if we replace \hat{u}_i by u_i for all $i \in N_{n,x}$. Therefore if u_i 's are exchangeable conditional on X , the permutation-based inference is valid in finite samples. The asymptotic validity of the permutation inference also follows when u_i 's satisfy the conditions specified for Y_i 's in the section III C.

C. Asymptotic Validity

When the random vector $Y = (Y_i)_{i=1}^n$ is exchangeable—the joint distribution of $Y_\pi = (Y_{\pi(i)})_{i=1}^n$ is the same as that of $Y = (Y_i)_{i=1}^n$ (and consequently, the joint distribution of $(\hat{e}_{\pi(i)})_{i=1}^n$ is the same as that of $(\hat{e}_i)_{i=1}^n$) for any permutation $\pi \in \Pi_n$ —the confidence interval is valid in finite samples. However, when Y is exchangeable, we have $C(G'_n) = 0$ for any graph G'_n such that for each $i \in N_n$, $d'_n(i) \leq n - d'_n(i)$, $d'_n(i)$ denoting the degree of i in G'_n . Therefore, this represents a strong form of graph irrelevance of Y . For validity of the confidence intervals, we need to cover the case $C(G_n) \neq 0$ as well.

This paper's main result establishes conditions under which the confidence interval $\mathcal{C}_\alpha(G_n)$ is asymptotically valid, even when Y fails to be exchangeable. The following definition clarifies which properties of the graph G_n are relevant for this purpose. Recall that a degree (denoted by $d_n(i)$) of a vertex i refers to the number of the edges vertex i has in G_n , and the maximum degree (denoted by $d_{mx,n}$), the maximum over the degrees of the vertices in N_n . We define

$$d_{avi,n} = \frac{1}{n} \sum_{i \in N_n: d_n(i) \geq 1} \frac{1}{d_n(i)},$$

that is, the average of the inverse degrees. Also define $d_{mx,n,3}$ to be the maximum number of the vertices within three edges from a fixed vertex. We call $d_{mx,n,3}$ the maximum 3-degree of graph G_n .

Definition 1. For a positive integer $d_{mx,n,3}$, let $\mathcal{G}_n(d_{mx,n,3})$ be the collection of graphs having the maximum 3-degree equal to $d_{mx,n,3}$.

Let us introduce the population versions of \hat{e}_i 's, \hat{a}_i 's, and \hat{a}_i^c 's,

$$e_i = \frac{Y_i - \mathbf{E}[Y_i]}{v_n}$$

and

$$a_i = \frac{1}{d_n(i)} \sum_{j \in N_n(i)} e_j \text{ and } a_i^c = \frac{1}{|N_n \setminus \bar{N}_n(i)|} \sum_{j \in N_n \setminus \bar{N}_n(i)} e_j.$$

The following assumption is concerned with the nondegeneracy of the limiting distribution of the test statistic:

Assumption 1 (Nondegeneracy and Moment Conditions).

There exist small $c > 0$ and large $M > 0$ such that the following is satisfied for all $n \geq 1$:

- i. $v_n^2 > c$.
- ii. For $q_i = e_i\{a_i - e_i\gamma\}$,

$$\mathbf{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{i \in N_n} (q_i - \mathbf{E}q_i) \right)^2 \right] > c \text{ and}$$

$$\frac{d_{avi,n}}{|\tilde{N}_n|} \sum_{ij \in \tilde{N}_n} \mathbf{E}[e_i^2 e_j^2] > c. \tag{11}$$

- iii. $\max_{1 \leq i \leq n} \mathbf{E}|Y_i|^8 < M$.

The first condition in equation (11) ensures that the variance of the leading term in the asymptotic linear representation is positive. The second condition in equation (11) is a technical, mild condition that essentially requires that $d_{avi,n}$ be bounded away from 0 from some large n on. This condition is mostly satisfied by many random graph models (such as Erdős-Rényi or Barabási-Albert graphs) whose degree distribution becomes tight and nondegenerate in the limit. To see this, we let

$$N_{n,d} = \{i \in N_n : d_n(i) = d\}$$

and rewrite

$$d_{avi,n} = \sum_{d=1}^{d_{mx,n}} \frac{1}{d} \frac{|N_{n,d}|}{n} \geq \frac{1}{k} \sum_{d=1}^k \frac{|N_{n,d}|}{n}$$

for any fixed positive integer $k \leq d_{mx,n}$. Therefore, $d_{avi,n}$ is bounded away from 0 whenever there is a positive fraction of nodes having degrees less than or equal to k for some fixed k . This condition is violated when all but an asymptotically negligible fraction of nodes in the network become either isolated or have a degree that goes to infinity as $n \rightarrow \infty$.

Assumption 2 (Symmetry in Means).

- i. $\mathbf{E}Y_i$'s are the same for all i 's in N_n .
- ii. For each $d = 1, \dots, d_{mx,n}$, there exists $r_{n,d} \in \mathbf{R}$ such that $\mathbf{E}q_i = r_{n,d}$ for all $i \in N_{n,d}$.

Assumption 2i requires that the expected values of Y_i 's are the same across i 's, which is weaker than the assumption that Y_i 's are identically distributed. Assumption 2ii requires that the expectation of q_i be the same for all i 's having the same degree. When i and j have the same degree, q_i and q_j are sums of the same number of random variables of the form $e_i e_k$ and $e_j e_l$. For example, suppose that Y_i 's have the same variance, and for each d , the correlation between Y_i and \bar{Y}_i is the same for all $i \in N_{n,d}$. Then assumption 2ii is satisfied.

For example, consider the data-generating process for Y_i as in equation (4) in the example of treatment outcomes through network interference. Since (D_i, η_i) 's are i.i.d. across i 's, the joint distribution of $((\bar{D}_j, D_j, \eta_j)_{j \in N_n(i)}, D_i, \eta_i)$ is the same across all i 's that have the same degree. This means that the joint distribution of (Y_i, \bar{Y}_i) is the same across all i 's that have the same degree. Thus assumption 2ii is satisfied in this case. (The i.i.d. assumption for $[D_i, \eta_i]_{i \in N_n}$ and the linearity in the model also ensure that assumption 2i is satisfied.)

Also, consider the residual graph concordance introduced in a preceding section. Recall that since the sample version of GC remains the same regardless of whether we use \hat{u}_i 's or u_i 's, u_i plays the role of Y_i , and the subset $N_{n,x}$ of i 's such that $X_i = x$ plays the role of N_n . Assumption 2 is satisfied as long as the conditional distribution of (u_i, \bar{u}_i) given $X_i = x$ is the same across $i \in N_{n,x}$ having the same degree.

We introduce a notion of a joint dependence pattern for Y shaped by graph G_n . We say that $Y = (Y_i)_{i=1}^n$ has a dependency graph $G_n = (N_n, E_n)$ if for any two disjoint subsets $A, A' \subset N_n$ having no edge in E_n such that one end vertex is in A and the other end vertex is in A' , two random vectors $(Y_i)_{i \in A}$ and $(Y_i)_{i \in A'}$ are independent. (See Penrose, 2003.) When Y has G_n as a dependency graph, it implies that any two disjoint sets of Y_i 's having no edge between the two sets are independent. Therefore, Y_i and Y_j can be correlated only when they are adjacent. The joint dependence of linked pairs Y_i and Y_j can be heterogeneous across the linked pairs.⁴

⁴Under the dependency graph assumption, we have $\gamma^c = 0$; hence we can impose it in the estimator of GC by setting $\hat{\gamma}^c = 0$. This does not alter the asymptotic distribution and the variance estimators. We keep the original definition of GC as involving γ^c there because a different inference procedure that does not invoke a dependency graph may be possible.

Let us define the class of joint distributions of Y to be considered. Let \mathcal{P}_n be the collection of joint distributions of Y .

Definition 2. For $c, M > 0$ and a graph G_n , we define $\mathcal{P}_n(G_n; c, M)$ to be the set of the joint distributions of $Y = (Y_i)_{i=1}^n$ such that under each $P \in \mathcal{P}_n(G_n; c, M)$, assumptions 1 and 2 are satisfied with (c, M) and G_n , and G_n is a dependency graph for Y .

