Measuring Differences in Stochastic Network Structure

Eric Auerbach*

February 10, 2020

Abstract

How can one determine whether a community-level treatment, such as the introduction of a social program or trade shock, alters agents' incentives to form links in a network? This paper proposes analogues of a two-sample Kolmogorov-Smirnov test, widely used in the literature to test the null hypothesis of "no treatment effects," for network data. It first specifies a testing problem in which the null hypothesis is that two networks are drawn from the same random graph model. It then describes two randomization tests based on the magnitude of the difference between the networks' adjacency matrices as measured by the $2 \to 2$ and $\infty \to 1$ operator norms. Power properties of the tests are examined analytically, in simulation, and through two real-world applications. A key finding is that the test based on the $\infty \to 1$ norm can be substantially more powerful than that based on the $2 \to 2$ norm for the kinds of sparse and degree-heterogeneous networks common in economics.

1 Introduction

Researchers collect network data to learn about the agent interactions driving many economic phenomena. This is often done by first drawing a sample of agents and then surveying pairs of agents about potential relationships. For instance, to understand how information about new technologies and social programs diffuse within a village, researchers might survey villagers about sources of loans or advice (see for instance Conley and Udry 2010; Banerjee et al. 2013;

*Department of Economics, Northwestern University. E-mail: eric.auerbach@northwestern.edu. I thank Manuel Arellano, Jon Auerbach, Lori Beaman, Vincent Boucher, Yong Cai, Ivan Canay, Christina Chung, Ivan Fernandez-Val, Eric Gautier, Ben Golub, Joel Horowitz, Chuck Manski, Angelo Mele, Roger Moon, Aureo de Paula, Sida Peng, Stephen Nei, Mikkel Sølvsten, Alireza Tahbaz-Salehi, and Chris Udry for helpful suggestions.

Beaman et al. 2018). To learn about the social interactions driving scholastic achievement in a school, researchers might ask students about their friends or classmates (see for instance Bramoullé et al. 2009; Calvó-Armengol et al. 2009; Goldsmith-Pinkham and Imbens 2013).

When the data are collected via social survey, it is rare to think that these reported relationships record all of the relevant interactions between agents. Instead, the reported relationships are likely also influenced by completely unrelated features of the data collection process. For example, agents may overreport relationships corresponding to recent but otherwise irrelevant interactions, survey enumerators may have heterogeneous beliefs about what interactions constitute a particular relationship, or researchers may incorrectly match reported names to agents in the sample. The implication of this idiosyncratic variation is that were the researcher to return and resurvey the agents on a different day, or use a slightly different survey methodology, the reported relationships could change substantially even if the underlying interactions between agents that drive the phenomena of interest have not.

A common solution to this problem is to model the reported relationships as random variables. However, treating network data as stochastic complicates their analysis because it is not immediately clear which features of the data are informative about actual interactions between agents and which are merely a coincidence of idiosyncratic variation. For instance, to evaluate the impact of a social policy on agents' incentives to form economic relationships, researchers might survey agents before and after the policy is implemented and compare the observed configurations of relationships (see recently Comola and Prina 2015; Banerjee et al. 2016; 2017; Heß et al. 2018). But when the data are stochastic, a skeptical researcher may be concerned that all of the observed changes in the network can be explained by idiosyncratic features of the data collection process and are unrelated to the change in policy. That is, the observed differences between the networks are statistically insignificant.[1]

How can one determine whether the realized differences between two stochastic networks are statistically significant? To address the question, this paper considers a testing problem in which the null hypothesis is that two networks are drawn from the same random graph model. That is, the underlying propensity for agents to report links is the same even if the

[1]Significance tests are sometimes misused in the social sciences. For instance, they may be incorrectly interpreted as all-purpose indicators of scientific relevance rather than a statement about a specific statistical hypothesis (see Abadie 2018; Andrews and Kasy 2017; Ziliak and McCloskey 2008, for recent discussions).

configurations of relationships observed by the researcher are nominally different.

How should one measure differences between stochastic networks in practice? A common way to evaluate the hypothesis of "no treatment effects" when a treatment and control are each associated with a random vector of outcomes is to compare their empirical distribution functions (see generally Imbens and Rubin 2015; Abadie and Cattaneo 2018). A justification for this approach is that the empirical distribution function is sensitive to a large class of potential differences between random vectors. This paper proposes an analogous use of operator norms to measure differences between stochastic networks. A key contribution of this paper is to motivate and justify the proposal.

1.1 Outline

Section 2 describes a model and testing problem. The model is based on various dyadic fixed effects regression models popular in the econometrics literature (see recently Dzemski 2014; Candelaria 2016; Charbonneau 2017; Graham 2017; Jochmans 2017; Toth 2017; Verdier 2017; Jochmans and Weidner 2019; Gao 2019). The null hypothesis is that two networks are drawn from the same model. Potential applications include tests of the hypothesis of no treatment effects, tests of common assumptions in the network economics literature such as network stationarity, the non-existence of network externalities, and link reciprocity/undirectedness, detecting breaks in dynamic network structure, and assessing model goodness-of-fit.

Section 3 outlines a randomization test. The test is based on an implication of the model, that under the null hypothesis the joint distribution of links is invariant to exchanging the weight of a link in one network for its identically indexed counterpart in the other. It requires the researcher to choose a measure of discrepancy between the networks. That is, a test statistic. Any collection of network statistics can be used to construct the test statistic. Example test statistics include the mean absolute difference in the degree distributions, eigenvector centralities, or clustering coefficients of the two networks. Economic theory will sometimes determine exactly which test statistic should be used.

In many cases, however, the researcher might not know which test statistic to use. Limiting attention to an arbitrary statistic may cause the researcher to ignore potentially relevant information in the data contrary to the null hypothesis; considering too many test statistics may result in a multiple comparisons problem. To guard against these possibilities, Section 4 considers two test statistics that produce tests powerful against a large class of alternative hypotheses. The first test statistic is based on the $2 \to 2$ operator norm (also known as the spectral norm) of the entry-wise difference between the networks' adjacency matrices. The second test statistic is based on the $\infty \to 1$ operator norm (closely related to the cut norm of Frieze and Kannan 1999) of the entry-wise difference between the networks' adjacency matrices.[2] Intuitively, these norms are matrix analogues of the empirical distribution function statistics (for example, the Kolmogorov-Smirnov or Cramér-von Mises test statistics) commonly used to measure differences between random vectors, see Section 4.1.

Section 4.2 describes some power properties of the two tests using concentration results from the random matrix theory literature. It provides a class of sequences of alternative hypotheses (in which the dimensions of the adjacency matrices tend towards infinity) such that the tests are consistent: the probability of (correctly) rejecting the null hypothesis when it is false approaches one uniformly over this class. When the networks are weighted, the results do not require restrictions on the weights, which to my knowledge is new.

If the tests based on the $2 \to 2$ and $\infty \to 1$ norms had similar power properties, then the former might be preferred in practice because the $2 \to 2$ norm is straightforward to compute, while even approximating the $\infty \to 1$ norm up to a constant factor requires tools from the semidefinite programming literature. However, a key contribution of this paper is to demonstrate that the latter test can be considerably more powerful for alternatives in which there is nontrivial heterogeneity in the row-variances of the networks' adjacency matrices. Such row-heteroskedasticity may occur when the networks are sparse or have heavy-tailed degree distributions. Both features are defining characteristics of many social and economic networks (see generally Jackson 2008, Chapter 3).

Intuitively, the $2 \to 2$ norm may have low power under row-heteroskedasticity because the weight vector that maximizes this norm places most of its weight on the entries corresponding to the rows of the adjacency matrices with the highest variances. The test thus potentially ignores differences between the two networks that occur in the low-variance rows. The $\infty \to 1$

[2] For positive integers p and q, the $p \to q$ operator norm of a matrix is the largest product of the matrix and a unit weight vector. The magnitude of the product is measured using the q-vector norm. The magnitude of the weight vector is measured using the p-vector norm (see Aliprantis and Border 2006, Chapter 6).

norm addresses the problem by using the ∞-vector norm instead of the 2-vector norm to define the unit weight vector. The entries of the weight vector that maximize this norm take values in $\{-1, 1\}$ and thus by definition necessarily place the same absolute weight on every entry. Consequently, if there are sufficiently large differences between the two networks in the low-variance rows, the test based on the $\infty \to 1$ norm will detect them. In some sense, this logic behind the ∞-vector norm is related to that behind the 1 or 0-vector norm penalizations common in the high-dimensional regression literature. Instead of imposing a sparse solution, however, the ∞-vector norm imposes a dense one.

Related tests for network data have been proposed in the computer science and statistics literatures. Ghoshdastidar et al. (2017a;b) propose a test statistic based on the $2 \to 2$ norm for a different testing problem; they also rely on multiple copies of each network. A literature on random dot product graph models (see Tang et al. 2017; Nielsen and Witten 2018) relies on a particular low-dimensional dot-product structure that often fails to characterize social and economic networks. Another application of randomization-based inference includes tests for network interference (see recently Aronow 2012; Athey et al. 2018; Leung 2016; Song 2018; Puelz et al. 2019). Rather than study the influence of a treatment on network structure, this literature studies the influence of a network on agents' exposure to treatment.

Section 5 concludes with two empirical demonstrations. Proofs are in an appendix.

1.2 Main methodological contributions

I highlight three main methodological contributions to the network econometrics literature. The first contribution is the use of operator norms to measure differences in stochastic network structure nonparametrically. While the idea is based on an established mathematics literature (see generally Lovász 2012), to my knowledge this is the first application to network inference. I do not know of other results for arbitrarily weighted and directed networks. Analogous tests for random vectors (for example, the Kolmogorov-Smirnov test) are ubiquitous in the social sciences.

The second contribution is to identify the problem of row-heteroskedasticity. Heteroskedasitcity is a well-studied phenomenon in the random vector setting, but to my knowledge this is the first paper to demonstrate how it can cause inferences based on the spectra of random matrices to be unreliable. The spectrum of an adjacency matrix is routinely used in economic network theory (see for instance Golub and Jackson 2012; Bramoullé et al. 2014). Researchers are advised to be careful when interpreting the spectra of adjacency matrices when the networks are modeled as stochastic.

The third contribution is to propose the ∞-norm penalty as a robust alternative to the 2-norm when measuring differences in network structure using operator norms. It is well known in the penalized regression literature that choosing a penalty function with a "cusp" or "kink" at 0 can induce sparsity in the optimizing weight vector (see for instance Belloni et al. 2014). However, this is to my knowledge the first setting in which a penalty function with a "bump" is chosen specifically to induce density. The idea may also be useful for matrix estimation problems in which operator norms play a central role. Examples include clustering, topic modeling, network density estimation, and matrix completion. These problems are becoming increasingly common in economic research.

2 Framework

Sections 2.1 and 2.2 describes a model and testing problem. Section 2.3 discusses key assumptions. Section 2.3 provides examples of potential empirical applications.

2.1 Model

It is without loss to consider undirected unipartite networks. These networks are defined on a set of N agents referred to as a *community* and indexed by $[N] := \{1, 2, ..., N\}$. Every pair of agents in a community is endowed with two real-valued random variables, each corresponding to a stochastic social relationship. For example, one weight might correspond to whether two agents are friends, another might give the amount of trade between them, etc. The variable $D_{ij,t}$ for $t = 1, 2$ records the realized relationship t between agents i and j. The $N \times N$ dimensional symmetric adjacency matrix D_t contains $D_{ij,t}$ in the $ijth$ and jith entries.

Directed or bipartite networks are incorporated in the following way. These networks are generally defined on a set of N_1 agents and N_2 markets indexed by $[N_1]$ and $[N_2]$ respectively. Every agent-market pair is endowed with two real-valued random variables, each corresponding to a stochastic social relationship. For example, one weight might correspond to whether the agent is employed in the market, another might give the amount of profit the agent makes in the market, etc. The variable $D^{\star}_{ij,t}$ records the realized relationship t between agent i and market j. The $N_1 \times N_2$ dimensional matrix $D^{\star}_t$ contains $D^{\star}_{ij,t}$ in the $ijth$ entry. This asymmetric rectangular adjacency matrix is then transformed into a symmetric square one by setting

$$D_t = \begin{bmatrix} 0_{N_1 \times N_1} & D^{\star}_t \\ (D^{\star}_t)^T & 0_{N_2 \times N_2} \end{bmatrix}$$

where $(\cdot)^T$ is the transpose operator, $0_{N_1 \times N_1}$ is an $N_1 \times N_1$ matrix of zeros, and D_t is a $N \times N$ symmetric matrix with $N = N_1 + N_2$. Thus the focus on undirected unipartite networks (i.e. symmetric and square adjacency matrices) is without loss.

The entries of D_1 and D_2 are assumed be equal to 0 on the main diagonal (i.e, no self-links) and mutually independent above the main diagonal. This independence assumption is common in the network econometrics literature as discussed in more detail below. It can also be relaxed as discussed in Section 2.3.1. The marginal distribution of $D_{ij,t}$ is denoted by $F_{ij,t}$ and the $N \times N$ dimensional matrix F_t contains $F_{ij,t}$ in the ijth entry. A generic matrix of distribution functions F_t is referred to as a *random graph model*.

