

Testing for Differences in Stochastic Network Structure

Eric Auerbach*

August 14, 2021

Abstract

How can one determine whether a 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 $2 \rightarrow 2$ and $\infty \rightarrow 1$ 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 $\infty \rightarrow 1$ norm can be substantially more powerful for the kinds of sparse and degree-heterogeneous networks common in economics.

1 Introduction

This paper proposes analogues of a two-sample Kolmogorov-Smirnov (KS) test for networks. The KS test is a standard way to assess whether two random vectors come from the same distribution and has many applications in economics. One is detecting distributional treatment effects. While tests that compare averages or ranks may ignore key differences between distributions, the KS test with enough data can detect any fixed difference.

*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, Bryan Graham, Joel Horowitz, Guido Imbens, Chuck Manski, Angelo Mele, Roger Moon, Aureo de Paula, Sida Peng, Stephen Nei, Mikkel Sølvsten, Alireza Tahbaz-Salehi, Chris Udry, and anonymous referees for helpful suggestions.

What is an analogous way to detect differences between networks? This paper proposes tests to assess whether two stochastic networks come from the same random graph model. While tests that compare measures of density, clustering, or centrality (see for instance Banerjee et al. 2018) may ignore key differences between models, the proposed tests, like the KS test for distributions, with enough data can detect any fixed difference.

Section 2 describes the model, testing problem, and applications. The model is a non-parametric version of a class of dyadic regression models popular in the literature. The null hypothesis is that two networks are drawn from the same model. Applications include tests for network non-stationarity, treatment effects, externalities, and more.

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. The test controls size by construction. Power depends on a choice of test statistic.

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 $2 \rightarrow 2$ operator norm (also known as the spectral norm or radius) of the difference between the networks' adjacency matrices. The second test statistic is based on the $\infty \rightarrow 1$ operator norm. The tests are easy to implement and quick to compute for networks connecting hundreds of agents. A key result is that while both tests with enough data will detect any fixed difference between models, the test based on the $\infty \rightarrow 1$ norm can be much more powerful when there is nontrivial heterogeneity in the row-variances of the adjacency matrices. Such row-heteroskedasticity may occur when the networks are sparse or have heavy-tailed degree distributions. This characterizes many social and economic networks (see for instance Jackson and Rogers 2007).

Section 5 provides two empirical demonstrations using real-world social network data. In both examples, the $\infty \rightarrow 1$ norm has sufficient power to detect the relevant difference, corroborating theoretical results. Alternatives are less reliable. R code for implementation can be found on my website.¹ Section 6 concludes with some example extensions.

¹<https://sites.google.com/site/ericjauerbach/>

2 Framework

2.1 Model

I focus on undirected unipartite networks defined on a set of N agents referred to as a *community* and indexed by $[N] := \{1, 2, \dots, N\}$. 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. The variable $D_{ij,t}$ for $t \in [2]$ records the realized relationship t between agents i and j . The $N \times N$ symmetric adjacency matrix D_t contains $D_{ij,t}$ in the ij th and ji th entries.

I suppose the networks to be compared are defined on the same community of agents. In many settings, the community is defined so that this is the case. For instance, to study trade networks, economists often define the unit of analysis to be countries and the links to be the amount of trade between countries in a year, even if the agents that actually engage in trade are people and firms that change over time. Testing for differences between networks defined on two different communities generally requires more structure describing exactly how the communities ought to be compared. This is left to future work.

Directed or bipartite networks can be 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. The variable $D_{ij,t}^*$ records the realized relationship t between agent i and market j . The $N_1 \times N_2$ matrix D_t^* contains $D_{ij,t}^*$ in the ij th entry. This asymmetric rectangular adjacency matrix is transformed into a symmetric square one

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

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 an

$N \times N$ symmetric matrix with $N = N_1 + N_2$. It follows that the focus on undirected unipartite networks (symmetric and square adjacency matrices) is without loss. This point is also made by Bandeira and Van Handel (2016), Section 3.1.