The following theorem is the main result of this paper:

Theorem 1. Suppose that for each $n \geq 1$, $G_n \in \mathcal{G}_n(d_{mx,n}, 3)$ satisfying that as $n \rightarrow \infty$,

$$\frac{d_{mx,n,3}^4}{n} \rightarrow 0. \quad (12)$$

Then for each $c, M > 0$, and for each sequence $P_n \in \mathcal{P}_n(G_n; c, M)$,

$$\lim_{n \rightarrow \infty} |P_n\{C(G_n) \in \mathcal{C}_\alpha(G_n)\} - (1 - \alpha)| = 0.$$

A notable aspect of theorem 1 is that despite the fact that the permutation test statistic does not involve centering explicitly, the confidence interval is still asymptotically valid regardless of whether $C(G_n) = 0$.

The condition in equation (12) allows for the maximum degree $d_{mx,n}$ to increase at a certain polynomial rate. One should not interpret the condition (12) as part of a description of the way the network forms and grows in reality. The condition should be used to gauge the finite sample environment in which the validity of inference justified assuming large n becomes reliable in finite samples.

It is not hard to see that the linear network interference example in section IID satisfies the dependency graph assumption. The dependency graph assumption is also compatible with the notion of homophily among people in the following sense. Homophily refers to the tendency of people being associated with each other more often when they are of the same social group such as race or religion. Suppose that we have two race categories, W, B , where each individual i is given race indicator $D_i \in \{W, B\}$, and the observed graph G_n is a subgraph of a very large and sparse graph, say, $G_L = (N_L, E_L)$ where L is much larger than n . Also, assume that (D_i, D_j) 's have the same joint distribution across $ij \in E_L$ and the same joint distribution across $ij \notin E_L$ such that $i \neq j$. Homophily says that for a person i with race $D_i = t$, person j , if the person is a friend of person i , is more likely to be found to be of the same race than of the different race, that is, for $ij \in E_L$,

$$P\{D_j \neq t | D_i = t\} < P\{D_j = t | D_i = t\}.$$

The dependency graph then imposes that a person j who is not a friend of i is as likely to be of the same race as to

be of the different race. In other words, there is no correlation of races between people who are not linked. This latter assumption is plausible when Y_i 's are locally dependent so that the law of the large numbers holds. For $t \in \{W, B\}$, let $Y_i(t) = 1\{D_i = t\}$ and

$$\bar{Y}_{i,L}^c(t) = \frac{1}{|N_L \setminus \bar{N}_L(i)|} \sum_{j \in N_L \setminus \bar{N}_L(i)} Y_j(t).$$

Then for $t_1, t_2 \in \{W, B\}$, and for $j \in N_L \setminus \bar{N}_L(i)$,

$$\begin{aligned} P\{D_i = t_1, D_j = t_2\} &= \mathbf{E}[Y_i(t_1)Y_j(t_2)] \\ &= \mathbf{E}[Y_i(t_1)\bar{Y}_{i,L}^c(t_2)] \\ &\approx \mathbf{E}[Y_i(t_1)\mathbf{E}[\bar{Y}_{i,L}^c(t_2)]] \\ &= P\{D_i = t_1\}P\{D_j = t_2\}, \end{aligned}$$

where the approximation above comes from the law of the large numbers. The approximation error will be negligible when L is much larger than n .⁵

The proposal of this paper is still applicable when Y exhibits strong dependence through several common shocks (or aggregate shocks) that affect all the nodes globally, as long as Y_i 's have G_n as a dependency graph conditional on the common shocks. In this case, when it comes to measuring the relevance of G_n in explaining the cross-sectional dependence of Y_i 's, this paper's view is that our focus should be on the conditional version of graph concordance given common shocks, where the probability (in both the definition of graph concordance and the coverage probability of the confidence set) is now replaced by the conditional probability given the common shocks. There are two reasons for this view. First, our interest is not merely in detecting the presence of cross-sectional dependence but the presence of a dependence pattern that is shaped by a given network. Therefore, it is better to condition on the source of cross-sectional dependence that arises for a reason unrelated to the network. Second, as we make only a single observation of Y , there is no hope of recovering unconditional probability in general that takes into account variations in the common shocks, because such variations are not observed in the data.⁶

D. Testing for the Graph Concordance of Cross-Sectional Observations

Suppose that we would like to test whether $C(G_n) > 0$. The null and alternative hypotheses of interest in this paper take the following form:

$$H_0 : C(G_n) \leq 0, \text{ against } H_1 : C(G_n) > 0.$$

⁵This does not mean that the dependency graph assumption is innocuous, because it assumes more than pairwise independence of Y_i 's which are not linked.

⁶See Andrews (2005) for examples of linear models where inference based on unconditional probability is possible using only cross-sectional observations despite the presence of common shocks.

Similarly as before, we define

$$T_1 = \frac{\sqrt{n}\hat{C}(G_n)}{\hat{\sigma}_+},$$

which is a version of T with the restriction $C(G_n) = 0$ imposed. For critical values, we use T_π defined previously and find a critical value $c_{\alpha,1}$ as follows:

$$c_{\alpha,1} = \inf \left\{ c \in \mathbf{R} : \frac{1}{|\Pi_n|} \sum_{\pi \in \Pi_n} 1\{T_\pi \leq c\} > 1 - \alpha \right\},$$

which is a one-sided version of c_α . Then we can perform the test by rejecting the null hypothesis if and only if $T_1 > c_{\alpha,1}$.

The null hypothesis is weaker than the exchangeability of Y . Hence, the permutation test does not preserve finite sample validity in all the cases of the null hypothesis. The following theorem shows that the permutation test controls the size asymptotically under the null hypothesis.

Let $\mathcal{P}_{n,0}(G_n)$ be the collection of the probabilities under which $C(G_n) \leq 0$, and let $\mathcal{P}_{n,00}(G_n)$ be the collection of the probabilities under which $C(G_n) = 0$. Also, let $\mathcal{P}_{n,1}(G_n; b)$ be the collection of the probabilities under which $C(G_n) \geq b$.

Theorem 2. (i) *Suppose that the conditions of theorem 1 hold. Then for each $c, M > 0$, and for each sequence $P_n \in \mathcal{P}_n(G_n; c, M) \cap \mathcal{P}_{n,0}(G_n)$,*

$$\limsup_{n \rightarrow \infty} P_n \{T_1 > c_{\alpha,1}\} \leq \alpha.$$

Furthermore, for each sequence $P_n \in \mathcal{P}_n(G_n; c, M) \cap \mathcal{P}_{n,00}(G_n)$,

$$\lim_{n \rightarrow \infty} |P_n \{T_1 > c_{\alpha,1}\} - \alpha| = 0.$$

(ii) *For each $c, M > 0$, and for each sequence $P_n \in \mathcal{P}_n(G_n; c, M) \cap \mathcal{P}_{n,1}(G_n; b)$ with $b > 0$,*

$$\liminf_{n \rightarrow \infty} P_n \{T_1 > c_{\alpha,1}\} = 1.$$

The proof of theorem 2 comes from the arguments used for proving theorem 1. Details are omitted for brevity.