A concrete example of a random graph model is

$$D_{ij,t} = f_t(\alpha_{i,t}, \alpha_{j,t}, w_{ij,t}) + \varepsilon_{ij,t}$$

where $\alpha_{i,t}$ is an agent-specific effect, $w_{ij,t}$ are agent-pair attributes, f_t is a community link function, and $\varepsilon_{ij,t}$ is an error that is independently distributed across agent-pairs with marginal distribution $G_{ij,t}$ (see for example Dzemski 2014; Candelaria 2016; Charbonneau 2017; Graham 2017; Jochmans 2017; Toth 2017; Verdier 2017; Jochmans and Weidner 2019; Gao 2019). Importantly, the framework of this paper treats the effects $\{\alpha_{i,t}\}_{i\in[N],t\in[2]}$ and attributes $\{w_{ij,t}\}_{i,j\in[N],t\in[2]}$ as non-stochastic. That is, if these variables are thought to be drawn from some joint distribution, then the random graph model is defined conditional on their realization. This modeling strategy of conditioning on the fixed effects and attributes follows exactly the cited literature (see in particular Graham 2019, Section 6.3 for a discussion).

The only remaining source of randomness are the $\{\varepsilon_{ij,t}\}_{i\in[N],t\in[2]}$ and so the $\{D_{ij,t}\}_{i,j\in[N],t\in[2]}$ are independent random variables with marginals given by $F_{ij,t}(s) = G_{ij,t}(s - f_t(\alpha_{i,t}, \alpha_{j,t}, w_{ij,t}))$. The random graph model F_t is thus parametrized by the agent-specific effects, agent-pair attributes, community link function, and distribution of idiosyncratic errors. Informally, if a treatment alters any of these parameters, then the framework characterizes the change as a "treatment effect.[3]" This is formalized by the statement of the testing problem below. Other notions of a random graph model imply other definitions of a treatment effect, see Section 2.3, but their study is left to future work.

2.2 Testing problem

The problem considered in this paper is to test the null hypothesis

$$H_0 : F_{ij,1} = F_{ij,2} \text{ for every } i, j \in [N]$$

against the alternative

$$H_1 : F_{ij,1} \neq F_{ij,2} \text{ for some } i, j \in [N].$$

In words, H_0 is the hypothesis that D_1 and D_2 are drawn from the same random graph model.

This testing problem is nonstandard relative to the classic goodness-of-fit testing literature because $F_{ij,1}$ and $F_{ij,2}$ are allowed to vary arbitrarily across $i, j \in [N]$ under H_0. That is, the entries within an adjacency matrix are not assumed to be identically distributed. Link heterogeneity is thought to be an indispensable feature of stochastic networks in economics (see generally Jackson 2008, Chapter 4). Tests that fail to account for this heterogeneity (i.e. tests that treat D_1 and D_2 as vectors of length $N(N-1)/2$ with identically distributed entries) are generally underpowered. A real-world example of this is the first empirical

[3]Not every difference in model parameters can be detected using D_1 and D_2. The difference between the parameters must yield a sufficiently large difference in F_1 and F_2. See Section 4.2.

demonstration of Section 5.

For the concrete example of Section 2.1, the problem is equivalent to testing the hypothesis that $G_{ij,1}(s - f_1(\alpha_{i,1}, \alpha_{j,1}, w_{ij,1})) = G_{ij,2}(s - f_2(\alpha_{i,2}, \alpha_{j,2}, w_{ij,2}))$ for every $s \in \mathbb{R}$ and $i, j \in [N]$. The hypothesis may be false whenever the two random graph models have different agent-specific effects, agent-pair attributes, community link functions, or distribution of idiosyncratic errors. Distinguishing between these parameters is a different testing problem and generally requires more assumptions. For example, if the $\{\varepsilon_{ij,t}\}_{i \in [N_T], t \in [2]}$ are assumed to be identically distributed, $\alpha_{i,1} = \alpha_{i,2}$, and $w_{ij,1} = w_{ij,2}$ for every $i, j \in [N]$, then the problem reduces to a test of whether the link functions f_1 and f_2 are the same. Alternatively, if the $\{\varepsilon_{ij,t}\}_{i \in [N_T], t \in [2]}$ are assumed to be identically distributed, $f_1 = f_2$, and $w_{ij,1} = w_{ij,2}$ for every $i, j \in [N]$, then the problem reduces to a test of whether $\alpha_{i,1} = \alpha_{i,2}$ for every $i \in [N]$. See the discussion after Theorem 4 in Section 4.2.3 below for more details.

Thus the testing problem is specifically designed to detect differences between D_1 and D_2 that are not the result of the idiosyncratic errors $\{\varepsilon_{ij,t}\}_{i,j \in [N], t \in [2]}$. The logic behind this test is that any change in the distribution of idiosyncratic errors, agent fixed effects, linking function, etc. reflects different incentives for agents to form or report links. Differences due to the idiosyncratic errors may only reflect measurement or misreporting error, as motivated in the introduction. Of course, this may not be the only potentially interesting hypothesis about the random graph models of D_1 and D_2. Alternative testing problems, see for example Section 2.3, are left to future work.

2.3 Discussion of key assumptions

As discussed above, a potential drawback of the hypothesis testing framework is that the researcher must fix a hypothesis to test. Sections 2.1 and 2.2 define one notion of what it means for the structure of two stochastic networks to be the same. Different assumptions correspond to different notions of network similarity. I discuss three examples, but many others exist.

2.3.1 Example 1: Independence assumption

The assumption that the entries of D_1 and D_2 are mutually independent above the main diagonal may appear restrictive. However, it is consistent with the cited literature that conditions on any driver of link formation that is correlated across agent-pairs and treats it as a fixed parameter of the model (see Graham 2019, Section 6.3). The only remaining variation satisfies the independence assumption. Intuitively, this error may represent the types of survey or measurement error discussed in the introduction.

Alternatively, the entries of D_1 and D_2 may be dependent. For instance, $D_{ij,r}$ and $D_{kl.s}$ may be correlated if $\{i, j\}$ and $\{k, l\}$ share an index or are nearby in some latent space. This is a different testing problem. Depending on how the dependence is modeled, the randomization procedure of Section 3 may not be well-suited for this problem. A test might instead be constructed by choosing a test statistic that concentrates under the null hypothesis and using asymptotic approximations or large deviation inequalities to construct a critical value. This strategy can be applied to the operator norm statistics of Section 4 for a wide variety of dependence structures (see for example Vershynin 2010). Since the incorporation of dependence is unrelated to the main idea of this paper to measure network structure nonparametrically, I leave this extension to future work.

2.3.2 Example 2: Same communities assumption

The framework of this paper does not need the indices of D_1 and D_2 to refer to literally the exact same set of agents. Rather what is required is that the researcher can specify a correspondence from one community to the other such that the links between corresponding agents are comparable. For instance, if the researcher observes two networks connecting two distinct communities of students, the researcher might define two students to be comparable if they have the same GPA, extra-curricular activities, and class schedule. Let the vector $x_{i,t}$ characterize the GPA, extra-curricular activities, and class schedule of student i in community t, $[N_1]$ index the first community of students, and $[N_2]$ index the second community.

Then it is straightforward to extend the framework to test

$$H_0^b : F_{ij,1} = F_{kl,2} \text{ for every } i, j \in [N_1], k, l \in [N_2] \text{ such that } (x_{i,1}, x_{j,1}) = (x_{k,2}, x_{l,2})$$

against the alternative

$$H_1^b : F_{ij,1} \neq F_{kl,2} \text{ for some } i, j \in [N], k, l \in [N_2] \text{ such that } (x_{i,1}, x_{j,1}) = (x_{k,2}, x_{l,2}).$$

A potentially interesting complication arises when the agent characteristics used for the comparisons $\{x_{i,t}\}_{i\in[N_t],t\in[2]}$ are either unobserved or only partially observed. While the results of this paper can be extended by considering every possible comparison that is consistent with the observed data and combining the tests in the usual way, the resulting test may be underpowered in practice.

2.3.3 Example 3: Sharp null hypothesis

The problem considered in this paper is a test of the sharp null hypothesis that two random graph models agree for every agent-pair. Analogous testing problems are commonly specified in the treatment effects literature to detect any potential difference between treatment and control units (see generally Imbens and Rubin 2015; Abadie and Cattaneo 2018).

Sometimes the researcher is only interested in a smaller class of potential differences between F_1 and F_2. For instance, the researcher may want to know if two networks have the same distribution of average degrees or clustering coefficients. The randomization test of Section 3 may not be well-suited for these problems. However, a test of such average null hypotheses may often be constructed by first calculating sample analogues and then approximating their distribution in the usual way (see Graham 2019, Section 7). The focus of this paper on measuring differences in network structure nonparametrically using operator norms is not generally relevant for these testing problems, and so they are not considered in this paper.

2.4 Example applications

I sketch three potential applications of the framework. Other examples (not discussed here) include detecting breaks in dynamic network structure and assessing model goodness-of-fit.

2.4.1 Application 1: effect of an exogenous event

The researcher observes two networks and hypothesizes that their structure has changed in response to an exogenous event. For example, Goyal et al. (2006) collect information about coauthorships between economists and argue that the profession has become more interconnected in response to new research technologies such as the internet. Fowler (2006) studies bill cosponsorships in the US Senate and argues that the body has become more partisan after the election of the 104th congress. Banerjee et al. (2016) survey villagers before and after the introduction of a microfinance program and posit that the program has disincentivized certain types of economic relationships. A skeptical researcher may be concerned that while the observed networks are nominally different, they correspond to the same underlying patterns of interactions between agents as described by a random graph model. The first demonstration in Section 5 is based on this application.

2.4.2 Application 2: influence of research design

The researcher makes seemingly arbitrary decisions about data collection and processing that influences exactly which relationships between agents are observed. Here "seemingly arbitrary" refers to decisions that are not exactly determined by economic theory. For example, to understand how information diffuses in a village, it may not be clear whether the researcher should survey agents about their social connections, economic connections, previous conversations about similar information, potential future conversations about different information, etc. (see for instance Conley and Udry 2010; Banerjee et al. 2013; Beaman et al. 2018). To study social interactions amongst high school students, it is common to collect data by asking students to list their friends, but researchers often assume that any nomination of one agent by another indicates a social relationship between both agents (see for instance Bramoullé et al. 2009; Calvó-Armengol et al. 2009; Goldsmith-Pinkham and

Imbens 2013). Reciprocity of social relationships has important implications for the magnitude of peer effects, predictions about who is the "key player," and more (see Comola and Fafchamps 2014).[4] The second demonstration in Section 5 is based on this application.

2.4.3 Application 3: network externalities

The researcher posits that the observed network is generated by a model with externalities. Network externalities here refer to models in which the existence of a link between two agents depends on other links realized in the network (see for instance Sheng 2012; de Paula, Richards-Shubik, and Tamer 2014; Leung 2015; Menzel 2015; Ridder and Sheng 2015; de Paula 2016; Griffith 2016; Badev 2017; Boucher and Mourifié 2017; Mele 2017; Mele and Zhu 2017; Gualdani 2017; Leung 2019). The model also allows links to be explained by idiosyncratic variation, and the researcher would like to test the null hypothesis that the observed configuration of network links can be explained solely by this variation. That is, test for the existence of network externalities.

The test requires two networks. Two networks are necessary because there are no testable implications of network externalities with only a single network in this setting. One can potentially use networks defined on different communities as in Section 2.3, Example 2. I illustrate the application with the following model motivated by Bloch and Jackson (2007) (see Graham 2015, Section 2)

$$D_{ij,t} = \mathbb{1}\left\{\alpha_{ij} + \gamma_{ij} \sum_{k=1}^{N} D_{ik,t} D_{jk,t} - \varepsilon_{ij,t} \geq 0\right\}$$

where $\varepsilon_{ij,t}$ is independent, identically distributed, and mean-zero. The parameters α_{ij} and γ_{ij} do not vary with t. In this model, agents with many friends in common are more likely to become friends, and the agents first draw $\{\varepsilon_{ij,t}\}_{i \neq j}$ and then choose links so that the link formation rule is satisfied for every ij-pair.

The goal of the researcher is to test the null hypothesis

$$H_0^c : \gamma_{ij} = 0 \text{ for every } i, j \in [N]$$

[4]I thank Vincent Boucher for suggesting this example.

against the alternative

$$H_1^c : \gamma_{ij} \neq 0 \text{ for some } i, j \in [N].$$

Under H_0^c, D_1 and D_2 are drawn from the same random graph model (in the sense of Section 2.1), thus the randomization test proposed in Section 3 controls size in finite samples. I do not study the power properties of this test for network externalities in this paper.

3 Randomization procedure

I outline a randomization procedure to construct tests for the problem of Section 2 (see Lehmann and Romano 2006, Chapter 15). The procedure takes as given a test statistic $T(D_1, D_2)$. Any real-valued function of D_1 and D_2 can be used to construct the test statistic.

Economic theory sometimes suggests a particular test statistic. For example, in the network peer effects literature, differences in agent behavior are thought to be driven by a collection of average peer outcomes and characteristics. In the information diffusion literature, differences in the spread of information are thought to be driven by a collection of centrality statistics such as degree, eigenvector, or diffusion centrality. If $C(D_1)$ denotes a vector of agent-specific peer characteristics or centrality measures evaluated on D_1, then a potential test statistic is the root-mean-squared difference between $C(D_1)$ and $C(D_2)$. That is, $T(D_1, D_2) = ||C(D_1) - C(D_2)||_2$, where $|| \cdot ||_2$ is the vector 2-norm.