I suppose the entries of D_1 and D_2 are equal to 0 on the main diagonal (no self-links) and mutually independent above the main diagonal. The assumption that $\{(D_{ij,1}, D_{ij,2})\}_{i>j \in [N]}$ are independent across agent-pairs is strong but common in the dyadic fixed-effects regression literature, see below. The assumption that $D_{ij,1}$ and $D_{ij,2}$ are independent is also strong but can be dropped in many cases, see Online Appendix Section E. The marginal distribution of $D_{ij,t}$ is denoted $F_{ij,t}(\cdot)$ and the $N \times N$ matrix F_t contains $F_{ij,t}(\cdot)$ in the ij th entry. A generic matrix of distribution functions F_t is referred to as a *random graph model*.

A concrete example of an unweighted random graph model is $D_{ij,t} = \mathbb{1}\{\mu_{ij,t} \geq \varepsilon_{ij,t}\}$, where $\mu_{ij,t} = f_t(\alpha_{i,t}, \alpha_{j,t}, w_{ij,t})$, $\alpha_{i,t}$ 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 idiosyncratic error that is independently distributed across agent-pairs with marginal distribution $G_{ij,t}$ (see generally Graham 2019, Section 6.3). I treat 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 they are drawn from some distribution, the random graph model is defined conditional on their realization. Such conditioning is common in the fixed-effects literature.

The only remaining source of randomness are the $\{\varepsilon_{ij,t}\}_{i,j \in [N], t \in [2]}$ and so the off-diagonal entries $\{D_{ij,t}\}_{i>j \in [N], t \in [2]}$ are mutually independent. The marginal distribution of $D_{ij,t}$ is given by $F_{ij,t}(s) = 1 - G_{ij,t}(f_t(\alpha_{i,t}, \alpha_{j,t}, w_{ij,t}))$ for $s \in [0, 1]$. The random graph model F_t is parametrized by the agent effects, attributes, 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. This is formalized by the statement of the testing problem below. Not every difference in model parameters can be detected, however. To constitute evidence against the hypothesis of no treatment effects, the difference in parameters must yield a sufficiently large difference between F_1 and F_2 . Other random graph models may imply other definitions of a treatment effect. Their study is left to future work.

2.2 Testing problem

This paper tests the null hypothesis

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

against the alternative

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

In words, H_0 is the hypothesis that D_1 and D_2 are drawn from the same random graph model. For the concrete example of Section 2.1, the problem is equivalent to testing the hypothesis that $G_{ij,1}(f_1(\alpha_{i,1}, \alpha_{j,1}, w_{ij,1})) = G_{ij,2}(f_2(\alpha_{i,2}, \alpha_{j,2}, w_{ij,2}))$ for every $i, j \in [N]$. The hypothesis may be false whenever the two random graph models have different agent-specific effects, agent-pair attributes, community link functions, or distributions of idiosyncratic errors. Distinguishing between these parameters generally requires more structure. For example, if the $\{\varepsilon_{ij,t}\}_{i,j \in [N], t \in [2]}$ are 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. If the $\{\varepsilon_{ij,t}\}_{i,j \in [N], t \in [2]}$ are 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 Online Appendix Section B.2 for a discussion.

The logic behind this test is that any difference in the distribution of idiosyncratic errors, agent fixed effects, link function, etc. reflects different incentives for agents to form or report links. Differences due to the idiosyncratic errors reflect only policy-irrelevant statistical noise. As previously motivated, the test is specifically designed to detect a large class of differences between two random graph models. Other testing problems tailored to detect more specific differences are discussed below.

Related testing problems have been considered in the literature. Tang et al. (2017) and Nielsen and Witten (2018) propose tests that rely on a low-dimensional dot-product structure. Ghoshdastidar et al. (2017) propose a test based on the $2 \rightarrow 2$ norm that relies on many independent draws from the same random graph model. A related application of

randomization-based inference to networks tests for interference (see Aronow 2012; Athey et al. 2018). Rather than study the influence of a treatment on network structure, this literature studies the influence of a network on agents' exposure to a treatment.

2.3 Example applications

I describe five applications to testing problems in the economics literature. Details can be found in Online Appendix Section D.