IV. Monte Carlo Simulation Studies

A. Data-Generating Process

For the graph in the simulation study, we consider two classes of random graphs, each having size $n \in \{300, 600\}$. The first class is an Erdős-Rényi (E-R) random graph, where each pair of the vertices forms an edge with equal probability $p_n = \lambda/(n-1)$, where λ is chosen from $\{1, 3, 5\}$. Each vertex from this random graph has neighbors of size λ on average. The distribution of the size of the neighborhood is approximately a Poisson distribution with parameter λ when n is large. The second class is a Barabási-Albert (B-A)

TABLE 1.—DEGREE CHARACTERISTICS OF THE GRAPHS AND THE TRUE GCs USED IN THE SIMULATION STUDY

	E-R Graph				B-A Graph			
	$n = 300$		$n = 600$		$n = 300$		$n = 600$	
	$\lambda = 1$	$\lambda = 5$	$\lambda = 1$	$\lambda = 5$	$m = 1$	$m = 3$	$m = 1$	$m = 3$
Maximum degree	5	12	6	13	22	42	40	70
Average degree	0.960	4.807	0.920	5.133	1.927	5.620	1.950	5.810
True GC ($c = 0.3$)	0.056	0.073	0.052	0.072	0.078	0.068	0.078	0.068
True GC ($c = 0.6$)	0.233	0.171	0.210	0.159	0.260	0.163	0.262	0.167

The table gives the network characteristics of the graph that was used for the simulation study. The simulation study was based on a single generation of the random graphs. The E-R represents an Erdős-Rényi random graph (E-R graph) with probability equal to $p = \lambda/(n-1)$, where λ is chosen from 1, 3, 5 and the B-A represents a Barabási-Albert random graph (B-A graph) of preferential attachment, where the parameter m refers to the number of links each new node forms with other existing nodes. The true GCs were computed from 1 million simulations using the single realization of the random graphs.

random graph of preferential attachment. To generate this random graph, we began with an Erdős-Rényi random graph of size 20 with $\lambda = 1$. Then we let the graph grow by adding each vertex sequentially and let the vertex form edges with m other existing vertices. (We chose m from $\{1, 2, 3\}$.) The probability of a new vertex forming an edge with an existing vertex is proportional to the number of the neighbors of the existing vertex. We keep adding new vertices until the size of the graph becomes n . We generate the graphs once and fix them when we generate the outcomes. The graph characteristics used in the simulation study are summarized in table 1.

As for the data-generating process for the outcomes, we first generate $\{Y_i\}_{i=1}^n$ i.i.d. from $N(0, 1)$ under the null hypothesis. Under the alternative hypothesis, we generate Y_i as follows. We first generate $\{Y_i^*\}_{i=1}^n$ i.i.d. from $N(0, 1)$. Let $E = \{e_1, \dots, e_s\}$ be the set of edges in the graph generated as previously. We remove redundant edges from E (i.e., remove ji with $j < i$) and let M be a two-column matrix whose entries are of the form $[i_s, j_s]$ for $e_s = i_s j_s$. Let M be sorted on the first column so that $i_s \leq i_{s+1}$:

Step 1: For $s = 1$, such that $e_1 = i_1 j_1$, we draw $Z_1 \sim N(0, 1)$ and set

$$(Y_{i_1}, Y_{j_1}) = \sqrt{1 - c^2} \times (Y_{i_1}^*, Y_{j_1}^*) + c \times Z_1,$$

where c is a parameter that determines the strength of the stochastic dependence of linked outcomes. (When c is away from 0, the variable Z_1 serving as a common factor to Y_{i_1} and Y_{j_1} induces correlation between Y_{i_1} and Y_{j_1} , and this correlation increases in c other things being equal.) We replace $(Y_{i_1}^*, Y_{j_1}^*)$ by (Y_{i_1}, Y_{j_1}) and redefine the series $\{Y_i^*\}_{i=1}^n$.

Step s : For $s > 1$ such that $e_s = (i_s, j_s)$, we draw $Z_s \sim N(0, 1)$ and set

$$(Y_{i_s}, Y_{j_s}) = \sqrt{1 - c^2} \times (Y_{i_s}^*, Y_{j_s}^*) + c \times Z_s.$$

We replace $(Y_{i_s}^*, Y_{j_s}^*)$ by (Y_{i_s}, Y_{j_s}) and redefine the series $\{Y_i^*\}_{i=1}^n$.

This algorithm generates $\{Y_i\}_{i=1}^n$ with positive graph concordance when $c > 0$. When $c = 0$, this series is i.i.d. from $N(0, 1)$. As for the parameter c , we consider values of 0, 0.3, and 0.6. Note that c does not necessarily represent the correlation between Y_i and Y_j for $ij \in E_n$, because we keep adding $c \times Z_s$ to Y_i as we run along the neighbors of vertex i .

For the confidence intervals, the E-R and B-A graphs were generated once, and based on these graphs, the true GCs were computed using 1 million Monte Carlo simulations. The graph characteristics and the true GCs are presented in table 1.

The number of Monte Carlo simulations in the investigation of the finite sample properties was set to be 5,000 and the number of random permutations used to construct critical values was set to be 1,000.

B. Results and Discussion

The empirical coverage probabilities of permutation-based confidence intervals are reported in table 2. The probabilities were computed for nominal level 95%. When $c = 0$ or $c = 0.3$, the finite sample coverage probabilities are close to 0.95. However, when $c = 0.6$, the coverage probabilities are slightly lower than 0.95 when the graph becomes denser.

Table 2 also reports the mean length of the permutation-based confidence intervals from 5,000 simulations. The lengths of the confidence intervals are not comparable directly across different graphs and different c 's because the GCs are different. However, we can see that there is a substantial reduction in the mean length of the confidence intervals when we increase the sample size from 300 to 600, indicating more information from more samples. The reduction in the mean length is most substantial when the graph has fewer edges.

The results using confidence intervals based on asymptotic normal approximation are shown in table 3. Using asymptotic critical values leads to coverage probabilities lower than the nominal level of 0.95. This undercoverage becomes more severe when the graph becomes denser. For example, when $n = 300$ and a B-A graph with $m = 3$ was used, the asymptotic coverage probability in the case of $c = 0$ is only 0.7354, while that of the permutation confidence interval is 0.9496.

TABLE 2.—EMPIRICAL COVERAGE PROBABILITY AND MEAN LENGTH OF PERMUTATION-BASED CONFIDENCE INTERVALS OF THE GC AT 95% NOMINAL LEVEL

Coverage Probability		E-R			B-A		
c		$\lambda = 1$	$\lambda = 3$	$\lambda = 5$	$m = 1$	$m = 2$	$m = 3$
0	$n = 300$	0.9470	0.9512	0.9466	0.9528	0.9472	0.9496
	$n = 600$	0.9516	0.9534	0.9542	0.9476	0.9534	0.9500
0.3	$n = 300$	0.9518	0.9488	0.9496	0.9460	0.9534	0.9476
	$n = 600$	0.9494	0.9480	0.9508	0.9502	0.9514	0.9416
0.6	$n = 300$	0.9456	0.9462	0.9502	0.9488	0.9528	0.9498
	$n = 600$	0.9496	0.9500	0.9456	0.9528	0.9564	0.9578
Mean Length		E-R			B-A		
c		$\lambda = 1$	$\lambda = 3$	$\lambda = 5$	$m = 1$	$m = 2$	$m = 3$
0	$n = 300$	0.2283	0.2040	0.1979	0.2515	0.2221	0.2031
	$n = 600$	0.1503	0.1428	0.1234	0.1749	0.1434	0.1350
0.3	$n = 300$	0.2276	0.2022	0.1952	0.2506	0.2211	0.2007
	$n = 600$	0.1498	0.1415	0.1214	0.1741	0.1429	0.1326
0.6	$n = 300$	0.2189	0.1876	0.1847	0.2482	0.2101	0.1930
	$n = 600$	0.1458	0.1322	0.1161	0.1774	0.1401	0.1280

E-R represents an Erdős-Rényi random graph (E-R graph) with probability equal to $p = \lambda/(n - 1)$, where λ is chosen from 1, 3, 5, and B-A represents a Barabási-Albert random graph (B-A graph) of preferential attachment, where the parameter m refers to the number of the links each new node forms with other existing nodes.