Once the researcher has selected a test statistic, the randomization procedure constructs a test for H_0 in the following way. For any positive integer R, let $\{\rho_{ij}^r\}_{i>j\in[N], r\in[R]}$ be a collection of $\binom{N}{2} \times R$ independent Bernoulli random variables with mean 1/2. Define $\rho_{ij}^r = \rho_{ji}^r$ if $i < j$. Then for each $r \in [R]$, the randomized $N \times N$ adjacency matrices D_1^r and D_2^r are generated by exchanging $D_{ij,1}$ and $D_{ij,2}$ whenever ρ_{ij}^r equals 1. That is,

$$D_{ij,1}^r = D_{ij,1}\rho_{ij}^r + D_{ij,2}(1 - \rho_{ij}^r)$$
$$D_{ij,2}^r = D_{ij,1}(1 - \rho_{ij}^r) + D_{ij,2}\rho_{ij}^r$$

where $D_{ij,1}^r$ is the ijth entry of D_1^r. For any $\alpha \in [0, 1]$, the proposed α-sized test based on

$T(D_1, D_2)$ rejects H_0 if

$$(R+1)^{-1}\left(1+\sum_{r\in[R]}\mathbb{1}\left\{T(D_1^r, D_2^r) \geq T(D_1, D_2)\right\}\right) \leq \alpha$$

and fails to reject H_0 otherwise.

Since (D_1, D_2) and (D_1^r, D_2^r) have the same distribution under H_0, Lehmann and Romano (2006), Theorem 15.2.1 implies the following. When H_0 is true the probability of a (incorrect) rejection does not exceed α, or

$$P\left((R+1)^{-1}\left(1+\sum_{r\in[R]}\mathbb{1}\left\{T(D_1^r, D_2^r) \geq T(D_1, D_2)\right\}\right) \leq \alpha | H_0 \text{ is true}\right) \leq \alpha.$$

This is true for any test statistic $T(D_1, D_2)$.[5] When the researcher has a collection of different network statistics, the researcher can combine them into one test statistic, or combine the tests in the usual way. One can also test whether any number of adjacency matrices are drawn from the same random graph model by permuting all of the corresponding entries.

4 Two test statistics based on operator norms

Two test statistics based on operator norms are proposed in Section 4.1. Some large sample power properties of the resulting tests are characterized in Section 4.2.

4.1 Specification

The randomization procedure of Section 3 produces a test for the problem of Section 2 that controls the probability of (incorrectly) rejecting the null hypothesis when it is true using any test statistic. However, not every statistic produces a test that is powerful in the sense that it tends to (correctly) reject the null hypothesis when it is false. This section proposes two statistics that have power against a large class of alternative hypotheses. The statistics are based on $p \to q$ operator norms.

[5] The probability that this test (incorrectly) rejects the null hypothesis when it is true is less than α. One can modify the test so that this probability is exactly α. See Lehmann and Romano (2006), Chapter 15.2.

For any real $p, q \geq 1$, the test statistic based on the $p \to q$ operator norm is given by

$$T_{p\to q}(D_1, D_2) = \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_p=1} || \left[\mathbb{1}\{D_1 \leq s\} - \mathbb{1}\{D_2 \leq s\}\right] \varphi||_q$$

where $||\cdot||_p$ refers to the vector p-norm, φ is a N-dimensional column vector with real-valued entries, and for any real number s and matrix X, $\mathbb{1}\{X \leq s\}$ contains 1 in the ijth entry if $X_{ij} \leq s$ and 0 otherwise. The test of H_0 based on $T_{p\to q}$ is exactly as described in Section 3.

When the entries of D_1 and D_2 are $\{0, 1\}$-valued,

$$T_{p\to q}(D_1, D_2) = \max_{\varphi:||\varphi||_p=1} || (D_1 - D_2)\, \varphi||_q$$

which is the $p \to q$ operator norm of the entry-wise difference between two adjacency matrices. Intuitively, $T_{p\to q}(D_1, D_2)$ compares the collection of weighted degree distributions of D_1 and D_2, indexed by the weight vector φ and given by $\{\sum_{j\in[N]} D_{ij,t}\varphi_j\}_{i\in[N],\varphi:||\varphi||_p=1}$. This weighted degree distribution function can be viewed as a matrix analogue of the empirical distribution function for vectors. Instead of measuring the number of entries that fall below a point on the real line, however, it measures the magnitude of connections from each agent to the set of other agents (i.e. degree), as weighted by φ.

Not every $p \to q$ operator norm is either computable or produces a test that has power against a nontrivial class of alternatives in H_1. I thus focus on two particular choices of p and q.

The first test statistic is based on the $2 \to 2$ operator norm

$$T_{2\to 2}(D_1, D_2) = \max_{s\in\mathbb{R}} \max_{\varphi:\sum_t \varphi_t^2=1} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left[\mathbb{1}\{D_{ij,1} \leq s\} - \mathbb{1}\{D_{ij,2} \leq s\}\right] \varphi_j\right)^2}.$$

It can be computed by first fixing $s \in \mathbb{R}$ and then computing the largest spectral value of $[\mathbb{1}\{D_1 \leq s\} - \mathbb{1}\{D_2 \leq s\}]$. The outer maximization is then taken over the number of unique entries in D_1 and D_2 which is not larger than $N(N-1)$.

The $2 \to 2$ norm is a natural first choice because it is straightforward to compute and its statistical properties have been well studied in the random matrix theory literature (see generally Tao 2012). However, as I demonstrate below, the resulting test may have low power under row-heteroskedasticity: nontrivial variation in the row-variances of the adjacency matrices. Intuitively, the problem is that the weight vector φ that maximizes the program places excessive weight on the high-variance rows of $[\mathbb{1}\{D_1 \leq s\} - \mathbb{1}\{D_2 \leq s\}]$. See Section 4.2 for details. To address this problem, I consider a second test statistic.

The second test statistic is based on the $\infty \to 1$ operator norm

$$T_{\infty\to 1}(D_1, D_2) = \max_{s\in\mathbb{R}} \max_{\varphi:\max_{t\in[N]}|\varphi_t|=1} \sum_{i\in[N]} \left| \sum_{j\in[N]} [\mathbb{1}\{D_{ij,1} \leq s\} - \mathbb{1}\{D_{ij,2} \leq s\}]\, \varphi_j \right|.$$

The logic behind this test statistic is that the weight vector φ that maximizes this problem necessarily places the same absolute weight on every entry and is thus less sensitive to row-heteroskedasticity. Unfortunately, this norm cannot be computed to arbitrary precision except in trivial cases (see Håstad 2001). The proposed test is instead based on the semidefinite approximation

$$S_{\infty\to 1}(D_1, D_2) = \frac{1}{2} \max_{s\in\mathbb{R}} \max_{X\in\mathcal{X}_{2N}} \left\langle \begin{bmatrix} 0_{N\times N} & \Delta(s) \\ \Delta(s) & 0_{N\times N} \end{bmatrix}, X \right\rangle$$

where $\Delta(s) = \mathbb{1}\{D_1 \leq s\} - \mathbb{1}\{D_2 \leq s\}$, $\langle\cdot\rangle$ is the inner product operator (i.e. $\langle X, Y\rangle = \sum_{i=1}^{2N}\sum_{j=1}^{2N} X_{ij}Y_{ij}$), and $\mathcal{X}_{2N}$ is the set of all $2N \times 2N$ positive semidefinite matrices with diagonal entries equal to 1, see Goemans and Williamson (1994); Nesterov (1998). This statistic can be computed to arbitrary precision relatively quickly using programs available in many statistical software packages. [6]

Remarkably, $T_{\infty\to 1}$ and $S_{\infty\to 1}$ are equivalent up to a factor of 2 regardless of N. It is

[6] The results below use the sqlp function in the SDPT3 package for R (see Toh et al. 2012).

relatively straightforward to show that $T_{\infty\to 1} \leq S_{\infty\to 1}$ since

$$\begin{aligned}
T_{\infty\to 1} &= \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_\infty=1} ||\Delta(s)\varphi||_1 \\
&= \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_\infty=1} \max_{\psi:||\psi||_\infty=1} \langle \Delta(s), \varphi\otimes\psi\rangle \\
&= \frac{1}{2} \max_{s\in\mathbb{R}} \max_{\phi:||\phi||_\infty=1} \left\langle \begin{bmatrix} 0_{N\times N} & \Delta(s) \\ \Delta(s) & 0_{N\times N} \end{bmatrix}, \phi\otimes\phi \right\rangle \\
&\leq \frac{1}{2} \max_{s\in\mathbb{R}} \max_{X\in\mathcal{X}_{2N}} \left\langle \begin{bmatrix} 0_{N\times N} & \Delta(s) \\ \Delta(s) & 0_{N\times N} \end{bmatrix}, X \right\rangle = S_{\infty\to 1}
\end{aligned}$$

where $\otimes$ refers to the vector outer product operator, the first equality follows from taking $\psi = \operatorname{sign}(\Delta(s)\varphi)$, the second equality follows by choosing ϕ to be the concatenation of φ and ψ, and the inequality follows from $\phi\otimes\phi \in \mathcal{X}_{2N}$.

Grothendieck's inequality (see Krivine 1979; Alon and Naor 2006, as well as Appendix A.2) directly implies a corresponding lower bound

$$S_{\infty\to 1} \leq \frac{\pi}{2\ln\left(1+\sqrt{2}\right)} T_{\infty\to 1} \leq 1.783\, T_{\infty\to 1}.$$

These bounds are not tight, but currently the best available. They motivate $S_{\infty\to 1}$ as a computable alternative to $T_{\infty\to 1}$ in the proposed test of H_0.

To summarize, the α-sized test based on the $2 \to 2$ norm rejects H_0 if

$$(R+1)^{-1}\left(1+\sum_{r\in[R]} \mathbb{1}\left\{T_{2\to 2}(D_1^r, D_2^r) \geq T_{2\to 2}(D_1, D_2)\right\}\right) \leq \alpha.$$

The α-sized test based on the $\infty \to 1$ norm rejects H_0 if

$$(R+1)^{-1}\left(1+\sum_{r\in[R]} \mathbb{1}\left\{S_{\infty\to 1}(D_1^r, D_2^r) \geq S_{\infty\to 1}(D_1, D_2)\right\}\right) \leq \alpha.$$

4.2 Some large sample power properties

As discussed in Section 3, the α-sized tests based on $T_{2\to 2}$ and $S_{\infty\to 1}$ (incorrectly) reject the null hypothesis when it is true with probability less than α. This section provides conditions such that the tests (correctly) reject the null hypothesis when it is false. Specifically, it defines a class of sequences of alternative hypotheses such that the power of the tests tend to one uniformly over the class. Each sequence describes a collection of models and tests indexed by $N \in \mathbb{N}$. Each model is as described in Section 2. Each test is as described in Section 3. The parameters F_1, F_2, R, and α may all vary with N (subject to restrictions outlined below). Limits are with $N \to \infty$. The results mirror those for the Kolmogorov-Smirnov test provided in Chapter 14.2 of Lehmann and Romano (2006).

4.2.1 Assumptions and constructions

The assumptions of Section 2 are collected in the following statement.

Assumption 1 (Framework): The adjacency matrices D_1 and D_2 are $N \times N$ symmetric random matrices with independent upper diagonal entries and zeros on the diagonal. The null hypothesis to be tested is

$$H_0 : F_{ij,1} = F_{ij,2} \text{ for every } i, j \in [N]$$

where $F_{ij,t}(s) = P(D_{ij,t} \leq s)$ against the alternative

$$H_1 : F_{ij,1} \neq F_{ij,2} \text{ for some } i, j \in [N]. \quad \square$$

The following restriction on the test parameters is imposed.

Assumption 2 (Parameters):

$$\frac{\ln(R) - \ln(\alpha)}{\ln(N)} = O\left(1\right) \text{ and } \alpha(R+1) \geq 1. \quad \square$$

Assumption 2 nests three basic assumptions. The first is that the number of simulations used to construct the randomization distribution is not exponentially large relative to the number of agents. The second is that the size of the test is not exponentially small relative to the number of agents. The third assumption is that the size of the test is larger than the inverse of the number of simulations. I do not believe these assumptions to be restrictive in practice. The first two assumptions are used to bound the reference distributions generated by $\{T_{2\to 2}(D_1^r, D_2^r)\}_{r\in[R]}$ and $\{S_{\infty\to 1}(D_1^r, D_2^r)\}_{r\in[R]}$. See Section 4.2.4. The third assumption is required because $(R+1)^{-1}\left(1+\sum_{r\in[R]} \mathbb{1}\{T(D_1^r, D_2^r) \geq T(D_1, D_2)\}\right) \geq (R+1)^{-1}$ by construction. Thus if $(R+1)^{-1} > \alpha$, the test will mechanically fail to reject H_0 regardless of the choice of T or any difference between F_1 and F_2.

The following constructions are used.