2.3.1 Application 1: a test of link stationarity

Goyal et al. (2006) observe co-authorships between economists over time and argue that the profession has become more interconnected in response to new research technologies such as the internet. The above framework can be used to test whether these differences over time are statistically significant. Let $D_{ij,t}$ describe the existence of a co-authorship between economists i and j in time period t . Then $H_0 : F_{ij,1} = F_{ij,2}$ is the hypothesis of link stationarity that the distribution of co-authorship links does not change over time. The first example in Section 5 below demonstrates this application to testing link stationarity.

2.3.2 Application 2: a test for link heterogeneity

Banerjee et al. (2013) collect data on a dozen social and economic ties between villagers. Jackson et al. (2017) suggest that this data may “encode richer information than simply identifying whether two people are close or not.” The above framework can be used to test whether the differences between networks induced by different survey questions are statistically significant. Let $D_{ij,1}$ denote whether agents i and j have one type of connection (e.g. they report a friendship) and $D_{ij,2}$ denote whether agents i and j have another type of connection (e.g. they report having borrowed or lent money). Then $H_0 : F_{ij,1} = F_{ij,2}$ is the hypothesis of link homogeneity that the distribution of network links does not vary with the choice of survey question. The second example in Section 5 below demonstrates this application to testing link homogeneity.

2.3.3 Application 3: a test of no treatment effects

Rose (2004) finds that trade agreements such as the WTO do not significantly alter international trade. More broadly, the above framework can be used to test whether program participation leads to a statistically significant change in network structure. Let $D_{ij,1}$ describe the outcome of interest (e.g. logarithm of the total value of trade) when either i or j participates in the program (e.g. is a member of the WTO) and $D_{ij,2}$ describe the outcome of interest when neither agent participates in the program. Then $H_0 : F_{ij,1} = F_{ij,2}$ is the hypothesis of no treatment effects that the distribution of network links is unrelated to agent participation in the program. One can extend the framework to incorporate unobserved agent and agent-pair heterogeneity, see Online Appendix Section D.3.

2.3.4 Application 4: a test for endogenous link formation

Goldsmith-Pinkham and Imbens (2013) consider a model of student GPA and link formation, and test whether a determinant of GPA also drives link formation in the network. They propose a one-sample parametric test for such endogenous link formation. The above framework provides an alternative two-sample nonparametric test. Let $D_{ij,1}$ indicate whether there is a friendship between students i and j when they are socially proximate (e.g. they participate in similar extracurricular activities) and $D_{ij,2}$ indicate whether there is a friendship between students i and j when they are not socially proximate. Then $H_0 : F_{ij,1} = F_{ij,2}$ is the hypothesis of exogenous link formation that the distribution of network links is unrelated to the agents' social proximity. See Online Appendix Section D.8 for a related one-sample test.

2.3.5 Application 5: a test for network externalities

Pelican and Graham (2020) consider a model of link formation in which the decision for two agents to link may depend on the existence of other links in the network. They propose a one-sample parametric test for such network externalities. The above framework provides an alternative two-sample nonparametric test. For example, suppose that the network links are observed in an equilibrium satisfying $D_{ij,t} = \mathbb{1}\{\alpha_{ij} + \gamma_{ij} \sum_{k=1}^N D_{ik,t} D_{jk,t} \geq \varepsilon_{ij,t}\}$ where α_{ij} is fixed and $\varepsilon_{ij,t}$ is independently distributed across agent-pairs with marginal distribution

G_{ij} . Then $H_0 : \gamma_{ij} = 0$ is the hypothesis of no network externalities. It implies that both D_1 and D_2 are drawn from the random graph model F with $F_{ij}(s) = 1 - G_{ij}(\alpha_{ij})$ for $s \in [0, 1]$. The issue of multiple or no equilibria that complicates the estimation of γ_{ij} does not impact the validity of this test because there are no network externalities under H_0 .

2.3.6 Potential problems with the assumptions

While the framework can be widely applied, its assumptions may be problematic in some settings. First, the test requires data from at least two networks defined on the same community of agents. Second, the test is sensitive to any non-idiosyncratic changes in the distribution of network links, whether or not they are policy relevant. Third, while the tests proposed below have the power to detect many potential differences between networks, they may have low power against specific alternatives. Ultimately, researchers should not rely exclusively on any one model or test. This paper offers one general way to determine whether the observed differences between two stochastic networks are statically significant, but it should be used in conjunction with other tests and estimation procedures in practice.