TABLE 3.—EMPIRICAL COVERAGE PROBABILITY AND MEAN LENGTH OF CONFIDENCE INTERVALS OF THE GC FROM ASYMPTOTIC NORMAL DISTRIBUTION AT 95% NOMINAL LEVEL

Coverage Probability		E-R			B-A		
c		$\lambda = 1$	$\lambda = 3$	$\lambda = 5$	$m = 1$	$m = 2$	$m = 3$
0	$n = 300$	0.9292	0.8912	0.8228	0.9112	0.8262	0.7354
	$n = 600$	0.9416	0.9330	0.8786	0.9224	0.8732	0.7914
0.3	$n = 300$	0.9326	0.8860	0.8334	0.9072	0.8212	0.7404
	$n = 600$	0.9376	0.9282	0.8730	0.9264	0.8614	0.7938
0.6	$n = 300$	0.9268	0.8884	0.8300	0.9080	0.8236	0.7518
	$n = 600$	0.9414	0.9284	0.8668	0.9322	0.8848	0.8018
Mean Length		E-R			B-A		
c		$\lambda = 1$	$\lambda = 3$	$\lambda = 5$	$m = 1$	$m = 2$	$m = 3$
0	$n = 300$	0.2111	0.1620	0.1205	0.2125	0.1300	0.0889
	$n = 600$	0.1439	0.1303	0.0895	0.1586	0.1006	0.0702
0.3	$n = 300$	0.2104	0.1606	0.1191	0.2119	0.1294	0.0876
	$n = 600$	0.1435	0.1292	0.0881	0.1579	0.1000	0.0689
0.6	$n = 300$	0.2023	0.1490	0.1126	0.2096	0.1227	0.0844
	$n = 600$	0.1396	0.1206	0.0841	0.1606	0.0981	0.0667

The confidence intervals from asymptotic normal distribution are not stable. For example, see the case of E-R graphs with $\lambda = 5$ and $c = 0$, where the coverage probability is only 0.8228 (with $n = 300$) and 0.8786 (with $n = 600$), when the nominal coverage probability is 0.95. Also see the case of B-A graphs with $m = 3$ with $c = 0$, where the coverage probability is only 0.7354 (with $n = 300$) and 0.7914 (with $n = 600$).

This illustrates the merit of the permutation-based approach over the asymptotic method.

V. Empirical Application: Social Networks in Indian Villages

A. Data

The data used here originated from the data used in Banerjee et al. (2013). In 2006, the social network data were collected for 75 rural villages in Karnataka, a state in southern India. The data contained household information such as caste, village leader indicator, and various amenities (e.g., roofing material, type of latrine). The social network data were collected along twelve dimensions in terms of visiting each other; kinship; borrowing or lending money, rice, or kerosene; giving or exchanging advice; or going to the

place of prayer together. In 2007, a microfinancing institution, Bharatha Swamukti Samsthe (BSS), began operations in 43 of these villages and collected data on who joined the microfinancing program. (See Banerjee et al., 2013, for details.)

Out of 75 villages in the original sample, we selected villages that have at least one minority caste or tribe household. The resulting villages contain a total of 10,176 households. As for the definitions of the social network, we considered three. The graph $G_{n,ALL}$ is defined to be the social network where two households are linked if and only if any of the twelve dimensions in the social network data occurred between the households. The graph $G_{n,EE}$ is defined to be the social network where two households are linked if and only if material exchanges (borrowing or lending rice, kerosene, or money) occurred, and the graph $G_{n,SC}$ is defined to be the social network where two households are linked if and only if some social activities (such as seeking advice or going to

TABLE 4.—SOCIAL NETWORK CHARACTERISTICS FROM DATA ON INDIAN VILLAGES

Homophily Study					
	Networks	Size	Maximum Degree	Average Degree	Median Degree
	$G_{n,EE}$	10,176	56	5.7545	5
	$G_{n,SC}$	10,176	87	8.4754	7
	$G_{n,ALL}$	10,176	90	9.1824	8
Microfinance Study					
Data Set	Networks	Size	Maximum Degree	Average Degree	Median Degree
A	$G_{n,EE}$	7,537	56	5.80	5
	$G_{n,SC}$	7,537	87	8.65	7
	$G_{n,ALL}$	7,537	90	9.39	8
B	$G_{n,EE}$	5,915	56	5.88	5
	$G_{n,SC}$	5,915	71	8.73	7
	$G_{n,ALL}$	5,915	80	9.47	8

The network $G_{n,EE}$ is defined based on observed material exchanges between households (such as borrowing or lending rice, kerosene, or money). The network $G_{n,SC}$ is defined based on social activities (such as seeking advice or going to the same temple or church). The network $G_{n,ALL}$ is the union of the two networks $G_{n,EE}$ and $G_{n,SC}$. The microfinancing study uses two data sets, A and B. Data set A consists of households in the villages that have at least one minority household. Data set B consists of households in data set A that are in villages whose participation rate in microfinancing is at least 10%.

TABLE 5.—ESTIMATED GC OF MINORITY CASTES/TRIBES AND MICROFINANCING DECISIONS ALONG THE SOCIAL NETWORKS

GC for Minority					
Networks	GC	Confidence Intervals			
		99%	95%	90%	
$G_{n,EE}$	0.5657	[0.5141,0.6174]	[0.5271,0.6043]	[0.5331,0.5984]	
$G_{n,SC}$	0.5936	[0.5385,0.6488]	[0.5512,0.6361]	[0.5583,0.6290]	
$G_{n,ALL}$	0.5813	[0.5231,0.6395]	[0.5370,0.6256]	[0.5446,0.6180]	
GC for Microfinancing					
Data Set	Networks	GC	Confidence Intervals		
			99%	95%	90%
A	$G_{n,EE}$	0.1386	[0.1006,0.1765]	[0.1105,0.1666]	[0.1156,0.1615]
	$G_{n,SC}$	0.1302	[0.0974,0.1630]	[0.1055,0.1549]	[0.1091,0.1513]
	$G_{n,ALL}$	0.1257	[0.0940,0.1574]	[0.1002,0.1512]	[0.1043,0.1471]
B	$G_{n,EE}$	0.1345	[0.0872,0.1818]	[0.1022,0.1668]	[0.1087,0.1603]
	$G_{n,SC}$	0.1225	[0.0881,0.1569]	[0.0957,0.1493]	[0.0999,0.1451]
	$G_{n,ALL}$	0.1182	[0.0832,0.1532]	[0.0924,0.1440]	[0.0964,0.1401]

the same temple or church) occurred. The network statistics are given in table 4.

B. Application 1: Homophily along Castes/Tribes

Jackson (2014) observed homophily of households from the same Indian village data when the social network is defined based on borrowing or lending of kerosene and rice. Similarly in Jackson (2014), we define a minority caste or tribe household to be the household reported to be a scheduled caste or scheduled tribe, which is considered for affirmative action by the Indian government Jackson, 2014. Here we use the GC of this minority indicators along with social networks as a measure of homophily and provide formal inference results.

The results are shown in table 5. The estimated GC is from 0.5657 to 0.5936. The GC is very high, indicating strong homophily along the castes. This strong homophily is robust to alternative definitions of the social networks. In particular, the estimated GC is significantly larger than zero, regardless of whether the social network is that of economic exchanges

or of social activities. The permutation-based confidence intervals are narrow around the point estimates, supporting the strong homophily.

C. Application 2: Graph Concordance of Microfinancing Decisions

Banerjee et al. (2013) investigated the role of the social networks in microfinancing decisions by households, modeling the microfinancing decisions using a logit model. We search for evidence that the social network plays any role in explaining the cross-sectional dependence of microfinancing decisions without modeling any parametric or nonparametric functional relationships among the variables. The social network can play a role in affecting the cross-sectional pattern of microfinancing decisions in two ways. First, it can serve as a channel for information diffusion from one household to another. Second, the edges in the social network merely reflect the common social context that the linked households share and affect the microfinancing decisions. A necessary implication of either way would be that the GC of

microfinancing decisions along the social network should be positive. Otherwise the role of the social network in either way may lack empirical support.

The estimation results of GC are reported in table 5. The estimated GC ranges from 0.1182 to 0.1386. The confidence intervals of all the GCs are substantially away from 0. This suggests that the social network plays a role in explaining the microfinancing decisions. Interestingly, the GC is strongest when we considered $G_{n,EE}$ —the social network defined in terms of material exchanges among households in the minority groups.

VI. Conclusion

This paper proposes a new measure that gauges the relevance of a given graph in explaining the cross-sectional dependence of an observed random vector laid on the graph. Examples include homophily on social networks and network interference of treatments. This paper develops a method of permutation inference for the measure and shows that this method exhibits two-tier validity.

The inference procedure does not require a complex set of conditions or a delicate choice of tuning parameters. Furthermore, the procedure does not presume any functional relationships among observables, and hence is suitable for a diagnostic purpose in investigating the role of a graph in explaining the cross-sectional dependence of an observed random vector.