Constructions:

$$\Delta(s) = \mathbb{1}\{D_1 \leq s\} - \mathbb{1}\{D_2 \leq s\}$$

$$\nu_{ij}(s) = F_{ij,1}(s) + F_{ij,2}(s) - 2F_{ij,1}(s)F_{ij,2}(s)$$

$$\tau = \max_{s\in\mathbb{R}} \max_{i\in[N]} \sqrt{\sum_{j\in[N]} \nu_{ij}(s)}$$

$$\sigma = \max_{s\in\mathbb{R}} \sum_{i\in[N]} \sqrt{\sum_{j\in[N]} \nu_{ij}(s)}$$

$$T_{2\to 2}(F_1, F_2) = \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_2=1} ||\left(F_1(s) - F_2(s)\right)\varphi||_2$$

$$T_{\infty\to 1}(F_1, F_2) = \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_\infty=1} ||\left(F_1(s) - F_2(s)\right)\varphi||_1. \quad \square$$

In words, $\nu_{ij}(s)$ is the variance of $\Delta_{ij}(s)$, $\sqrt{\sum_{j\in[N]} \nu_{ij}(s)}$ is the root of the ith row-variance of $\Delta(s)$, $\max_{i\in[N]} \sqrt{\sum_{j\in[N]} \nu_{ij}(s)}$ and $\sum_{i\in[N]} \sqrt{\sum_{j\in[N]} \nu_{ij}(s)}$ are the maximum and average root-row-variance of $\Delta(s)$ respectively, and τ and σ are the maximum of the maximum and average root-row-variances taken over s. Under H_0 and certain regularity conditions, $T_{2\to 2}(D_1, D_2)$ is proportional to τ and $T_{\infty\to 1}(D_1, D_2)$ is proportional to σ with high probability. This is demonstrated by Lemmas 1 and 2 in Section 4.2.4 below.

$T_{2\to 2}(F_1, F_2)$ and $T_{\infty\to 1}(F_1, F_2)$ are the two test statistics applied to the (matrix of) distribution functions F_1 and F_2. These metrics quantify the extent to which H_0 is violated.

Larger values correspond to more extreme violations. Under certain regularity conditions, $T_{2\to 2}(F_1, F_2)$ large relative to τ or $T_{\infty\to 1}(F_1, F_2)$ large relative to σ eventually results in a rejection of H_0. This is the content of Theorems 1 and 2 below. The results are analogous to Theorem 14.2.2 of Lehmann and Romano (2006) which provides sufficient conditions for the Kolmogorv-Smirnov test to be consistent.

4.2.2 Consistency

Assumptions 1 and 2 are assumed throughout. The consistency result for the test based on the $2 \to 2$ norm is given by Theorem 1.

Theorem 1: The power of the α-sized test that rejects H_0 whenever

$$(R+1)^{-1}\left(1 + \sum_{r\in[R]} \mathbb{1}\{T_{2\to 2}(D_1^r, D_2^r) \geq T_{2\to 2}(D_1, D_2)\}\right) \leq \alpha$$

tends to one uniformly over all alternatives that satisfy $T_{2\to 2}(F_1, F_2)/\tau \to \infty$ and $\tau/\sqrt{\ln(N)} \to \infty$. That is,

$$\inf_{\substack{F_1,F_2: T_{2\to 2}(F_1,F_2)/\tau\to\infty \\ \text{and } \tau/\sqrt{\ln(N)}\to\infty}} P\left((R+1)^{-1}\left(1 + \sum_{r\in[R]} \mathbb{1}\{T_{2\to 2}(D_1^r, D_2^r) \geq T_{2\to 2}(D_1, D_2)\}\right) \leq \alpha\right) \to 1. \quad \square$$

The hypothesis of Theorem 1 has two rate conditions. The first rate condition is that $T_{2\to 2}(F_1, F_2)/\tau \to \infty$. This condition implies that the size of the violation of H_0 (as given by $T_{2\to 2}(F_1, F_2)$) exceeds the magnitude of the test statistic $T_{2\to 2}(D_1, D_2)$ under H_0 (which is on the order of τ) with high probability. When F_t is sufficiently dense in the sense that for some $s \in \mathbb{R}$, $F_{ij,t}(s)$ is uniformly bounded away from 0 and 1, one can show that it follows from $\sqrt{N}T_{2\to 2}(F_1, F_2) \to \infty$. An analogous rate condition appears in Theorem 14.2.2 of Lehmann and Romano (2006).

The second rate condition is that $\tau/\sqrt{\ln(N)} \to \infty$. This condition implies that the reference distribution generated by $\{T_{2\to 2}(D_1^r, D_2^r)\}_{r\in[R]}$ concentrates below τ. It is satisfied if F_t is sufficiently dense in the sense that, for some $s \in \mathbb{R}$, $\frac{N}{\ln(N)}F_{ij,t}(s)$ and $\frac{N}{\ln(N)}(1 - F_{ij,t}(s))$

are uniformly bounded away from 0. In other words, agents have on expectation at least $\ln(N)$ connections. This suggests that the test based on the $2 \to 2$ norm may be poorly suited for sparse networks in which agents have on expectation only a bounded number of connections. See Section 4.3 below.

The consistency result for the test based on the $\infty \to 1$ norm is given by Theorem 2.

Theorem 2: The power of the α-sized test that rejects H_0 whenever

$$(R+1)^{-1}\left(1+\sum_{r\in[R]}\mathbb{1}\{S_{\infty\to1}(D_1^r,D_2^r)\geq S_{\infty\to1}(D_1,D_2)\}\right)\leq\alpha$$

tends to one uniformly over all alternatives that satisfy $T_{\infty\to1}(F_1,F_2)/\sigma\to\infty$ and $\sigma/\sqrt{\ln(N)}\to\infty$. That is,

$$\inf_{\substack{F_1,F_2:T_{\infty\to1}(F_1,F_2)/\sigma\to\infty\\ \text{and } \sigma/\sqrt{\ln(N)}\to\infty}} P\left((R+1)^{-1}\left(1+\sum_{r\in[R]}\mathbb{1}\{S_{\infty\to1}(D_1^r,D_2^r)\geq S_{\infty\to1}(D_1,D_2)\}\right)\leq\alpha\right)\to1. \quad \square$$

The two rate conditions in the hypothesis of Theorem 2 are similar to those in the hypothesis of Theorem 1. The first is that $T_{\infty\to1}(F_1,F_2)/\sigma\to\infty$ (or equivalently, that $S_{\infty\to1}(F_1,F_2)/\sigma\to\infty$). This condition implies that the size of the violation of H_0 (as given by $T_{\infty\to1}(F_1,F_2)$) exceeds the magnitude of the test statistic under H_0 (which is on the order of σ) with high probability. When F_t is dense in the sense that for some $s\in\mathbb{R}$, $F_{ij,1}(s)$ is uniformly bounded away from 0 and 1, it also follows from $\sqrt{N}T_{\infty\to1}(F_1,F_2)\to\infty$.

The second rate condition is that $\sigma/\sqrt{\ln(N)}\to\infty$. This condition implies that that the reference distribution generated by $\{S_{\infty\to1}(D_1^r,D_2^r)\}_{r\in[R]}$ concentrates below σ. It is satisfied if for some $s\in\mathbb{R}$, $\frac{N^3}{\ln(N)}F_{ij,t}(s)$ and $\frac{N^3}{\ln(N)}(1-F_{ij,t}(s))$ are uniformly bounded away from 0. In contrast to the hypothesis of Theorem 1, it is sufficient that agents have on expectation at least $\ln(N)/N^3$ connections, which covers settings in which agents have a bounded number of connections. In contrast to the second rate condition of Theorem 1, this second rate condition is unlikely to be restrictive in practice.

Theorems 1 and 2 predict two scenarios in which the test based on the $\infty\to1$ norm is potentially more powerful than that based on the $2 \to 2$ norm, in the sense that the hypothesis of Theorem 2 is satisfied but that of Theorem 1 is not. The first scenario is under network sparsity. An example of this is when $F_1 \wedge (1-F_1)$ and $F_2 \wedge (1-F_2)$ are uniformly on the order of $1/N$. One can verify in this case that $\tau/\sqrt{\ln(N)} \to 0$ but $\sigma/\sqrt{\ln(N)} \to \infty$. The second scenario is under degree-heterogeneity. An example of this is when, for some small positive integer K, $F_{ij,1} = F_{ij,2}$ is on the order of a constant if $i \wedge j \leq K$ but $F_{ij,1} \neq F_{ij,2}$ is on the order of $1/\sqrt{N}$ when $i \wedge j > K$. Intuitively, the first K agents have an order of magnitude more links than the other $N-K$ agents. One can verify in this case that $T_{2\to 2}(F_1, F_2)/\tau \to 0$ but $T_{\infty\to 1}(F_1, F_2)/\sigma \to \infty$.[7] Simulation evidence supporting these predictions is provided in Section 4.3.

4.2.3 Additional results

I supplement Theorems 1 and 2 with two additional results. The first result is that the rate conditions $T_{2\to 2}(F_1, F_2)/\tau \to \infty$ and $T_{\infty\to 1}(F_1, F_2)/\sigma \to \infty$ are close to necessary. This is made precise in the statement of Theorem 3 and is analogous to Theorem 14.2.3 of Lehmann and Romano (2006), which demonstrates a similar result for the Kolmogorov-Smirnov test. The second result is a weaker version of the consistency results of Theorems 1 and 2, given by Theorem 4. The hypothesis of this theorem does not specify rate conditions that depend on operator norms, and may be easier to interpret and apply in practice. I demonstrate its use with the concrete example of Section 2.

Theorem 3: For any sequence of positive real numbers $\delta_N \to \infty$

$$\inf_{\substack{F_1, F_2 : \delta_N [T_{2\to 2}(F_1,F_2)/\tau] \to \infty \\ \text{and } \tau/\sqrt{\ln(N)} \to \infty}} P\left((R+1)^{-1} \left(1 + \sum_{r \in [R]} \mathbb{1}\{T_{2\to 2}(D_1^r, D_2^r) \geq T_{2\to 2}(D_1, D_2)\} \right) \leq \alpha \right) \to \alpha$$

[7] Alternatively, the test based on the $2 \to 2$ norm is potentially more powerful under degree-heterogeneity when the differences between F_1 and F_2 occur in a small number of high-variance rows. Even in this case, making inferences based on a small number of high-variance observations (i.e. "outliers") is not recommended.

and

$$\inf_{\substack{F_1,F_2:\delta_N[T_{\infty\to 1}(F_1,F_2)/\sigma]\to\infty \\ \text{and } \sigma/\sqrt{\ln(N)}\to\infty}} P\left((R+1)^{-1}\left(1+\sum_{r\in[R]}\mathbb{1}\{S_{\infty\to 1}(D_1^r,D_2^r)\geq S_{\infty\to 1}(D_1,D_2)\}\right)\leq\alpha\right)\to\alpha. \quad \square$$

The statement of Theorem 3 differs from that of Theorems 1 and 2 in two ways. The first is that the rate conditions in the subscript under the infima have been changed from $[T_{2\to 2}(F_1,F_2)/\tau]\to\infty$ and $[T_{\infty\to 1}(F_1,F_2)/\sigma]\to\infty$ to $\delta_N\,[T_{2\to 2}(F_1,F_2)/\tau]\to\infty$ and $\delta_N\,[T_{\infty\to 1}(F_1,F_2)/\sigma]\to\infty$ respectively. That is, the infima are taken over a (slightly) larger class of sequences of in H_1. The second difference is the conclusion: the power of the tests no longer tend to one. In fact, the limiting power of the tests may be no greater than α. To prove the theorem, the main work is in constructing a sequence of alternatives such that $T_{2\to 2}(F_1,F_2)/\tau\to 0$ slower than any given sequence, and $T_{2\to 2}(D_1,D_2)$ and $T_{2\to 2}(D_1^r,D_2^r)$ converge to the same nondegenerate distribution. Intuitively, Theorem 3 states that the tests proposed in Theorems 1 and 2 cannot detect differences between random graph models that are too similar, in the sense that $T_{2\to 2}(F_1,F_2)$ or $T_{\infty\to 1}(F_1,F_2)$ are too close to 0.

The second result is that under certain conditions the power of the tests from Theorems 1 and 2 tend to one whenever the difference between F_1 and F_2 is "regular" in the sense that there exist two nontrivially sized subcommunities $I_N,J_N\subseteq[N]$ such that the probabilities that any agent in I_N links to any agent in J_N all either increase or decrease with t.

Theorem 4: Suppose there exists $I_N,J_N\subseteq[N]$ with $\liminf_{N\to\infty}\frac{|I_N|\wedge|J_N|}{N}>0$, $s\in\mathbb{R}$, and $\rho_N>0$ such that for all $i\in I_N$ and $j\in J_N$ either $(F_{ij,1}(s)-F_{ij,2}(s))>\rho_N$ or $(F_{ij,1}(s)-F_{ij,2}(s))<\rho_N$. Then the power of the test from Theorem 1 converges to one if $\rho_N N/\ln(N)\to\infty$. The power of the test from Theorem 2 converges to one if $\rho_N N\to\infty$. $\square$

The benefit of Theorem 4 relative to Theorems 1 and 2 is that its hypothesis does not use rate conditions that depend on operator norms. To illustrate its use, I sketch two simple testing problems with models based on the concrete example from Section 2. Recall for this example that $F_{ij,t}(s)=G_{ij,t}(s-f_t(\alpha_{i,t},\alpha_{j,t},w_{ij,t}))$. Suppose for example that the idiosyncratic errors are identically distributed for all agent-pairs and networks, the agent-pair attributes are the same for the two networks, and the community link function has the form $f(\alpha_{i,t}, \alpha_{j,t}, w_{ij}) = \Lambda(\alpha_{i,t} + \alpha_{j,t} + w_{ij}\beta)$ for some unknown vector β and strictly monotonic function Λ. Then the hypothesis of Theorem 4 is satisfied if there exists an I_N with $\liminf_{N\to\infty} I_N/N > 0$ such that $|\alpha_{i,1} - \alpha_{i,2}| > \rho_N$ for all $i \in I_N$. That is, under these conditions, the tests of Theorems 1 and 2 eventually (correctly) reject the null hypothesis that the two networks have the same collection of agent-specific effects when it is false.