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. Examples include differences in centrality measures such as agent degree, eigenvector centrality, or clustering. Certain centrality measures may direct the power of the test towards specific alternatives.

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,

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

where $D_{ij,1}^r$ is the ij th 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

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

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 that when H_0 is true the probability of an (incorrect) rejection does not exceed α , or

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

This is true for any test statistic $T(D_1, D_2)$. When the researcher has a collection of network statistics, the tests can be combined in the usual way. One can similarly test whether any number of adjacency matrices are drawn from the same random graph model by permuting all of the corresponding entries.

4 Two tests based on operator norms

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 that it tends to (correctly) reject the null hypothesis when it is false. This section proposes two statistics based on operator norms that have power against a large class of alternative hypotheses.

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

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

where $\|\cdot\|_p$ refers to the vector p -norm, φ is an 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 ij th entry if $X_{ij} \leq s$ and 0 otherwise. When D_1 and D_2 are unweighted (i.e. they have $\{0, 1\}$ -valued entries), $T_{p \rightarrow q}(D_1, D_2) = \max_{\varphi: \|\varphi\|_p=1} \|(D_1 - D_2)\varphi\|_q$ which is the $p \rightarrow q$ operator norm of the entry-wise difference between the two adjacency matrices. Intuitively, $T_{p \rightarrow q}(D_1, D_2)$ 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 is a matrix analogue to 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 community as weighted by φ .

Not every $p \rightarrow q$ operator norm is computable or produces a test that has power against a nontrivial class of alternatives, which is why I focus on two specific choices of p and q . The first test statistic is based on the $2 \rightarrow 2$ operator norm (also known as the spectral norm or radius). The $2 \rightarrow 2$ norm is a natural first choice because it is straightforward to compute in $O(N^3)$ time via singular value decomposition and its statistical properties have been well studied in the random matrix theory literature. However, I show below that 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 may place excessive weight on the high-variance rows of $[\mathbb{1}\{D_1 \leq s\} - \mathbb{1}\{D_2 \leq s\}]$.

The second test statistic is based on the $\infty \rightarrow 1$ operator norm. 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 so is less sensitive to row-heteroskedasticity. However, computing this norm is NP-hard and so the proposed test is instead based on the semidefinite approximation

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

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 Alon and Naor (2006). This statistic can be computed in $O(N^{3.5})$ time using programs available in many statistical software packages. Its use is justified by $T_{\infty \rightarrow 1} \leq S_{\infty \rightarrow 1} \leq 2T_{\infty \rightarrow 1}$ for any D_1, D_2 , and N . See Appendix Section A.2.

4.2 Consistency

As discussed in Section 3, the α -sized tests based on $T_{2 \rightarrow 2}$ and $S_{\infty \rightarrow 1}$ (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 (as described in Section 2) and tests (as described in Section 3) indexed by $N \in \mathbb{N}$. The parameters F_1, F_2, R , and α may all vary with N subject to the restrictions below. Limits are with $N \rightarrow \infty$.

4.2.1 Assumptions and constructions

The following assumptions on the size of the test are imposed.

Assumptions:

$$-\ln(\alpha) = O(\ln(N)) \text{ and } \alpha R \geq 2. \quad \square$$

I do not believe them to be restrictive in practice. The first is that the size of the test is not exponentially small relative to the number of agents. The second is that the size of the test is larger than twice the inverse of the number of simulations. They are required because if α is too small, the test will mechanically fail to reject H_0 regardless of the choice of T or difference between F_1 and F_2 . They follow when α is fixed and R tends to infinity with N .

The following constructions are used.