REFERENCES

Andrews, D. W. K., “Cross-Section Regression with Common Shocks,” *Econometrica* 73 (2005), 1551–1585.
 Aronow, P., and C. Samii, “Estimating Average Causal Effects under Interference between Units,” working paper (2015). arXiv: 1305.6156v2[mathST].
 Athey, S., D. Eckles, and G. W. Imbens, “Exact p -Values for Network Interference,” NBER working paper 21313 (2015).
 Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, “The Diffusion of Microfinance,” *Science* 341:6144 (2013), 1236498.
 Cameron, A. C., and P. K. Trivedi, *Regression Analysis of Count Data*, 2nd ed. (2013).
 Canay, I. J., P. Romano, and A. M. Shaikh, “Randomization Tests under an Approximate Symmetry Assumption,” *Econometrica* 85 (2017), 1013–1030.
 Chung, E., and J. P. Romano, “Exact and Asymptotically Robust Permutation Tests,” *Annals of Statistics* 41 (2013), 484–507.
 Cliff, A., and J. Ord, “Testing for Spatial Autocorrelation among Regression Residuals,” *Geographic Analysis* 4 (1972), 267–284.
 Coleman, J. S., “Relational Analysis: The Study of Social Organizations with Survey Methods,” *Human Organization* 17 (1958), 28–36.
 Currarini, S., M. O. Jackson, and P. Pin, “An Economic Model of Friendship: Homophily, Minorities and Segregation,” *Econometrica* 77 (2009), 1003–1045.
 Daniels, H. E., “The Relation between Measures of Correlation in the Universe of Sample Permutation,” *Biometrika* 33 (1944), 129–135.
 Delgado, M. A., “Testing Serial Independence Using the Sample Distribution Function,” *Journal of Time Series Analysis* 17 (1996), 271–285.
 Fershtman, C., and N. Gandal, “Direct and Indirect Knowledge Spillovers: The ‘Social Network’ of Open-Source Projects,” *RAND Journal of Economics* 42 (2011), 70–91.
 Friedman, J. H., and L. C. Rafsky, “Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests,” *Annals of Statistics* 7 (1979), 697–717.

Hoeffding, W., “The Large-Sample Power of Tests Based on Permutations of Observations,” *Annals of Mathematical Statistics* 23 (1952), 169–192.
 Hsiao, C. M., H. Pesaran, and A. Pick, “Diagnostic Tests of Cross-Section Independence for Limited Dependent Variable Panel Models,” *Oxford Bulletin of Economics and Statistics* 74 (2012), 253–277.
 Jackson, M. O., “Networks in the Understanding of Economic Behaviors,” *Journal of Economic Perspectives* 28 (2014), 3–22.
 Jackson, M. O., and A. Wolinsky, “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory* 71 (1996), 44–74.
 Janssen, A., “Studentized Permutation Tests for Non-i.i.d. Hypotheses and the Generalized Behrens-Fisher Problem,” *Statistics and Probability Letters* 36 (1997), 9–21.
 Kelejian, H. H., and I. R. Prucha, “On the Asymptotic Distribution of the Moran I Test Statistic with Applications,” *Journal of Econometrics* 104 (2001), 219–257.
 Lehmann, E. L., and J. P. Romano, *Testing Statistical Hypotheses* (New York: Springer, 2005).
 Leung, M. P., “Treatment and Spillover Effects under Network Interference,” University of California working paper (2016).
 Moran, P., “Notes on Continuous Stochastic Phenomena,” *Biometrika* 37 (1950), 17–23.
 Neuhaus, G., “Conditional Rank Tests for the Two-Sample Problem under Random Censorship,” *Annals of Statistics* 21 (1993), 1760–1779.
 Penrose, M., *Random Geometric Graphs* (New York: Oxford University Press, 2003).
 Pesaran, M. H., “General Diagnostic Tests for Cross Section Dependence in Panels,” CESifo working paper 1229 (2004).
 Robinson, P. M., “Correlation Testing in Time Series, Spatial and Cross-Sectional Data,” *Journal of Econometrics* 147 (2008), 5–16.
 Romano, J. P., “Bootstrap and Randomization Tests of Some Non-parametric Hypotheses,” *Annals of Statistics* 17 (1989), 141–159.

APPENDIX

I. A Matrix Formulation of Permutation-Based Inference

This section gives a reformulation of test statistics and critical values using vectors and matrices. While this formulation is less illuminating, it is convenient for programming the inference procedure using programs such as Matlab because the programs are known to run faster when the algorithms are properly vectorized.

Define A_n to be the $n \times n$ adjacency matrix whose (i, j) th entry is given by 1 if $ij \in E_n$ and 0 otherwise. Introduce the degree vector $\mathbf{d}_n = A_n \mathbf{1}$, where $\mathbf{1}$ denotes the $n \times 1$ vector of ones. Let $\mathbf{d}_{n,inv}$ be an $n \times 1$ vector whose i th entry is $d_n(i)^{-1}$ if $d_n(i) \geq 1$ and 0 otherwise, where $d_n(i)$ is the i th entry of \mathbf{d}_n . Also let $\mathbf{d}_{n,inv}^c$ be an $n \times 1$ vector whose i th entry is $(n - 1 - d_n(i))^{-1}$.

Let $\hat{\mathbf{e}}$ be the n -dimensional vector whose i th entry is given by \hat{e}_i . Define

$$\hat{\mathbf{e}}_A = (A_n \hat{\mathbf{e}}) \odot \mathbf{d}_{n,inv} \text{ and } \hat{\mathbf{e}}_{A^c} = (A_n^c \hat{\mathbf{e}}) \odot \mathbf{d}_{n,inv}^c,$$

where A_n^c is the $n \times n$ matrix whose (i, j) th entry is given by $1\{a_{ij} = 0\} - 1\{i = j\}$, a_{ij} is the (i, j) th entry of A_n , and \odot denotes the Hadamard product. Then we can write

$$\hat{C}(G_n) = \mathbf{1}^T (\hat{\mathbf{e}} \odot (\hat{\mathbf{e}}_A - \hat{\mathbf{e}}_{A^c})) / n.$$

Now, let us turn to $\hat{\sigma}_+^2$. First, define the matrix $M_0 = \mathbf{d}_n \mathbf{1}^T - \mathbf{1} \mathbf{d}_n^T$, and let M be the matrix whose (i, j) th entry is 1 if the (i, j) th entry of M_0 is zero, and 0 otherwise. Define

$$\mathbf{s}_n = M \mathbf{1},$$

so that the i th entry of \mathbf{s}_n is equal to the number of the vertices having the same degree as vertex i . Define $\hat{\mathbf{e}}_\tau$ to be the subvector of $\hat{\mathbf{e}}$ that collects only the entries having the corresponding entries of \mathbf{s}_n greater than or equal to 1. We similarly define the subvector $\hat{\mathbf{e}}_{A,\tau}$ of $\hat{\mathbf{e}}_A$ using \mathbf{s}_n . Define M_τ to be the submatrix of M that selects only entries (i, j) of M such that the i th and the j th entries of \mathbf{s}_n are greater than or equal to 1. Let \mathbf{s}_τ be the subvector of \mathbf{s}_n that selects only the entries that are greater than or equal to 1. Then let n_τ be the number of entries in \mathbf{s}_τ . Let $\mathbf{s}_{\tau,inv}$ be the $n_\tau \times 1$ vector whose i th entry is given by the inverse of the i th entry of \mathbf{s}_τ .