Alternatively, suppose that the idiosyncratic errors are identically distributed, the agent-specific effects and agent-pair attributes are the same for the two networks, and the community link function has the form $f_t(\alpha_i, \alpha_j, w_{ij}) = \Lambda_t(\alpha_i, \alpha_j) + w_{ij}\beta$ for some functions $\{\Lambda_t\}_{t\in[2]}$ and vector β. Then the hypothesis of Theorem 4 is satisfied if $\Lambda_1(\alpha_i, \alpha_j)$ and $\Lambda_2(\alpha_i, \alpha_j)$ disagree on $I_N \times J_N$ with $\liminf_{N\to\infty}(I_N \wedge J_N)/N > 0$. That is, under these conditions, the tests from Theorems 1 and 2 eventually (correctly) reject the null hypothesis that the two networks have the same community link function when it is false.

4.2.4 Two key lemmas

Proofs of Theorems 1-4 can be found in the appendix. They rely on bounds for the reference distributions generated by $\{T_{2\to2}(D_1^r, D_2^r)\}_{r\in[R]}$ and $\{S_{\infty\to1}(D_1^r, D_2^r)\}_{r\in[R]}$, although the lemmas in the appendix are stated and proved for an arbitrary collection of matrices with independent, bounded, and mean-zero entries.

Lemma 1: For any fixed $\alpha \in [0,1]$ with probability at least $1-\alpha$

$$\hat{\tau} \leq \max_{r\in[R]} T_{2\to2}(D_1^r, D_2^r) \leq C_{2\to2}\tau + O\left(\sqrt{\ln(N)} + \sqrt{\ln(R/\alpha)}\right)$$

where $\hat{\tau} = \max_{s\in\mathbb{R}} \max_{i\in[N]} \sqrt{\sum_{j\in[N]} \Delta_{ij}(s)^2}$. $\square$

Lemma 2: For any fixed $\alpha \in [0,1]$ with probability at least $1-\alpha$

$$\hat{\sigma} \leq \max_{r\in[R]} S_{\infty\to1}(D_1^r, D_2^r) \leq C_{\infty\to1}\sigma + O\left(\sqrt{\ln(N)} + \sqrt{\ln(R/\alpha)}\right)$$

where $\hat{\sigma} = \max_{s\in\mathbb{R}} \sum_{i\in[N]} \sqrt{\sum_{j\in[N]} \Delta_{ij}(s)^2}$. $\square$

Specifically, the lemmas hold for $C_{2\to2} > 2$ and $C_{\infty\to1} > \frac{\pi}{\ln(1+\sqrt{2})} > 3.57$. They demonstrate that the reference distributions for $T_{2\to2}$ and $S_{\infty\to1}$ generated by randomly exchanging the corresponding entries of the two adjacency matrices are highly likely to be on the order of magnitude of τ and σ no matter the underlying F_1 and F_2. The upper bounds follow from the general recipe for bounding operator norms of random matrices specified in Chapter 2.3 of Tao (2012). The first step is to apply a variant of Talagrand's inequality (see Boucheron, Lugosi, and Massart 2013, Corollary 6.10) to bound the variation of these statistics around their expectation. The second step is to bound the expectations using results from Bandeira and Van Handel (2016) and Gittens and Tropp (2009).[8] Related arguments are also used recently by Moon (2019). The corresponding lower bound for $\max_{r\in[R]} T_{2\to2}(D_1^r, D_2^r)$ has been known to the random matrix theory literature for some time. To my knowledge, the bound on $\max_{r\in[R]} S_{\infty\to1}(D_1^r, D_2^r)$ is original.

4.3 Simulation evidence

Section 4.2.2 predicts that the test based on the $\infty \to 1$ norm is potentially more powerful than that based on the $2 \to 2$ norm for sparse and degree-heterogeneous alternatives. This subsection provides supporting evidence from two Monte Carlo experiments. It considers the case of unweighted unipartite networks with no loops (symmetric, binary, and hollow adjacency matrices) for simplicity. The purpose of this section is not to simulate data that mimics real-world networks (see instead Section 5), but rather to assess the predictions in a controlled environment. Other simulations (not reported) yielded qualitatively identical results.

4.3.1 The sparse experiment

Sparsity is a common feature of social and economic networks. For example, in many social surveys it is common for agents to report less than a dozen connections. To examine the impact of network sparsity on the power of the two tests, I consider two Erdös-Renyi graph models. In these models, the adjacency matrices are $\{0,1\}$-valued with $P(D_{ij,t} = 1) =$

[8] Both steps can be modified to allow for many types of link dependence as suggested in Section 2.3, Example 1 (see Vershynin 2010, for an example).

$1 - F_{ij,1}(0) = \frac{8}{N}$ and $1 - F_{ij,2}(0) = \frac{5}{N}$ for every $i, j \in [N]$. Agents in the first network have approximately 60% more links than agents in the second network, violating H_0. Applying the two tests to data simulated from the models with $N = 50/100$ and $R = 10000$ yields an average p-value for the test based on the $2 \to 2$ norm of approximately 0.070/0.020 and an average p-value for the test based on the $\infty \to 1$ norm of approximately 0.049/0.013. The test based on the $\infty \to 1$ norm is clearly more powerful, but not dramatically so.

4.3.2 The degree heterogeneity experiment

Degree heterogeneity is another common feature of social and economic networks. For example, in many production and collaboration networks it is common for a small number of agents to have an order of magnitude more links than the median agent. To examine the impact of degree heterogeneity on the power of the two tests, I consider two second-order stochastic blockmodels. In these models, $P(D_{ij,t} = 1) = 1 - F_{ij,t}(0)$, with $F_{1j,1}(0) = F_{1j,2}(0) = .5$ for all $j \in [N]$ and $1 - F_{ij,1}(0) = .02$ and $1 - F_{ij,2}(0) = .08$ for any $i, j \in [N] \setminus [1]$. Agents in the first network have approximately 400% percent more links than in the second network, violating H_0. However, the high degree agent, agent 1, has approximately the same number of links. Applying the two tests to data simulated from the models with $N = 50/100$ and $R = 10000$ yields an average p-value for the test based on the $2 \to 2$ norm of approximately 0.521/0.204 and an average p-value for the test based on the $\infty \to 1$ norm of approximately 0.001/0.000. The test based on the $\infty \to 1$ norm appears to be substantially more powerful.

5 Two empirical demonstrations

I provide two empirical demonstrations using publicly available data. The first demonstration is about measuring a change in network structure over time. A sample of high school students are surveyed annually about their social connections. The networks appear to be less connected and more clustered over time, perhaps because the students place increasing value on having friends in common as they age. The problem is to test whether the observed changes in the reported relationships can be explained by the same random graph model. In this demonstration one can think of the age of the students as the “treatment” and the null

Table 1: Glasgow Social Network Descriptive Statistics

		Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter
First Wave					
	Mean	3.49	0.09	0.35	14
	SD	1.69	0.20		
Third Wave					
	Mean	3.64	0.07	0.42	20
	SD	1.80	0.21		

This table compares two social networks surveyed from 135 Scottish secondary students in the "Teenage Friends and Lifestyle Study." Two students are linked if one student reports being friends or best friends with the other. The First Wave corresponds to all the links reported in the first survey and the Third Wave corresponds to all of the links in the third survey. It describes the means and standard deviations for four measures of network structure: the sequence of agent degrees, eigenvector centralities, clustering coefficients, and diameters of the largest connected component.

hypothesis as the definition of "no treatment effects."

The data comes from the "Teenage Friends and Lifestyle Study" (see Michell and West 1996) in which the researchers survey 160 Scottish students about friendship links during their second, third, and fourth years of secondary school.[9] This example uses the social network surveyed from the first and third waves when the students are respectively 13 and 15 years old. Only students who appear in all three waves are included, yielding a final sample size of $N = 135$.

Descriptive statistics about the two networks are provided in Table 1. These statistics are often used in the economics literature to characterize network structure (see Jackson 2008, Chapter 2). Specifically, the table describes the means and standard deviations of four measures of network structure: the sequence of agent degrees $\{\sum_{j\in[N]} D_{ij}\}_{i\in[N]}$, eigenvector centralities, clustering coefficients $\frac{\sum_{i,j,k\in[N]} D_{ij}D_{ik}D_{jk}}{\sum_{i,j,k\in[N]} D_{ij}D_{ik}}$, and diameters of the largest connected component. The last two measures are scalars and so standard deviations are not reported. The table indicates that while the total number of links appear to be roughly the same for

[9] The data can be found at `https://www.stats.ox.ac.uk/~snijders/siena/Glasgow_data.htm` and is similar in spirit to though considerably less comprehensive than the restricted-use National Longitudinal Study of Adolescent to Adult Health dataset commonly used to study peer effects in economics.

Table 2: Glasgow Network Test Results

R = 10000	Average Degree	Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter	$2 \to 2$ Norm	$\infty \to 1$ Norm
p-value	0.50	0.89	0.19	0.01	0.06	0.01	0.01

This table compares two social networks surveyed from 135 Scottish secondary students in the "Teenage Friends and Lifestyle Study." Two students are linked if one student reports being friends or best friends with the other. The two social networks compared correspond to the collections of links surveyed in the first and third waves of the study. It describes p-values for tests based on the absolute difference in average degree, the mean squared difference in agent degrees, the mean squared difference in eigenvector centralities, absolute difference in clustering coefficients, absolute difference in diameters, $2 \to 2$ norm of the entry-wise differences between the two networks' adjacency matrices, and semidefinite approximation to the $\infty \to 1$ norm of the entry-wise differences between the two networks' adjacency matrices.

both networks, the second network is less interconnected than the first network (as measured by the eigenvector centrality, clustering, and diameter statistics). That these differences are unlikely to be generated by the same random graph model is demonstrated in Table 2.

Table 2 reports the p-value of the randomization test proposed in Section 3 using seven test statistics. Specifically, it reports the p-value

$$(R+1)^{-1}\left(1+\sum_{r\in[R]} \mathbb{1}\left\{T(D_{1,r}, D_{2,r}) \geq T(D_1, D_2)\right\}\right).$$

Each column of the table contains the p-value associated with one of the seven test statistics. The first five test statistics are the absolute difference in average degree, the mean squared difference in agent degrees, the mean squared difference in eigenvector centralities, absolute difference in clustering coefficients, and absolute difference in diameters of the two networks. The last two test statistics are $T_{2\to 2}$ and $S_{\infty\to 1}$ as defined in Section 4.1. Table 2 indicates that the large differences in the clustering and diameters between the two networks are unlikely to be generated by the same random graph model, but this difference would not be detected using the tests based on degree or eigenvector centrality. The implausibility of the null hypothesis is clearly indicated by the tests based on $T_{2\to 2}$ and $S_{\infty\to 1}$.

The second demonstration is about comparing networks generated by different survey questions. A sample of households in a village are surveyed about multiple types of relation-

Table 3: India Social and Economic Network Descriptive Statistics

		Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter
Social Network					
	Mean	3.92	0.26	0.13	6
	SD	3.17	0.22		
Economic Network					
	Mean	4.94	0.27	0.19	5
	SD	3.65	0.21		

This table compares two social networks from village 10 in Banerjee et al. (2013). Two agents are linked in the first network if they "engage socially" and linked in the second if they "borrow or lend money, rice, or kerosene." It describes the means and standard deviations for four measures of network structure: the sequence of agent degrees, eigenvector centralities, clustering coefficients, and diameters of the largest connected component.

ships. The researcher hypothesizes that these survey questions provide the same information about the underlying interactions between agents, so that the researcher may treat them as interchangeable for the purposes of estimating treatment spillovers or modeling information diffusion. In other words, the differences between the networks can be explained by the same random graph model. This is, for example, a key assumption made by Banerjee et al. (2013).

The data for this demonstration comes from Banerjee et al. (2013), in which the researchers survey information about a dozen social and economic connections between households for 75 villages in rural India.[10] This demonstration uses data from the $N = 77$ households in village 10 and compares the social network in which two households are linked if a member of one of the households indicates that they "engage socially" with a member of the other household to the economic network in which two households are linked if a member of one of the households indicates that they "borrow money from," "borrow kerosene or rice from," "lend kerosene or rice to," or lend money to" a member of the other household. Table 3 contains the same summary statistics for the two networks as given in Table 1.

Table 3 describes two key differences between the social and economic networks. The first difference is that the surveyed households have (on average) approximately one more

[10]The data can be found at `https://hdl.handle.net/1902.1/21538`.