Constructions:

$$\begin{aligned}\Delta(s) &= \mathbb{1}\{D_1 \leq s\} - \mathbb{1}\{D_2 \leq s\} \\ \nu_{ij}(s) &= F_{ij,1}(s) + F_{ij,2}(s) - 2F_{ij,1}(s)F_{ij,2}(s) \\ \tau &= \max_{s \in \mathbb{R}} \max_{i \in [N]} \sqrt{\sum_{j \in [N]} \nu_{ij}(s)} \\ \sigma &= \max_{s \in \mathbb{R}} \sum_{i \in [N]} \sqrt{\sum_{j \in [N]} \nu_{ij}(s)} \\ T_{2 \rightarrow 2}(F_1, F_2) &= \max_{s \in \mathbb{R}} \max_{\varphi: \|\varphi\|_2=1} \|(F_1(s) - F_2(s)) \varphi\|_2 \\ T_{\infty \rightarrow 1}(F_1, F_2) &= \max_{s \in \mathbb{R}} \max_{\varphi: \|\varphi\|_\infty=1} \|(F_1(s) - F_2(s)) \varphi\|_1. \quad \square\end{aligned}$$

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 i th 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 . $T_{2 \rightarrow 2}(F_1, F_2)$ and $T_{\infty \rightarrow 1}(F_1, F_2)$ are the test statistics applied to the (matrix of) distribution functions F_1 and F_2 . They quantify the extent to which H_0 is violated with larger values corresponding to more extreme violations. Under certain regularity conditions, $T_{2 \rightarrow 2}(F_1, F_2)$ large relative to τ or $T_{\infty \rightarrow 1}(F_1, F_2)$ large relative to σ eventually results in a rejection of H_0 . See Theorems 1 and 2 below.

4.2.2 Results

The main consistency results are given by Theorems 1 and 2 below. Additional results about the power properties of the tests can be found in Online Appendix Section B.

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

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

tends to one uniformly over all alternatives satisfying $T_{2 \rightarrow 2}(F_1, F_2)/\tau \rightarrow \infty$ and $\tau/\sqrt{\ln(N)} \rightarrow \infty$. \square

Theorem 1 has two rate conditions. The first rate condition is that $T_{2 \rightarrow 2}(F_1, F_2)/\tau \rightarrow \infty$. It implies that the size of the violation of H_0 (as given by $T_{2 \rightarrow 2}(F_1, F_2)$) exceeds the magnitude of the test statistic under H_0 (which is on the order of τ , see below). When F_t is sufficiently dense in the sense that $F_{ij,t}(s)$ is uniformly bounded away from 0 and 1 for some $s \in \mathbb{R}$, it follows from $\sqrt{N}T_{2 \rightarrow 2}(F_1, F_2) \rightarrow \infty$.

The second rate condition is that $\tau/\sqrt{\ln(N)} \rightarrow \infty$. It implies that the reference distribution generated by $\{T_{2 \rightarrow 2}(D_1^r, D_2^r)\}_{r \in [R]}$ concentrates below τ . It is satisfied if F_t is sufficiently dense in the sense that $\frac{N}{\ln(N)}F_{ij,t}(s)$ and $\frac{N}{\ln(N)}(1 - F_{ij,t}(s))$ are uniformly bounded away from 0 for some $s \in \mathbb{R}$. For unweighted and undirected networks, this second condition requires agents to have in expectation at least $\ln(N)$ connections.

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

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

tends to one uniformly over all alternatives satisfying $T_{\infty \rightarrow 1}(F_1, F_2)/\sigma \rightarrow \infty$ and $\sigma/\sqrt{\ln(N)} \rightarrow \infty$. \square

Theorem 2 also has two rate conditions. They admit a similar interpretation to those in Theorem 1, but suggest different properties of F_1 and F_2 . The first rate condition $T_{\infty \rightarrow 1}(F_1, F_2)/\sigma \rightarrow \infty$ follows if $F_{ij,t}(s)$ is uniformly bounded away from 0 and 1 for some $s \in \mathbb{R}$ and $\sqrt{N}T_{\infty \rightarrow 1}(F_1, F_2) \rightarrow \infty$. The second rate condition $\sigma/\sqrt{\ln(N)} \rightarrow \infty$ follows if $\frac{N^3}{\ln(N)}F_{ij,t}(s)$ and $\frac{N^3}{\ln(N)}(1 - F_{ij,t}(s))$ are uniformly bounded away from 0 for some $s \in \mathbb{R}$. For unweighted and undirected networks, this second condition only requires agents to have in expectation at least $\ln(N)/N^2$ connections.