Now define A_n^* to be the $n \times n$ matrix whose (i, j) th entry is 1 if the (i, j) -entry of $I_n + A_n + A_n^2 + A_n^3$ is positive, and 0 otherwise. Let

$$\hat{\gamma} = \mathbf{1}^\top (\hat{\mathbf{e}} \odot \hat{\mathbf{e}}_A) / n,$$

and define

$$\begin{aligned} \hat{\mathbf{q}} &= \hat{\mathbf{e}} \odot (\hat{\mathbf{e}}_A - \hat{\gamma} \hat{\mathbf{e}}) - M_\tau(\hat{\mathbf{e}}_\tau \odot (\hat{\mathbf{e}}_{A,\tau} - \hat{\gamma} \hat{\mathbf{e}}_\tau)) \odot \mathbf{s}_{\tau,inv}, \\ \hat{\delta}^2 &= \mathbf{1}^\top ((\hat{\mathbf{q}} \hat{\mathbf{q}}^\top) \odot A_n^*) \mathbf{1}, \\ \hat{\delta}_1^2 &= \hat{\mathbf{q}}^\top \hat{\mathbf{q}} / n. \end{aligned}$$

Then we construct

$$\hat{\delta}_+^2 = \hat{\delta}^2 \mathbf{1} \{\hat{\delta}^2 > 0\} + \hat{\delta}_1^2 \mathbf{1} \{\hat{\delta}^2 \leq 0\}.$$

Once this procedure is programmed, the permutation version is simply obtained by permuting the indices of $\hat{\mathbf{e}}$ and using the vector instead of $\hat{\mathbf{e}}$ in the above computations.

2. Asymptotic Analysis through Conditioning on Permutations

Once we fix a permutation $\pi \in \Pi_n$, the permuted observations inherit their dependence structure and heterogeneity of marginals under the relabeling of observations through π . When the relabeling of the observations does not alter the limiting properties of the test statistic, the asymptotic analysis can be performed by considering a fixed permutation asymptotically first and then by taking care of the randomness of the permutation.

We introduce some notation. Let $\mathcal{X}_n \subset \mathbf{R}^m$ be a measurable set from which random vector X_n takes values. Let \mathbb{T}_n be a finite collection of transforms $\tau_n : \mathcal{X}_n \rightarrow \mathcal{X}_n$, and $t_n : \mathcal{X}_n \rightarrow \mathbf{R}^m$ be a Borel measurable map. Suppose that $T_{1n} \in \mathbb{T}_n$ and $T_{2n} \in \mathbb{T}_n$ are independent and identically distributed, and independent of X_n , where T_{1n} is drawn from the uniform distribution on \mathbb{T}_n .

The following theorem is theorem 3.2 of Hoeffding (1952) (adapted to notation here).

Theorem A1 (Hoeffding). *Suppose that F is the CDF of a random vector in \mathbf{R}^m and that for any continuity points $t_1, t_2 \in \mathbf{R}^m$ of F ,*

$$\left| P \left\{ \begin{bmatrix} t_n(T_{1n}X_n) \\ t_n(T_{2n}X_n) \end{bmatrix} \leq t \right\} - P \left\{ \begin{bmatrix} Q \\ Q' \end{bmatrix} \leq t \right\} \right| = o(1), \quad t = (t'_1, t'_2)',$$

as $n \rightarrow \infty$, where Q and Q' are independent and identically distributed as F .

Then for any $\eta > 0$ and for any continuity points $t_1, t_2 \in \mathbf{R}^m$ of F ,

$$P \{ |P \{ t_n(T_n X_n) \leq t | X_n \} - F(t) | > \eta \} \rightarrow 0, \quad t = (t'_1, t'_2)',$$

as $n \rightarrow \infty$.

Let $q_n : \mathcal{X}_n \rightarrow \mathbf{R}^m$ be a Borel measurable map and let $\lambda_n : \mathcal{X}_n \rightarrow \mathbf{R}^m$ be an m -dimensional positive-definite matrix-valued map. Define for $\tau_n \in \mathbb{T}_n$,

$$h_{\tau_n} = \text{Var}(q_n(T_n X_n) | T_n = \tau_n).$$

We introduce the following lemma:

Lemma 1. *Suppose that T_n is independent of X_n and is drawn from the uniform distribution on \mathbb{T}_n . Suppose further that for each $\varepsilon \in (0, 1)$, there exists a sequence of subsets $\mathbb{T}_n(\varepsilon) \subset \mathbb{T}_n$ and $\tilde{\mathbb{T}}_n(\varepsilon) \subset \mathbb{T}_n \times \mathbb{T}_n$ such that $|\mathbb{T}_n(\varepsilon)|/|\mathbb{T}_n| \rightarrow 1$ and $|\tilde{\mathbb{T}}_n(\varepsilon)|/|\mathbb{T}_n|^2 \rightarrow 1$ as $n \rightarrow \infty$, and a continuous, strictly increasing function $\varphi : [0, \infty) \rightarrow [0, \infty)$ (not depending on n) such that $\varphi(0) = 0$, where the sets $\mathbb{T}_n(\varepsilon)$ and $\tilde{\mathbb{T}}_n(\varepsilon)$ and the function φ satisfy the following conditions, C1 and C2:*

C1. For any fixed sequences $(\tau_{1n}, \tau_{2n}) \in \tilde{\mathbb{T}}_n(\varepsilon)$,

$$\sup_{t \in \mathbf{R}^m} \left| P \left\{ \begin{bmatrix} h_{\tau_{1n}}^{-1/2} q_n(\tau_{1n} X_n) \\ h_{\tau_{2n}}^{-1/2} q_n(\tau_{2n} X_n) \end{bmatrix} \leq t \right\} - \Phi_{2m}(t) \right| \leq \varphi(\varepsilon) + o(1),$$

as $n \rightarrow \infty$, where Φ_{2m} is the CDF of $N(0, I_{2m})$.

C2. For any fixed sequence $\tau_n \in \mathbb{T}_n(\varepsilon)$,

$$\|h_{\tau_n}^{-1/2} (\lambda_n^{-1/2}(\tau_n X_n) - h_{\tau_n}^{1/2})\| \leq \varphi(\varepsilon) + o_P(1), \quad (\text{A1})$$

as $n \rightarrow \infty$.

Then for any $\eta > 0$, as $n \rightarrow \infty$,

$$\sup_{t \in \mathbf{R}^m} P \{ |P \{ \lambda_n^{-1/2}(T_n X_n) q_n(T_n X_n) \leq t | X_n \} - \Phi_m(t) | > \eta \} \rightarrow 0, \quad (\text{A2})$$

where Φ_m is the CDF of $N(0, I_m)$.

Proof. Define $\Delta_n(\tau_n) = h_{\tau_n}^{-1/2} \{ \lambda_n^{-1/2}(\tau_n X_n) - h_{\tau_n}^{1/2} \}$ and write

$$\begin{aligned} & \{ \lambda_n^{-1/2}(\tau_n X_n) - h_{\tau_n}^{-1/2} \} q_n(\tau_n X_n) \\ &= \{ \lambda_n^{-1/2}(\tau_n X_n) - h_{\tau_n}^{-1/2} \} h_{\tau_n}^{1/2} h_{\tau_n}^{-1/2} q_n(\tau_n X_n). \end{aligned} \quad (\text{A3})$$

Define $\mathbf{1}_n(X_n) = \mathbf{1} \{ \|h_{\tau_n}^{-1/2} q_n(\tau_n X_n)\| \leq 1/\varphi^{1/4}(\varepsilon) \}$. Since

$$\{ \lambda_n^{-1/2}(\tau_n X_n) - h_{\tau_n}^{-1/2} \} h_{\tau_n}^{1/2} = -(\Delta_n(\tau_n) + I_{2m})^{-1} \Delta_n(\tau_n),$$

condition C2 implies that for any $\tau_n \in \mathbb{T}_n(\varepsilon)$,

$$\| \{ \lambda_n^{-1/2}(\tau_n X_n) - h_{\tau_n}^{-1/2} \} q_n(\tau_n X_n) \| \mathbf{1}_n(X_n) \leq \varphi^{3/4}(\varepsilon) + o_P(1),$$

as $n \rightarrow \infty$. Therefore, for any sequence of pairs (τ_{1n}, τ_{2n}) from $\tilde{\mathbb{T}}_n(\varepsilon) \equiv \mathbb{T}_n^2(\varepsilon) \cap \tilde{\mathbb{T}}_n(\varepsilon)$, and for any $t \in \mathbf{R}^{2m}$,