Table 4: India Network Test Results

R = 10000	Average Degree	Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter	$2 \to 2$ Norm	$\infty \to 1$ Norm
p-value	0.00	0.58	0.73	0.05	0.55	0.25	0.05

This table compares two social networks from village 10 in Banerjee et al. (2013). Two agents are linked in the first network if they "engage socially" and linked in the second if they "borrow or lend money, rice, or kerosene." It describes p-values for tests based on the absolute difference in average degree, the mean squared difference in agent degrees, the mean squared difference in eigenvector centralities, absolute difference in clustering coefficients, absolute difference in diameters, $2 \to 2$ norm of the entry-wise differences between the two networks' adjacency matrices, and semidefinite approximation to the $\infty \to 1$ norm of the entry-wise differences between the two networks' adjacency matrices.

economic link than social link (five instead of four links). The second difference is that there is more clustering in the economic network. Table 4 below demonstrates that these differences are unlikely to be explained by the same random graph model.

Table 4 contains the same information as Table 2 but for the Banerjee et al. (2013) data. It indicates that the differences between the average degrees and clustering coefficients of the two networks are unlikely to be explained by the null hypothesis. However, this difference would not be detected by a reasonably-sized test based on the $2 \to 2$ norm because $T_{2\to2}(D_1, D_2)$ is in roughly the third quartile of its reference distribution. On the other hand, $S_{\infty\to1}$ is firmly in the upper decile of its reference distribution, and so this test statistic provides more evidence against the null hypothesis.

References

Abadie, A. (2018). Statistical non-significance in empirical economics. Technical report, National Bureau of Economic Research.

Abadie, A. and M. D. Cattaneo (2018). Econometric methods for program evaluation. *Annual Review of Economics 10*, 465–503.

Aliprantis, C. and K. Border (2006). *Infinite dimensional analysis: a hitchhikers guide.* Springer.

Alon, N. and A. Naor (2006). Approximating the cut-norm via grothendieck's inequality. *SIAM Journal on Computing 35*(4), 787–803.

Andrews, I. and M. Kasy (2017). Identification of and correction for publication bias. Technical report, National Bureau of Economic Research.

Aronow, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research 41*(1), 3–16.

Athey, S., D. Eckles, and G. W. Imbens (2018). Exact p-values for network interference. *Journal of the American Statistical Association 113*(521), 230–240.

Badev, A. (2017). Discrete games in endogenous networks: Equilibria and policy. *arXiv preprint arXiv:1705.03137*.

Bandeira, A. S. and R. Van Handel (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability 44*(4), 2479–2506.

Banerjee, A., E. Breza, E. Duflo, and C. Kinnan (2017). Do credit constraints limit entrepreneurship? heterogeneity in the returns to microfinance. *Buffett Institute Global Poverty Research Lab Working Paper* (17-104).

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. O. Jackson (2016). Changes in social network structure in response to exposure to formal credit markets.

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

Beaman, L. A., A. BenYishay, J. Magruder, and A. M. Mobarak (2018). Can network theory-based targeting increase technology adoption?

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*(2), 29–50.

Bloch, F. and M. O. Jackson (2007). The formation of networks with transfers among players. *Journal of Economic Theory 133*(1), 83–110.

Boucher, V. and I. Mourifié (2017). My friend far, far away: a random field approach to exponential random graph models. *The Econometrics Journal 20*(3), S14–S46.

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of econometrics 150*(1), 41–55.

Bramoullé, Y., R. Kranton, and M. D'amours (2014). Strategic interaction and networks. *American Economic Review 104*(3), 898–930.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Candelaria, L. E. (2016). A semiparametric network formation model with multiple linear fixed effects. *Duke University*.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal 20*(3), S1–S13.

Comola, M. and M. Fafchamps (2014). Estimating mis-reporting in dyadic data: Are transfers mutually beneficial?

Comola, M. and S. Prina (2015). Treatment effect accounting for network changes: Evidence from a randomized intervention. *Forthcoming Review of Economics and Statistics*.

Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review*, 35–69.

de Paula, A. (2016). Econometrics of network models. Technical report, cemmap working paper, Centre for Microdata Methods and Practice.

de Paula, A., S. Richards-Shubik, and E. T. Tamer (2014). Identification of preferences in network formation games.

Dzemski, A. (2014). An empirical model of dyadic link formation in a network with unobserved heterogeneity. Technical report, University of Manheim Working Paper.

Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship networks. *Political Analysis 14*(4), 456–487.

Frieze, A. and R. Kannan (1999). Quick approximation to matrices and applications. *Combinatorica 19*(2), 175–220.

Gao, W. Y. (2019). Nonparametric identification in index models of link formation. *Journal of Econometrics*.

Ghoshdastidar, D., M. Gutzeit, A. Carpentier, and U. von Luxburg (2017a). Two-sample hypothesis testing for inhomogeneous random graphs. *arXiv preprint arXiv:1707.00833*.

Ghoshdastidar, D., M. Gutzeit, A. Carpentier, and U. von Luxburg (2017b). Two-sample tests for large random graphs using network statistics. *arXiv preprint arXiv:1705.06168*.

Gittens, A. and J. A. Tropp (2009). Error bounds for random matrix approximation schemes. *arXiv preprint arXiv:0911.4108*.

Goemans, M. X. and D. P. Williamson (1994). .879-approximation algorithms for max cut and max 2sat. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pp. 422–431. ACM.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics 127*(3), 1287–1338.

Goyal, S., M. J. Van Der Leij, and J. L. Moraga-González (2006). Economics: An emerging small world. *Journal of political economy 114*(2), 403–412.

Graham, B. S. (2015). Methods of identification in social networks. *Annu. Rev. Econ. 7*(1), 465–485.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Graham, B. S. (2019). Network data. *Forthcoming Econometrics Handbook Chapter*.

Griffith, A. (2016). Random assignment with non-random peers: A structural approach to counterfactual treatment assessment.

Gualdani, C. (2017). An econometric model of network formation with an application to board interlocks between firms.

Håstad, J. (2001). Some optimal inapproximability results. *Journal of the ACM (JACM) 48*(4), 798–859.

Heß, S. H., D. Jaimovich, and M. Schündeln (2018). Development projects and economic networks: Lessons from rural gambia.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jackson, M. (2008). *Social and economic networks*, Volume 3. Princeton university press Princeton.

Jochmans, K. (2017). Two-way models for gravity. *Review of Economics and Statistics 99*(3), 478–485.

Jochmans, K. and M. Weidner (2019). Fixed-effect regressions on network data. *Econometrica 87*(5), 1543–1560.

Krivine, J.-L. (1979). Constantes de grothendieck et fonctions de type positif sur les spheres. *Advances in Mathematics 31*(1), 16–30.

Lehmann, E. L. and J. P. Romano (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Leung, M. (2016). Treatment and spillover effects under network interference. *Available at SSRN 2757313*.

Leung, M. P. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics 188*(1), 182–195.

Leung, M. P. (2019). A weak law for moments of pairwise stable networks. *Journal of Econometrics*.

Lovász, L. (2012). *Large networks and graph limits*, Volume 60. American Mathematical Soc.

Mele, A. (2017). A structural model of dense network formation. *Econometrica 85*(3), 825–850.

Mele, A. and L. Zhu (2017). Approximate variational estimation for a model of network formation.

Menzel, K. (2015). Strategic network formation with many agents.

Michell, L. and P. West (1996). Peer pressure to smoke: the meaning depends on the method. *Health education research 11*(1), 39–49.

Moon, H. R. (2019). A uniform bound of the operator norm of random element matrices and operator norm minimizing estimation. *arXiv preprint arXiv:1905.01096*.

Nesterov, Y. (1998). Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software 9*(1-3), 141–160.

Nielsen, A. M. and D. Witten (2018). The multiple random dot product graph model. *arXiv preprint arXiv:1811.12172*.

Puelz, D., G. Basse, A. Feller, and P. Toulis (2019). A graph-theoretic approach to randomization tests of causal effects under general interference. *arXiv preprint arXiv:1910.10862*.

Ridder, G. and S. Sheng (2015). Estimation of large network formation games.

Sheng, S. (2012). Identification and estimation of network formation games.

Song, K. (2018). Measuring the graph concordance of locally dependent observations. *Review of Economics and Statistics 100*(3), 535–549.

Tang, M., A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli 23*(3), 1599–1630.

Tao, T. (2012). *Topics in random matrix theory*, Volume 132. American Mathematical Soc.

Toh, K.-C., M. J. Todd, and R. H. Tütüncü (2012). On the implementation and usage of sdpt3–a matlab software package for semidefinite-quadratic-linear programming, version 4.0. In *Handbook on semidefinite, conic and polynomial optimization*, pp. 715–754. Springer.

Toth, P. (2017). Semiparametric estimation in networks with homophily and degree heterogeneity.

Verdier, V. (2017). Estimation and inference for linear models with two-way fixed effects and sparsely matched data. *Review of Economics and Statistics* (0).

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.

Ziliak, S. and D. N. McCloskey (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives.* University of Michigan Press.

A Proofs

A.1 Theorem 1

Proof of Lemma 1: Let X be an arbitrary $N \times N$ dimensional matrix. The lower bound follows from

$$\max_{\varphi \in \mathcal{S}^N} \sum_{i \in [N]} \left(\sum_{j \in [N]} X_{ij} \varphi_j \right)^2 \geq \max_{\varphi \in \mathcal{E}^N} \sum_{i \in [N]} \left(\sum_{j \in [N]} X_{ij} \varphi_j \right)^2 = \max_{j \in [N]} \sum_{i \in [N]} X_{ij}^2$$

where $\mathcal{S}^N$ is the N-dimensional hypersphere $\{\varphi \in \mathbb{R}^N : \sum_{t \in [N]} \varphi_t^2 = 1\}$, $\mathcal{E}^N$ is the usual set of basis vectors in $\mathbb{R}^N$ $\{\varphi \in \mathbb{R}^N : \sum_{t \in [N]} \varphi_t^2 = \sum_{t \in [N]} |\varphi_j| = 1\}$, and the inequality follows from $\mathcal{E}^N \subset \mathcal{S}^N$. Consequently, if $\{X_s\}_{s \in S}$ is any collection of $N \times N$ dimensional matrices indexed by a set S then

$$\max_{s \in S} \max_{\varphi \in \mathcal{S}^N} \sqrt{\sum_{i \in [N]} \left(\sum_{j \in [N]} X_{ij,s} \varphi_j \right)^2} \geq \max_{s \in S} \max_{j \in [N]} \sqrt{\sum_{i \in [N]} X_{ij,s}^2}$$

where $X_{ij,s}$ is the ijth entry of the matrix X_s.

The upper bound follows inequalities by Talagrand (see Boucheron, Lugosi, and Massart 2013, Theorem 6.10) and Bandeira and Van Handel (2016). Specifically, let X be an $N \times N$ dimensional random symmetric matrix with independent and mean-zero entries above the diagonal and zeros on the main diagonal. The entries of X are absolutely bounded by 1. Then for any $\varepsilon > 0$, Talagrand's inequality implies

$$P\left(\max_{\varphi \in \mathcal{S}^N} \sqrt{\sum_{i \in [N]} \left(\sum_{j \in [N]} X_{ij} \varphi_j \right)^2} - E\left[\max_{\varphi \in \mathcal{S}^N} \sqrt{\sum_{i \in [N]} \left(\sum_{j \in [N]} X_{ij} \varphi_j \right)^2} \right] > \varepsilon \right) \leq \exp\left(-\varepsilon^2/2\right).$$

since $\max_{\varphi \in \mathcal{S}^N} \sqrt{\sum_{i \in [N]} \left(\sum_{j \in [N]} X_{ij} \varphi_j \right)^2}$ is convex in X by the triangle inequality.

Corollary 3.2 to Theorem 1.1 of Bandeira and Van Handel (2016) implies

$$E\left[\max_{\varphi \in \mathcal{S}^N} \sqrt{\sum_{i \in [N]} \left(\sum_{j \in [N]} X_{ij} \varphi_j \right)^2} \right] \leq (1+\gamma) \left[2 \max_{j \in [N]} \sqrt{\sum_{i \in [N]} E\left[X_{ij}^2\right]} + \frac{6}{\sqrt{\ln(1+\gamma)}} \sqrt{\ln(N)} \right]$$

for any $\gamma \in [0, 1/2]$. Consequently, for any real positive integer S, collection of $N \times N$ dimensional random symmetric matrices $\{X_s\}_{s\in[S]}$ such that each matrix X_s satisfies the above conditions, and $\gamma \in [0, 1/2]$

$$\max_{s\in[S]} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} X_{ij}\varphi_j\right)^2} > \varepsilon + (1+\gamma)2 \max_{s\in[S]} \max_{j\in[N]} \sqrt{\sum_{i\in[N]} E\left[X_{ij,s}^2\right]} + \frac{6(1+\gamma)}{\sqrt{\ln(1+\gamma)}}\sqrt{\ln(N)}$$

with probability less than $S\exp\left(-\varepsilon^2/2\right)$ by the union bound. Or equivalently, for any $\alpha \in [0, 1]$ and $\gamma \in [0, 1/2]$

$$\max_{s\in[S]} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} X_{ij}\varphi_j\right)^2} \leq \sqrt{-2\ln\left(\frac{\alpha}{S}\right)} + (1+\gamma)2 \max_{s\in[S]} \max_{j\in[N]} \sqrt{\sum_{i\in[N]} E\left[X_{ij,s}^2\right]} + \frac{6(1+\gamma)}{\sqrt{\ln(1+\gamma)}}\sqrt{\ln(N)}$$

with probability greater than $1 - \alpha$. The claim follows. □

Proof of Theorem 1: By two applications of the triangle inequality,

$$\begin{aligned}
\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,1}\leq s} - \mathbb{1}_{D_{ij,2}\leq s}\right)\varphi_j\right)^2} &\geq \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(F_{ij,1}(s) - F_{ij,2}(s)\right)\varphi_j\right)^2} \\
- \left|\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,1}\leq s} - F_{ij,1}(s)\right)\varphi_j\right)^2} \right. &+ \left. \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,2}\leq s} - F_{ij,2}(s)\right)\varphi_j\right)^2}\right|.
\end{aligned}$$