These results predict two scenarios in which the test based on the $\infty \rightarrow 1$ norm is potentially more powerful than that based on the $2 \rightarrow 2$ norm, in the sense that the rate conditions of Theorem 2 but not Theorem 1 are satisfied. The first scenario is network sparsity. For example, if $NF_{ij,t}(s)$ and $N(1 - F_{ij,t}(s))$ are uniformly bounded away from 0 for some $s \in \mathbb{R}$ and $t \in [2]$, then $\tau/\sqrt{\ln(N)} \rightarrow 0$ but $\sigma/\sqrt{\ln(N)} \rightarrow \infty$. This suggests that the test based on the $\infty \rightarrow 1$ norm may be better suited for networks in which agents have

in expectation only a few connections. The second scenario is degree-heterogeneity. For example, if $F_{ij,1} = F_{ij,2}$ is uniformly bounded away from 0 when $i \wedge j \leq K$ and $\sqrt{N}F_{ij,1} \neq \sqrt{N}F_{ij,2}$ is uniformly bounded from above and away from 0 when $i \wedge j > K$ for some fixed positive integer K then $T_{2 \rightarrow 2}(F_1, F_2)/\tau \rightarrow 0$ but $T_{\infty \rightarrow 1}(F_1, F_2)/\sigma \rightarrow \infty$. This suggests that the test based on the $\infty \rightarrow 1$ norm may be better suited for networks in which the differences between F_1 and F_2 do not necessarily occur in the high-variance rows.² Corroborating simulation evidence can be found in Online Appendix Section C.

5 Two empirical demonstrations

I provide two empirical demonstrations using publicly available data.³ In both settings the randomization test based on the $\infty \rightarrow 1$ norm is sufficiently powerful to detect the relevant difference in network structure. Alternatives are less reliable.

The first is a test of network stationarity as described in Application 1 of Section 2.3. A sample of high school students is surveyed annually about their social connections. The networks appear to be less connected and more clustered over time, potentially because the students place increasing value on having friends in common as they age. The problem is to test whether the observed changes in reported relationships are statistically significant.

The data comes from the “Teenage Friends and Lifestyle Study” (see Michell and West 1996) in which 160 Scottish students are surveyed about friendship links during their second through fourth years of secondary school.⁴ The demonstration compares the networks reported during the first and third years when the students are respectively 13 and 15 years old. Only students who appear in all years are included, yielding a sample size of $N = 129$.

Panel A of Table 1 describes the means and standard deviations of four network statistics: 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 connected component. The table indicates that while the total number of links appear to be roughly the same for both networks, the

²Alternatively, the test based on the $2 \rightarrow 2$ norm is potentially more powerful when the differences between F_1 and F_2 occur in a small number of high-variance rows.

³R code and data can be found on my website: <https://sites.google.com/site/ericjauerbach/>

⁴The data can be found at https://www.stats.ox.ac.uk/~snijders/siena/Glasgow_data.htm.

second network is less connected and more clustered than the first. Table 2 demonstrates that these differences are unlikely to be generated by the same random graph model.

Panel A of Table 2 reports the p-values of the randomization test described in Section 3 using 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 $T_{2 \rightarrow 2}$ and $S_{\infty \rightarrow 1}$ as described in Section 4.1. Panel A indicates that the differences in the clustering and diameter statistics are unlikely to be generated by the same random graph model. The implausibility of the null hypothesis is also clearly indicated by the tests based on $T_{2 \rightarrow 2}$ and $S_{\infty \rightarrow 1}$.

An alternative way to detect differences in network structure is a regression-based approach where one computes a vector of network statistics for both networks, specifies a regression model in which the network statistics depend on a constant, an indicator for one of the networks, and an idiosyncratic error, and conducts a Wald test (see for example Banerjee et al. 2018, Section 3).⁵ Panel A of Table 3 reports the OLS point estimates and p-values for three network statistics on the left-hand side: agent degree, eigenvector centrality, and agent clustering $\left\{ \frac{\sum_{j,k \in [N]} D_{ij} D_{ik} D_{jk}}{\sum_{j,k \in [N]} D_{ij} D_{ik}} \right\}_{i \in [N]}$. In all three regressions, the coefficient in front of the network indicator is not statistically significant. The regression-based tests do not detect the change in structure of the Glasgow networks.