$$\begin{aligned} & P \{ Z_{2n}^*(\tau_{1n}, \tau_{2n}) \leq t - \varphi^{3/4}(\varepsilon) \mathbf{1}_{2m} \} - b_n(\varepsilon; \tau_{1n}) - b_n(\varepsilon; \tau_{2n}) \\ & \leq P \{ Z_{2n}(\tau_{1n}, \tau_{2n}) \leq t \} \\ & \leq P \{ Z_{2n}^*(\tau_{1n}, \tau_{2n}) \leq t + \varphi^{3/4}(\varepsilon) \mathbf{1}_{2m} \} + b_n(\varepsilon; \tau_{1n}) + b_n(\varepsilon; \tau_{2n}), \end{aligned}$$

where $\mathbf{1}_{2m}$ is the $2m$ -dimensional vector of ones, $b_n(\varepsilon; \tau_n) \equiv P \{ \|h_{\tau_n}^{-1/2} q_n(\tau_n X_n)\| > 1/\varphi^{1/4}(\varepsilon) \}$,

$$Z_{2n}^*(\tau_{1n}, \tau_{2n}) \equiv \begin{bmatrix} h_{\tau_{1n}}^{-1/2} q_n(\tau_{1n} X_n) \\ h_{\tau_{2n}}^{-1/2} q_n(\tau_{2n} X_n) \end{bmatrix}, \quad \text{and}$$

$$Z_{2n}(\tau_{1n}, \tau_{2n}) \equiv \begin{bmatrix} \lambda_{\tau_{1n}}^{-1/2}(\tau_{1n} X_n) q_n(\tau_{1n} X_n) \\ \lambda_{\tau_{2n}}^{-1/2}(\tau_{2n} X_n) q_n(\tau_{2n} X_n) \end{bmatrix}.$$

Since $\mathbf{E} \{ \|h_{\tau_n}^{-1/2} q_n(\tau_n X_n)\|^2 \} = m$, using Chebyshev's inequality, we deduce that for all $\tau_n \in \mathbb{T}_n(\varepsilon)$, $b_n(\varepsilon; \tau_n) \leq m\varphi^{1/2}(\varepsilon)$. Hence by condition C1,

$$\begin{aligned} & \Phi(t - \varphi^{3/4}(\varepsilon) \mathbf{1}_{2m}) - 2m\varphi^{1/2}(\varepsilon) - \varphi(\varepsilon) + o(1) \\ & \leq P \{ Z_{2n}(\tau_{1n}, \tau_{2n}) \leq t \} \\ & \leq \Phi(t + \varphi^{3/4}(\varepsilon) \mathbf{1}_{2m}) + 2m\varphi^{1/2}(\varepsilon) + \varphi(\varepsilon) + o(1), \end{aligned}$$

as $n \rightarrow \infty$, or, for some constant $C_2 > 0$,

$$|P \{ Z_{2n}(\tau_{1n}, \tau_{2n}) \leq t \} - \Phi_{2m}(t)| \leq C_2 \varphi^{1/2}(\varepsilon) + o(1), \quad (\text{A4})$$

as $n \rightarrow \infty$. (Here the terms $o(1)$ are uniform over $t \in \mathbf{R}^{2m}$.)

Now, note that by Jensen's inequality,

$$\begin{aligned} & |P \{ Z_{2n}(T_{1n}, T_{2n}) \leq t \} - \Phi_{2m}(t)| \\ & \leq \mathbf{E} |P \{ Z_{2n}(T_{1n}, T_{2n}) \leq t | T_{1n}, T_{2n} \} - \Phi_{2m}(t)|. \end{aligned} \quad (\text{A5})$$

Since T_{1n} and T_{2n} are drawn from the uniform distribution on \mathbb{T}_n independent of X_n , the last expectation is equal to

$$\begin{aligned} & \frac{1}{|\mathbb{T}_n|^2} \sum_{\tau_1, \tau_2 \in \mathbb{T}_n} |P \{ Z_{2n}(\tau_1, \tau_2) \leq t \} - \Phi_{2m}(t)| \\ & \leq \max_{\tau_1, \tau_2 \in \mathbb{T}_n(\varepsilon)} |P \{ Z_{2n}(\tau_1, \tau_2) \leq t \} - \Phi_{2m}(t)| + \frac{|\mathbb{T}_n|^2 - |\tilde{\mathbb{T}}_n(\varepsilon)|}{|\mathbb{T}_n|^2}. \end{aligned}$$

The last term is $o(1)$ as $n \rightarrow \infty$. Since $\bar{\mathbb{T}}_n(\varepsilon)$ is a finite set, there exists a sequence (τ_{1n}, τ_{2n}) that achieves the maximum above, and by equation (A4), the absolute value on the left-hand side of equation (A5) vanishes, as we send $n \rightarrow \infty$ and $\varepsilon \rightarrow 0$, fulfilling the condition of theorem A1. Hence equation (A2) follows.

The convergence in condition C1 usually follows because for each fixed permutation, the dependence structure and marginal heterogeneity are inherited under the relabeling of the observations. The main requirement in condition C1 is that the correlation between $q_n(\tau_{1n}X_n)$ and $q_n(\tau_{2n}X_n)$ is small for most τ_{1n} 's and τ_{2n} 's. As for Condition C2, the matrix h_{τ_n} is usually the conditional variance of $q_n(T_n X_n)$ given $T_n = \tau_n$ such that for each fixed $\tau_n \in \mathbb{T}_n$, $h_{\tau_n}^{-1/2} q_n(\tau_n X_n) \xrightarrow{d} N(0, I_m)$, and the matrix $\lambda_n(\tau_n X_n)$ represents its consistent estimator for each fixed $\tau_n \in \mathbb{T}_n$. Then lemma 1 says that the permutation distribution of $\lambda_n^{-1/2}(T_n X_n) q_n(T_n X_n)$ is asymptotically the same as $N(0, I_m)$.

3. A Synopsis of the Proof of Theorem 1

In this section, we provide a synopsis of the proof of theorem 1. The full proofs of the results are in the online supplemental note of this paper.

First, define

$$C_\pi(G_n) = \frac{1}{n} \sum_{i \in N_n} \mathbf{E}[e_{\pi(i)} a_{i,\pi} | \pi] - \frac{1}{n} \sum_{i \in N_n} \mathbf{E}[e_{\pi(i)} a_{i,\pi}^c | \pi],$$

where $a_{i,\pi} = \frac{1}{d_n(i)} \sum_{j \in N_n(i)} e_{\pi(j)}$ if $d_n(i) \geq 1$, and $a_{i,\pi} = 0$ otherwise, and

$$a_{i,\pi}^c = \frac{1}{|N_n \setminus \bar{N}_n(i)|} \sum_{j \in N_n \setminus \bar{N}_n(i)} e_{\pi(j)}.$$

Also let

$$\zeta_n = \frac{1}{\sqrt{n}} \sum_{i \in N_n} (q_i - \mathbf{E}[q_i]), \text{ and } \zeta_{n,\pi} = \frac{1}{\sqrt{n}} \sum_{i \in N_n} (q_{i,\pi} - \mathbf{E}[q_{i,\pi} | \pi]).$$

Then we can show the following,

$$\begin{aligned} \sqrt{n}\{\hat{C}(G_n) - C(G_n)\} &= \zeta_n + o_P(1), \\ \sqrt{n}\{\hat{C}_\pi(G_n) - C_\pi(G_n)\} &= \zeta_{n,\pi} + o_P(1), \end{aligned}$$

uniformly over $\pi \in \Pi_n$. Furthermore, after some algebra, we can show that

$$\hat{\sigma}_n^2 = \text{Var}(\zeta_n) + o_P(1).$$

Once asymptotic linear representation is obtained, we turn to asymptotic normality. It is not hard to see that q_i 's have as a dependency graph the graph $G_{n,2} = (N_n, E_{n,2})$, where $ij \in E_{n,2}$ if and only if i and j are within three edges from each other. Using the central limit theorem for a sum of random variables having a dependency graph (e.g., theorem 2.4 of Penrose, 2003), we can show that

$$T \xrightarrow{d} N(0, 1).$$

The convergence rate in the normal approximation is $O(d_{mx,n,3}/n^{1/4})$, which explains the rate requirement in equation (12).