Since $\{\mathbb{1}_{D_1^r\leq s} - \mathbb{1}_{D_2^r\leq s}\}_{s\in\mathbb{R},r\in[R]}$ is a collection of no more than $RN(N-1)$ unique $N \times N$ dimensional random symmetric matrices with independent and mean-zero entries above the diagonal and zeros on the main diagonal, all absolutely bounded by 1, then Lemma 1 implies that for any $\alpha \in [0, 1]$ and $\gamma \in [0, 1/2]$

$$\begin{aligned}
\max_{r\in[R]} \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,1}^r\leq s} - \mathbb{1}_{D_{ij,2}^r\leq s}\right)\varphi_j\right)^2} &\leq \sqrt{-2\ln\left(\frac{\alpha}{RN(N-1)}\right)} \\
+ (1+\gamma)2 \max_{s\in[\mathbb{R}]} \max_{j\in[N]} \sqrt{\sum_{i\in[N]} \nu_{ij}(s)} &+ \frac{6(1+\gamma)}{\sqrt{\ln(1+\gamma)}}\sqrt{\ln(N)}
\end{aligned}$$

with probability greater than $1 - \alpha$, where $\nu_{ij}(s) = [F_{ij,1}(s) + F_{ij,2}(s) - 2F_{ij,1}(s)F_{ij,2}(s)]$. Since

$$\frac{\sqrt{-\ln(\alpha) + \ln(R) + \ln(N)}}{\max_{s\in[S]} \max_{j\in[N]} \sqrt{\sum_{i\in[N]} \nu_{ij}(s)}} \to 0$$

by Assumption 2 and the rate condition $\tau/\sqrt{\ln(N)} \to \infty$,

$$\frac{\max_{r\in[R]}\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{S}^N}\sqrt{\sum_{i\in[N]}\left(\sum_{j\in[N]}\left(\mathbb{1}_{D^r_{ij,1}\leq s}-\mathbb{1}_{D^r_{ij,2}\leq s}\right)\varphi_j\right)^2}}{\max_{s\in[\mathbb{R}]}\max_{j\in[N]}\sqrt{\sum_{i\in[N]}\nu_{ij}(s)}} = O_p(1).$$

The matrices $\{\mathbb{1}_{D_t\leq s}-F_t(s)\}_{s\in\mathbb{R},t\in\{1,2\}}$ also satisfy the hypothesis of Lemma 1, and so too

$$\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{S}^N}\sqrt{\sum_{i\in[N]}\left(\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,1}\leq s}-F_{ij,1}(s)\right)\varphi_j\right)^2}+\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{S}^N}\sqrt{\sum_{i\in[N]}\left(\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,2}\leq s}-F_{ij,2}(s)\right)\varphi_j\right)^2}$$
$$= O_p\left(\max_{s\in[S]}\max_{j\in[N]}\sqrt{\sum_{i\in[N]}\nu_{ij}(s)}\right).$$

The rate condition $T_{2\to 2}(F_1,F_2)/\tau \to \infty$ then implies

$$\frac{\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{S}^N}\sqrt{\sum_{i\in[N]}\left(\sum_{j\in[N]}\left(F_{ij,1}(s)-F_{ij,2}(s)\right)\varphi_j\right)^2}}{\max_{s\in[\mathbb{R}]}\max_{j\in[N]}\sqrt{\sum_{i\in[N]}\nu_{ij}(s)}} \to \infty$$

so

$$\frac{\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{S}^N}\sqrt{\sum_{i\in[N]}\left(\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,1}\leq s}-\mathbb{1}_{D_{ij,2}\leq s}\right)\varphi_j\right)^2}}{\max_{s\in[S]}\max_{j\in[N]}\sqrt{\sum_{i\in[N]}\nu_{ij}(s)}} \to \infty$$

and thus

$$\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{S}^N}\sqrt{\sum_{i\in[N]}\left(\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,1}\leq s}-\mathbb{1}_{D_{ij,2}\leq s}\right)\varphi_j\right)^2} \geq \max_{r\in[R]}\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{S}^N}\sqrt{\sum_{i\in[N]}\left(\sum_{j\in[N]}\left(\mathbb{1}_{D^r_{ij,1}\leq s}-\mathbb{1}_{D^r_{ij,2}\leq s}\right)\varphi_j\right)^2}$$

eventually. The claim follows. □

A.2 Theorem 2

The proof of Lemma 2 relies on the following inequality due to Grothendieck and Krivine (1979).

Theorem (Grothendieck): Let X be an arbitrary $N\times N$ dimensional real matrix such that

$$\max_{\varphi,\psi\in\mathbb{R}^N:||\varphi||_\infty,||\psi||_\infty\leq 1}\left|\sum_{i\in[N]}\sum_{j\in[N]}X_{ij}\varphi_j\psi_i\right| \leq 1.$$

Then

$$\max_{\varphi,\psi\in\mathcal{H}:||\varphi||_{\mathcal{H}},||\psi||_{\mathcal{H}}\leq 1}\left|\sum_{i\in[N]}\sum_{j\in[N]}X_{ij}<\tilde{\varphi}_i,\tilde{\psi}_j>_{\mathcal{H}}\right|\leq K=\frac{\pi}{2\ln\left(1+\sqrt{2}\right)}\leq 1.783$$

where $\mathcal{H}$ is an arbitrary Hilbert space and $||\cdot||_{\mathcal{H}}$ and $<\cdot,\cdot>_{\mathcal{H}}$ are the associated norm and inner product operators.

Proof of Lemma 2: Let X be an arbitrary $N\times N$ dimensional matrix. The lower bound then follows from

$$\begin{aligned}
&K\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}X_{ij}\varphi_j\right|=K\max_{\varphi,\psi\in\mathcal{C}^N}\sum_{i\in[N]}\sum_{j\in[N]}X_{ij}\varphi_j\psi_i\geq\max_{\varphi,\psi\in\mathcal{M}^N}\sum_{i\in[N]}\sum_{j\in[N]}X_{ij}\sum_{s\in[N]}\varphi_{js}\psi_{is}\\
&\geq\sum_{i\in[N]}\sum_{j\in[N]}X_{ij}\sum_{s\in[N]}\frac{X_{is}}{\sqrt{\sum_{s\in[N]}X_{is}^2}}\mathbb{1}\{j=s\}=\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}X_{ij}^2}
\end{aligned}$$

where $\mathcal{C}^N$ is the N-dimensional hypercube $\{-1,1\}^N$, $\mathcal{M}^N$ is the set of $N\times N$ matrices with rows of Euclidean length 1 $\{\Lambda\in\mathbb{R}^{N\times N}:\sum_{j=1}^N\Lambda_{ij}^2=1\forall i\in[N]\}$, and the first inequality is due to Grothendieck. Consequently, if $\{X_s\}_{s\in S}$ is any collection of $N\times N$ dimensional matrices indexed by a set S then

$$K\max_{s\in S}\max_{\varphi,\psi\in\mathcal{C}^N}\sum_{i\in[N]}\sum_{j\in[N]}X_{ij,s}\varphi_j\psi_i\geq\max_{s\in S}\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}X_{ij,s}^2}$$

where $X_{ij,s}$ is the ijth entry of matrix X_s.

The upper bound also follows from Talagrand's inequality and an inequality due to Gittens and Tropp (2009). Specifically, let X be an $N\times N$ dimensional random symmetric matrix with independent and mean-zero entries above the diagonal and zeros on the main diagonal. The entries of X are absolutely bounded by 1. Then for any $\varepsilon>0$, Talagrand's inequality implies

$$P\left(\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}X_{ij}\varphi_j\right|-E\left[\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}X_{ij}\varphi_j\right|\right]>\varepsilon\right)\leq\exp\left(-\varepsilon^2/2\right).$$

since $\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}X_{ij}\varphi_j\right|$ is convex in X by the triangle inequality.

Corollary 2 to Theorem 3 of Gittens and Tropp (2009) implies

$$E\left[\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}X_{ij}\varphi_j\right|\right]\leq 2\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}E\left[X_{ij}^2\right]}.$$

Consequently, for any real positive integer S and collection of $N\times N$ dimensional random symmetric matrices $\{X_s\}_{s\in[S]}$ such that each matrix X_s satisfies the above conditions

$$\max_{s\in[S]}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}X_{ij,s}\varphi_j\right| > \varepsilon + 2\max_{s\in[S]}\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}E\left[X_{ij,s}^2\right]}$$

with probability less than $S\exp\left(-\varepsilon^2/2\right)$ by the union bound. So for any $\alpha\in[0,1]$

$$\max_{s\in[S]}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}X_{ij,s}\varphi_j\right| \leq \sqrt{-2\ln\left(\frac{\alpha}{S}\right)} + 2\max_{s\in[S]}\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}E\left[X_{ij,s}^2\right]}$$

with probability greater than $1-\alpha$. The claim follows. $\square$

Proof of Theorem 2: By two applications of the triangle inequality,

$$\begin{aligned}\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,1}\leq s}-\mathbb{1}_{D_{ij,2}\leq s}\right)\varphi_j\right| &\geq \max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}\left(F_{ij,1}(s)-F_{ij,2}(s)\right)\varphi_j\right| \\ &- \left|\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,1}\leq s}-F_{ij,1}(s)\right)\varphi_j\right| + \max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,2}\leq s}-F_{ij,2}(s)\right)\varphi_j\right|\right|.\end{aligned}$$

Since $\{\mathbb{1}_{D_1^r\leq s}-\mathbb{1}_{D_2^r\leq s}\}_{s\in\mathbb{R},r\in[R]}$ is a collection of no more than $RN(N-1)$ unique $N\times N$ dimensional random symmetric matrices with independent and mean-zero entries above the diagonal and zeros on the main diagonal, all absolutely bounded by 1, then Lemma 2 implies that for any $\alpha\in[0,1]$

$$\max_{r\in[R]}\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,1}^r\leq s}-\mathbb{1}_{D_{ij,2}^r\leq s}\right)\varphi_j\right| \leq \sqrt{-2\ln\left(\frac{\alpha}{RN(N-1)}\right)} + 2\max_{s\in\mathbb{R}}\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}\nu_{ij}(s)}$$

with probability greater than $1-\alpha$ where $\nu_{ij}(s)=[F_{ij,1}(s)+F_{ij,2}(s)-2F_{ij,1}(s)F_{ij,2}(s)]$. Since

$$\frac{-\ln(\alpha)+\ln(R)+\ln(N)}{\max_{s\in\mathbb{R}}\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}\nu_{ij}(s)}}\to 0$$

by Assumption 2 and the rate condition $\sigma/\sqrt{\ln(N)}\to\infty$,

$$\frac{\max_{r\in[R]}\max_{s\in\mathbb{R}}\max_{\varphi\in\mathcal{C}^N}\sum_{i\in[N]}\left|\sum_{j\in[N]}\left(\mathbb{1}_{D_{ij,1}^r\leq s}-\mathbb{1}_{D_{ij,2}^r\leq s}\right)\varphi_j\right|}{\max_{s\in\mathbb{R}}\sum_{i\in[N]}\sqrt{\sum_{j\in[N]}\nu_{ij}(s)}} = O_p(1).$$

The matrices $\{\mathbb{1}_{D_t \leq s} - F_t(s)\}_{s\in\mathbb{R}, t\in\{1,2\}}$ also satisfy the hypothesis of Lemma 2, and so too

$$\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{C}^N} \sum_{i\in[N]} \left| \sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,1}\leq s} - F_{ij,1}(s)\right) \varphi_j \right| + \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{C}^N} \sum_{i\in[N]} \left| \sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,2}\leq s} - F_{ij,2}(s)\right) \varphi_j \right|$$
$$= O_p\left(\max_{s\in\mathbb{R}} \sum_{i\in[N]} \sqrt{\sum_{j\in[N]} \nu_{ij}(s)} \right).$$