The second demonstration is a test for link heterogeneity as described in Application 2 in Section 2.3. Households in a village are surveyed about multiple types of relationships. Different survey questions appear to reveal information about different types of connections between agents. The problem is to test whether the observed differences in the network structure induced by the different survey questions are statistically significant.

The data comes from Banerjee et al. (2013), who survey information about a dozen social and economic connections between households for 75 villages in rural India.⁶ This demonstration uses data from the 77 households in village 10 and compares the social network

⁵I thank an anonymous referee for suggesting the comparison. Formally, if S_{it} is a network statistic associated with agent i in network t such as i 's degree or eigenvector centrality, then the test is based on the linear model $S_{it} = \alpha + \beta \mathbb{1}_{t=2} + \varepsilon_{it}$. It is assumed that the errors $\{\varepsilon_{it}\}_{i \in [N], t \in [2]}$ are independent and identically distributed with normal marginals. The null hypothesis is $\beta = 0$.

⁶The data can be found at <https://hdl.handle.net/1902.1/21538>.

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.

Panel B of Table 1 describes the same statistics as Panel A, but for this second demonstration. It shows two main differences between the social and economic networks. The first difference is that households have on average one more economic link. The second difference is that there is more clustering in the economic network. Table 2 demonstrates that these differences are unlikely to be generated by the same random graph model.

Panel B of Table 2 describes the same tests as Panel A, but for this second demonstration. 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 $2 \rightarrow 2$ norm because $T_{2 \rightarrow 2}(D_1, D_2)$ is in the third quartile of its reference distribution. On the other hand, $S_{\infty \rightarrow 1}$ is firmly in the upper decile of its reference distribution, and so this test statistic provides evidence against the null hypothesis.

Panel B of Table 3 describes the same regressions as Panel A, but for this second application. The results from the tests based on agent degree and clustering also indicate a difference in the structure of the social and economic networks. The strength of this evidence, however, depends crucially on the regression model assumptions (linearity, normality, homoskedasticity, etc.) and choice of network statistic.

6 Conclusion

This paper considers a two-sample testing problem where the null hypothesis is that two networks are drawn from the same random graph model. It proposes tests based on the magnitude of the difference between the networks’ adjacency matrices as measured by the $2 \rightarrow 2$ and $\infty \rightarrow 1$ operator norms. Both tests with enough power can detect any fixed difference between the models. However, the test based on the $\infty \rightarrow 1$ norm can be much more

powerful for the kinds of sparse and degree-heterogeneous networks common in economics.

6.1 Example extensions

The above framework can be extended in many ways. I describe three examples. Details can be found in Online Appendix Section D.

6.1.1 Extension 1: a completely randomized experiment

Banerjee et al. (2018) collect data on connections between villagers in a sample of villages before and after a microfinance agency offers loans to villagers in some but not all of the villages. They find that the program disincentivizes the formation of certain types of connections. The above framework can be extended to test whether the observed differences between the treatment and control villages are statistically significant. Specifically, the researcher can first measure the magnitude of the change in each village’s network before and after treatment assignment using, for example, the operator norm based statistics from Section 4. The test statistic can then be the square of the average measurement for the treated villages minus the average measurement for the control villages. A reference distribution can be constructed by permuting the treatment status of the villages.

6.1.2 Extension 2: a one-sample test of independence

Fafchamps and Gubert (2007) study a risk-sharing network and argue that the surveyed links are not related to the respondents’ occupations. The above framework can be extended to test whether the network connections and respondent occupations are statistically independent. Specifically, let D_{ij} indicate whether agents i and j have a risk-sharing connection and W_{ij} denote whether agents i and j have the same occupation. The test statistic can then be the magnitude of the Hadamard product of D and W as measured by the $2 \rightarrow 2$ or semidefinite approximation to the $\infty \rightarrow 1$ operator norm. A reference distribution can be constructed by permuting the villager occupations.