From here on, we focus on T_π . For this, we use lemma 1. We take X_n and $q_n(X_n)$ in lemma 1 to be Y and $\zeta_{n,\pi}$, respectively, and take \mathbb{T}_n to be Π_n with the identification: $\tau_n X_n = (Y_{\pi(i)})_{i \in N_n}$. Also take $\lambda_n(X_n)$ and $h_{\tau_n}^2$ in lemma 1 to be $\hat{\sigma}_\pi^2$ and $h_{n,\pi}^2 \equiv \text{Var}(\zeta_{n,\pi} | \pi)$.

For asymptotic normality of $\zeta_{n,\pi}$, we first show that $h_{n,\pi}^2$ is asymptotically nondegenerate for most permutations π 's⁷, that is,

$$\frac{1}{|\Pi_n|} \sum_{\pi \in \Pi_n} 1 \left\{ h_{n,\pi}^2 > \frac{c}{2} \right\} \rightarrow 1, \text{ as } n \rightarrow \infty. \quad (\text{A6})$$

Then we apply the central limit theorem to $\zeta_{n,\pi}$. For this, note that $q_{i,\pi}$'s have as a dependency graph the graph $G_{n,3,\pi} = (N_n, E_{n,3,\pi})$, where $ij \in E_{n,3,\pi}$ if and only if $\pi(i)$ and $\pi(j)$ are within two edges from each other. Therefore we can apply the Berry-Esseen bound in theorem 2.4 of Penrose (2003) for each fixed sequence of permutations $\pi \in \Pi_n$. The bound is uniform over $\pi \in \Pi_n$, giving this convergence in distribution uniform over $\pi \in \Pi_n$. Note that $\text{Cov}(\zeta_{n,\pi_1}, \zeta_{n,\pi_2} | \pi_1, \pi_2)$ is small for most permutations if $\text{Cov}(q_{i,\pi_1}, q_{i,\pi_2} | \pi_1, \pi_2)$ is so. The latter asymptotic negligibility follows, because for most permutations π_1 and π_2 , q_{i,π_1} and q_{i,π_2} are independent.

Unlike T , the permutation test statistic T_π does not involve a centering by $C_\pi(G_n)$. (Compare equations [5] and [8].) Centering by $C_\pi(G_n)$ is not possible because it is unknown. However, the centering is not needed, because $C_\pi(G_n)$ is asymptotically negligible for most π 's, regardless of whether $C(G_n) = 0$. Indeed for any $\varepsilon > 0$,

$$\frac{1}{|\Pi_n|} \sum_{\pi \in \Pi_n} 1 \{ |\sqrt{n} C_\pi(G_n)| > \varepsilon \} = o(1), \text{ as } n \rightarrow \infty. \quad (\text{A7})$$

This result comes from the fact that for most permutations, the correlation between $e_{\pi(i)}$ and $a_{i,\pi}$ (or $a_{i,\pi}^c$) is 0 under the local dependence setup when $d_{mx,n}$ is small relative to n . Once we take $\Pi_n(\varepsilon)$ to be the set of permutations satisfying the inequalities in the indicators of equations (A6) and (A7), we obtain (C1) of lemma 1.

To show condition C2 of lemma 1, we need to deal with $\hat{\sigma}_\pi^2$. We show that $\hat{\sigma}_\pi^2$ is close to $h_{n,\pi}^2$ for most π 's in large samples. For this, we first define

$$q_{i,\pi}^* = \frac{1}{|S_n(i)|} \sum_{j \in S_n(i)} \mathbf{E}[q_{j,\pi} | \pi], \quad (\text{A8})$$

and let $\eta_{i,\pi} = q_{i,\pi} - \mathbf{E}[q_{i,\pi} | \pi]$ and $\tilde{\eta}_{i,\pi} = q_{i,\pi} - q_{i,\pi}^*$. Let $\sigma_{n,\pi}^2$ be the population version of $\hat{\sigma}_\pi^2$ (conditional on fixed π) defined by

$$\sigma_{n,\pi}^2 = \frac{1}{n} \sum_{i_1 \in N_n} \mathbf{E}[\tilde{\eta}_{i_1,\pi}^2 | \pi] + \frac{1}{n} \sum_{i_1 \in N_n} \sum_{i_2 \in N_{n,3}(i_1)} \mathbf{E}[\tilde{\eta}_{i_1,\pi} \tilde{\eta}_{i_2,\pi} | \pi].$$

Then after some algebra, we can show that for most permutations,

$$\hat{\sigma}_\pi^2 = \sigma_{n,\pi}^2 + o_P(1).$$

It remains to show that $\sigma_{n,\pi}^2$ is close to $h_{n,\pi}^2$. First, we write

$$\begin{aligned} h_{n,\pi}^2 - \tilde{h}_{n,\pi}^2 &= \frac{1}{n} \sum_{i_1 \in N_n} \sum_{i_2 \in N_n \setminus \{i_1\}} \mathbf{E}[\eta_{i_1,\pi} \eta_{i_2,\pi} | \pi], \\ \sigma_{n,\pi}^2 - \tilde{\sigma}_{n,\pi}^2 &= \frac{1}{n} \sum_{i_1 \in N_n} \sum_{i_2 \in N_{n,3}(i_1)} \mathbf{E}[\tilde{\eta}_{i_1,\pi} \tilde{\eta}_{i_2,\pi} | \pi], \end{aligned} \quad (\text{A9})$$

where $\tilde{h}_{n,\pi}^2 = \frac{1}{n} \sum_{i_1 \in N_n} \mathbf{E}[\eta_{i_1,\pi}^2 | \pi]$ and $\tilde{\sigma}_{n,\pi}^2 = \frac{1}{n} \sum_{i_1 \in N_n} \mathbf{E}[\tilde{\eta}_{i_1,\pi}^2 | \pi]$. Again the double sums on the right-hand side in equation (A9) are asymptotically negligible because $\eta_{i_1,\pi}$ and $\eta_{i_2,\pi}$ are uncorrelated for most permutations π . It remains to compare $\tilde{h}_{n,\pi}^2$ and $\tilde{\sigma}_{n,\pi}^2$. Observe that

$$\begin{aligned} \mathbf{E}[\eta_{i_1,\pi}^2 - \tilde{\eta}_{i_1,\pi}^2 | \pi] &= (q_{i_1,\pi}^* - \mathbf{E}[q_{i_1,\pi} | \pi]) \mathbf{E}[2q_{i_1,\pi} - \mathbf{E}[q_{i_1,\pi} | \pi] - q_{i_1,\pi}^* | \pi] \\ &= (q_{i_1,\pi}^* - \mathbf{E}[q_{i_1,\pi} | \pi]) (\mathbf{E}[q_{i_1,\pi} | \pi] - q_{i_1,\pi}^*) \\ &= -(q_{i_1,\pi}^* - \mathbf{E}[q_{i_1,\pi} | \pi])^2. \end{aligned} \quad (\text{A10})$$

However, for most π 's, $e_{\pi(i)}$ and $a_{i,\pi}$ are independent. Hence for such permutations, we have

$$\begin{aligned} \mathbf{E}[q_{i,\pi} | \pi] &= \mathbf{E}[e_{\pi(i)} | \pi] \mathbf{E}[a_{i,\pi} | \pi] \\ &= \mathbf{E}[e_{\pi(i)} | \pi] \frac{1}{d_n(i)} \sum_{j \in N_n(i)} \mathbf{E}[e_{\pi(j)} | \pi] = 0, \end{aligned}$$

because $\mathbf{E}[e_j] = 0$ for all $i \in N_n$, and hence $q_{i,\pi}^* = 0$ from equation (A8). Therefore, $\tilde{h}_{n,\pi}^2$ and $\tilde{\sigma}_{n,\pi}^2$ are close to each other for most permutations. Thus by lemma 1, we obtain that for all $t \in \mathbf{R}$,

$$P \{ \sqrt{n} \hat{C}_\pi / \hat{\sigma}_\pi \leq t | Y \} = \Phi(t) + o_P(1).$$

⁷When we say a statement holds "for most permutations" here, it means that the statement holds for all but an asymptotically negligible fraction of permutations.