The rate condition $T_{\infty\to1}(F_1, F_2)/\sigma \to \infty$ then implies

$$\frac{\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{C}^N} \sum_{i\in[N]} \left| \sum_{j\in[N]} \left(F_{ij,1}(s) - F_{ij,2}(s)\right) \varphi_j \right|}{\max_{s\in\mathbb{R}} \sum_{i\in[N]} \sqrt{\sum_{j\in[N]} \nu_{ij}(s)}} \to \infty$$

so

$$\frac{\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{C}^N} \sum_{i\in[N]} \left| \sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,1}\leq s} - \mathbb{1}_{D_{ij,2}\leq s}\right) \varphi_j \right|}{\max_{s\in\mathbb{R}} \sum_{i\in[N]} \sqrt{\sum_{j\in[N]} \nu_{ij}(s)}} \to \infty$$

and thus

$$\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{C}^N} \sum_{i\in[N]} \left| \sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,1}\leq s} - \mathbb{1}_{D_{ij,2}\leq s}\right) \varphi_j \right| \geq \max_{r\in[R]} \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{C}^N} \sum_{i\in[N]} \left| \sum_{j\in[N]} \left(\mathbb{1}_{D^r_{ij,1}\leq s} - \mathbb{1}_{D^r_{ij,2}\leq s}\right) \varphi_j \right|$$

eventually. The claim follows. $\square$

A.3 Theorem 3

Proof of Theorem 3: I demonstrate the claim for the test based on the $2 \to 2$ norm since the proof of that based on the $\infty \to 1$ norm is identical. The proof is constructive in that, for any sequence $\delta_N \to \infty$, it specifies a specific sequence of distribution function matrices F_1 and F_2, depending on δ_N, such that $\delta_N T_{2\to2}(F_1, F_2)/\tau \to \infty$ or

$$\delta_N \frac{\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(F_{ij,1}(s) - F_{ij,2}(s)\right) \varphi_j \right)^2}}{\max_{s\in[S]} \max_{j\in[N]} \sqrt{\sum_{i\in[N]} \left[F_{ij,1}(s) + F_{ij,2}(s) - 2F_{ij,1}(s)F_{ij,2}(s)\right]}} \to \infty$$

and

$$P\left((R+1)^{-1} \left(1 + \sum_{r\in[R]} \mathbb{1}\{T_{2\to2}(D^r_1, D^r_2) \geq T_{2\to2}(D_1, D_2)\} \right) \leq \alpha \right) \to \alpha.$$

The proof has three steps. The first step is to specify F_1 and F_2. For an arbitrary $\varepsilon > 0$, define $A_{1-\varepsilon} = [\lceil (1-\varepsilon)\, N \rceil]$ and $A_\varepsilon = [N] \setminus A_{1-\varepsilon}$. That is, let $A_{1-\varepsilon}$ index the first $\lceil (1-\varepsilon) N \rceil$ agents in the sample and A_ε the last $\lfloor \varepsilon N \rfloor$. Suppose $F_{ij,1} = F_{ij,2}$ for $i, j \in A_{1-\varepsilon}$ with $F_{ij,1}$ and $F_{ij,2}$ uniformly bounded away from 0 and 1, $F_{ij,1} = F_{ij,2} = 0$ for $i \in A_\varepsilon$ and $j \in A_{1-\varepsilon}$ (or $i \in A_{1-\varepsilon}$ and $j \in A_\varepsilon$), and $F_{ij,1} = 1 + F_{ij,2}$ for $i, j \in A_\varepsilon$.

The second step is to fix $\varepsilon = (\delta_N N)^{-1/2}$. Note that since $T_{2\to 2}$ is $O\left(N\varepsilon\right)$ and τ is $O(\sqrt{N})$ by construction from the first step, it follows that $T_{2\to 2}(F_1, F_2)/\tau \to 0$, but $\delta_N T_{2\to 2}(F_1, F_2)/\tau \to \infty$.

The third step is then to apply the triangle inequality twice. The first application gives

$$\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(\mathbb{1}_{D_{ij,1}\le s} - \mathbb{1}_{D_{ij,2}\le s}\right) \varphi_j \right)^2}$$
$$\ge \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_{1-\varepsilon}} \left(\sum_{j\in A_{1-\varepsilon}} \left(\mathbb{1}_{D_{ij,1}\le s} - \mathbb{1}_{D_{ij,2}\le s}\right) \varphi_j \right)^2} - \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_\varepsilon} \left(\sum_{j\in A_\varepsilon} \left(\mathbb{1}_{D_{ij,1}\le s} - \mathbb{1}_{D_{ij,2}\le s}\right) \varphi_j \right)^2}.$$

The second application gives

$$\max_{r\in[R]} \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in[N]} \left(\sum_{j\in[N]} \left(\mathbb{1}_{D^r_{ij,1}\le s} - \mathbb{1}_{D^r_{ij,2}\le s}\right) \varphi_j \right)^2}$$
$$\le \max_{r\in[R]} \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_{1-\varepsilon}} \left(\sum_{j\in A_{1-\varepsilon}} \left(\mathbb{1}_{D^r_{ij,1}\le s} - \mathbb{1}_{D^r_{ij,2}\le s}\right) \varphi_j \right)^2} + \max_{r\in[R]} \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_\varepsilon} \left(\sum_{j\in A_\varepsilon} \left(\mathbb{1}_{D^r_{ij,1}\le s} - \mathbb{1}_{D^r_{ij,2}\le s}\right) \varphi_j \right)^2}.$$

Both $\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_\varepsilon} \left(\sum_{j\in A_\varepsilon} \left(\mathbb{1}_{D_{ij,1}\le s} - \mathbb{1}_{D_{ij,2}\le s}\right) \varphi_j \right)^2}$ and $\max_{r\in[R]} \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_\varepsilon} \left(\sum_{j\in A_\varepsilon} \left(\mathbb{1}_{D^r_{ij,1}\le s} - \mathbb{1}_{D^r_{ij,2}\le s}\right) \varphi_j \right)^2}$ are bounded by $N\varepsilon$ by construction and thus are $o\left(\sqrt{N}\right)$ by the second step. On the other hand, $\max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_{1-\varepsilon}} \left(\sum_{j\in A_{1-\varepsilon}} \left(\mathbb{1}_{D_{ij,1}\le s} - \mathbb{1}_{D_{ij,2}\le s}\right) \varphi_j \right)^2}/\sqrt{N}$ and $\max_{r\in[R]} \max_{s\in\mathbb{R}} \max_{\varphi\in\mathcal{S}^N} \sqrt{\sum_{i\in A_{1-\varepsilon}} \left(\sum_{j\in A_{1-\varepsilon}} \left(\mathbb{1}_{D^r_{ij,1}\le s} - \mathbb{1}_{D^r_{ij,2}\le s}\right) \varphi_j \right)^2}/\sqrt{N}$ are bounded away from 0 by the lower bound in Lemma 1. Since these two are identically distributed and nondegenerate by construction, the result follows. $\square$

A.4 Theorem 4

Proof of Theorem 4: The claim is proven by checking the hypotheses of Theorems 1 and 2. This is done in three steps. The first step is to demonstrate that the assumption that there exists $I_N, J_N \subseteq [N]$ with $\liminf_{N\to\infty} \frac{|I_N|\wedge|J_N|}{N} > 0$ and $\rho_N > 0$ such that for all $i \in I_N, j \in J_N$ there exists s such that $(F_{ij,1}(s) - F_{ij,2}(s)) > \rho_N$ or $(F_{ij,1}(s) - F_{ij,2}(s)) < \rho_N$ implies that $T_{2\to2}(F_1, F_2) \geq \rho_N N$ and $T_{\infty\to1}(F_1, F_2) \geq \rho_N N^2$ eventually $(N \to \infty)$. Write $\delta = \liminf_{N\to\infty} \frac{|I_N|\wedge|J_N|}{N} > 0$. Then

$$\begin{aligned}
T_{2\to2}(F_1,F_2) &= \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_2=1} \left(\sum_{i\in[N]} \left(\sum_{j\in[N]} (F_{ij,1}(s) - F_{ij,2}(s))\varphi_j \right)^2 \right)^{1/2} \\
&\geq \left(\sum_{i\in I_N} \left(\sum_{j\in J_N} \frac{(F_{ij,1}(s) - F_{ij,2}(s))}{\sqrt{\sum_{j\in J_N} 1}} \right)^2 \right)^{1/2} \geq \rho_N N \delta^2
\end{aligned}$$

eventually and

$$\begin{aligned}
T_{\infty\to1}(\rho_N F_1, \rho_N F_2) &= \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_\infty=1} \sum_{i\in[N]} \left| \sum_{j\in[N]} (F_{ij,1}(s) - F_{ij,2}(s)\varphi_j \right| \\
&= \max_{s\in\mathbb{R}} \max_{\varphi:||\varphi||_\infty=1} \max_{\psi:||\psi||_\infty=1} \sum_{i\in[N]} \sum_{j\in[N]} (F_{ij,1}(s) - F_{ij,2}(s))\varphi_j\psi_i \\
&\geq \max_{s\in\mathbb{R}} \max_{\varphi\in\{0,1\}^N} \max_{\psi\in\{0,1\}^N} \left| \sum_{i\in[N]} \sum_{j\in[N]} (F_{ij,1}(s) - F_{ij,2}(s))\varphi_j\psi_i \right| \\
&\geq \max_{s\in\mathbb{R}} \left| \sum_{i\in I_N} \sum_{j\in J_N} (F_{ij,1}(s) - F_{ij,2}(s)) \right| \geq \rho_N N^2 \delta^2
\end{aligned}$$

eventually where the first inequality follows from the fact that for any $N \times N$ dimensional matrix X

$$\begin{aligned}
\max_{\varphi\in\{0,1\}^N} \max_{\psi\in\{0,1\}^N} \left| \sum_{i\in[N]} \sum_{j\in[N]} X_{ij}\varphi_j\psi_i \right| &= \max_{\varphi\in\{-1,1\}^N} \max_{\psi\in\{-1,1\}^N} \left| \sum_{i\in[N]} \sum_{j\in[N]} X_{ij} \left(\frac{\varphi_j+1}{2}\right) \left(\frac{\psi_i+1}{2}\right) \right| \\
&= \max_{\varphi\in\{-1,1\}^N} \max_{\psi\in\{-1,1\}^N} \left| \sum_{i\in[N]} \sum_{j\in[N]} \left[X_{ij}\varphi_j\psi_i + X_{ij} + X_{ij}\varphi_j + X_{ij}\psi_i\right]/4 \right| \\
&\leq \max_{\varphi\in\{-1,1\}^N} \max_{\psi\in\{-1,1\}^N} \sum_{i\in[N]} \sum_{j\in[N]} X_{ij}\varphi_j\psi_i = \max_{\varphi:||\varphi||_\infty=1} \max_{\psi:||\psi||_\infty=1} \sum_{i\in[N]} \sum_{j\in[N]} X_{ij}\varphi_j\psi_i.
\end{aligned}$$

where the last inequality follows from distributing the maximzation over the sum.

The second step is to observe that the assumption that there exists $I_N, J_N \subseteq [N]$ with $\liminf_{N\to\infty} \frac{|I_N|\wedge|J_N|}{N} > 0$ and $\rho_N > 0$ such that for all $i \in I_N, j \in J_N$ there exists s such that

$(F_{ij,1}(s) - F_{ij,2}(s)) > \rho_N$ or $(F_{ij,1}(s) - F_{ij,2}(s)) < \rho_N$ implies that

$$\delta\sqrt{\rho_N N} \leq \tau \leq \sqrt{2}\sqrt{\rho_N N} \text{ and } \delta\sqrt{\rho_N N^3} \leq \sigma \leq \sqrt{2}\sqrt{\rho_N N^3}$$

eventually. Without loss of generality suppose that $(F_{ij,1}(s) - F_{ij,2}(s)) > \rho_N$. The two upper bounds then follow from the fact that $F_{ij,t}(s)$ is bounded in $[0, 1]$. The first lower bound follows from

$$\begin{aligned}
\tau &= \max_{s\in\mathbb{R}} \max_{i\in[N]} \sqrt{\sum_{j\in[N]} (F_{ij,1}(s) + F_{ij,2}(s) - 2F_{ij,1}(s)F_{ij,2}(s))} \\
&\geq \max_{s\in\mathbb{R}} \max_{i\in I_N} \sqrt{\sum_{j\in J_N} ([F_{ij,1}(s) - F_{ij,2}(s)] + 2F_{ij,2}(s)(1 - F_{ij,1}(s)))}
\end{aligned}$$

and the fact that $2F_{ij,2}(s)(1 - F_{ij,1}(s))$ is not negative. Similarly

$$\begin{aligned}
\sigma &= \max_{s\in\mathbb{R}} \sum_{i\in[N]} \sqrt{\sum_{j\in[N]} (F_{ij,1}(s) + F_{ij,2}(s) - 2F_{ij,1}(s)F_{ij,2}(s))} \\
&\geq \max_{s\in\mathbb{R}} \sum_{i\in I_N} \sqrt{\sum_{j\in J_N} ([F_{ij,1}(s) - F_{ij,2}(s)] + 2F_{ij,2}(s)(1 - F_{ij,1}(s)))}
\end{aligned}$$

implies the second lower bound.

The third step is to observe that steps 1 and 2 imply that

$$T_{2\to2}(F_1, F_2)/\tau \geq \sqrt{\rho_N N}\delta^2/\sqrt{2} \text{ and } T_{\infty\to1}(F_1, F_2)/\sigma \geq \sqrt{\rho_N N}\delta^2/\sqrt{2}$$

eventually so that $T_{2\to2}(F_1, F_2)/\tau \to \infty$ and $T_{\infty\to1}(F_1, F_2)/\sigma \to \infty$ so long as $\rho_N N \to \infty$. Since

$$\sigma/\sqrt{\ln(N)} \geq \delta\sqrt{\rho_N N^3/\ln(N)}$$

eventually, $\rho_N N \to \infty$ also implies that $\sigma/\sqrt{\ln(N)} \to \infty$ and so the hypothesis of Theorem 2 is satisfied. Since

$$\tau/\sqrt{\ln(N)} \geq \delta\sqrt{\rho_N N/\ln(N)},$$

eventually, strengthening the rate condition to $\rho_N N/\ln(N) \to \infty$ implies that $\tau/\sqrt{\ln(N)} \to \infty$ so that the hypothesis of Theorem 1 is also satisfied. This demonstrates the proof. $\square$