Table 1: Descriptive Statistics

Panel A: Glasgow Networks		Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter	
First Wave		Mean	3.48	0.09	0.36	14.00
		SD	1.63	0.20	0.00	0.00
Third Wave		Mean	3.60	0.07	0.42	20.00
		SD	1.75	0.21	0.00	0.00
Panel B: India Networks		Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter	
Social Network		Mean	3.92	0.26	0.13	6.00
		SD	3.17	0.22	0.00	0.00
Economic Network		Mean	4.94	0.27	0.19	5.00
		SD	3.65	0.21	0.00	0.00

Panel A compares two networks from the “Teenage Friends and Lifestyle Study.” Panel B compares two networks from village 10 in Banerjee et al. (2013). Both panels describe 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.

6.1.3 Extension 3: a one-sample specification test

Jackson and Rogers (2007) argue that real-world social networks have features that are not explained by an Erdős-Renyi model of link formation. More broadly, the above framework can be extended to test the hypothesis that the network data is drawn from a given random graph model. Specifically, the researcher can first simulate network data from the random graph model and use the magnitude of the difference between the observed network data and the simulated network using, for example, the operator norm based statistics from Section 4 as a test statistic. A reference distribution can then be constructed by recomputing the test statistic on additional pairs of networks drawn from the random graph model.

Table 2: Randomization Tests

Panel A: Glasgow Networks	Average Degree	Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter	$2 \rightarrow 2$ Norm	$\infty \rightarrow 1$ Norm
p-value	0.59	0.96	0.20	0.02	0.08	0.01	0.00
Panel B: India Networks	Average Degree	Agent Degree	Eigenvector Centrality	Clustering Coefficient	Diameter	$2 \rightarrow 2$ Norm	$\infty \rightarrow 1$ Norm
p-value	0.00	0.58	0.73	0.05	0.55	0.25	0.05

Panel A compares two networks from the “Teenage Friends and Lifestyle Study.” Panel B compares two networks from village 10 in Banerjee et al. (2013). Both panels describe p-values for randomization 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, $2 \rightarrow 2$ norm of the entry-wise differences between the two networks’ adjacency matrices, and semidefinite approximation to the $\infty \rightarrow 1$ norm of the entry-wise differences between the two networks’ adjacency matrices. For all tests, $R = 10,000$.

Table 3: Regressions

Panel A: Glasgow Networks	Agent Degree	Eigenvector Centrality	Agent Clustering
intercept	3.48 (0.00)	0.09 (0.00)	0.21 (0.00)
coefficient	0.12 (0.58)	-0.02 (0.34)	-0.03 (0.14)
Panel B: India Networks	Agent Degree	Eigenvector Centrality	Agent Clustering
intercept	3.92 (0.00)	0.26 (0.00)	0.03 (0.00)
coefficient	1.01 (0.07)	0.01 (0.76)	0.02 (0.01)

Panel A compares two networks from the “Teenage Friends and Lifestyle Study.” Panel B compares two networks from village 10 in Banerjee et al. (2013). Both panels describe point estimates and p-values (in parentheses) for the OLS regression of agent-level network centrality measures agent degree, eigenvector centrality, and agent clustering on a constant and an indicator for the second network.

References

- Alon, N. and A. Naor (2006). Approximating the cut-norm via grothendieck’s inequality. *SIAM Journal on Computing* 35(4), 787–803.
- 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.
- 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., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science* 341(6144), 1236498.
- Banerjee, A. V., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2018). Changes in social network structure in response to exposure to formal credit markets. *Available at SSRN 3245656*.
- Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.
- 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.
- Fafchamps, M. and F. Gubert (2007). Risk sharing and network formation. *American Economic Review* 97(2), 75–79.
- Ghoshdastidar, D., M. Gutzeit, A. Carpentier, and U. von Luxburg (2017). Two-sample hypothesis testing for inhomogeneous random graphs. *arXiv preprint arXiv:1707.00833*.
- Gittens, A. and J. A. Tropp (2009). Error bounds for random matrix approximation schemes. *arXiv preprint arXiv:0911.4108*.

- 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.
- 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. (2019). Network data. *Forthcoming Econometrics Handbook Chapter*.
- Jackson, M. O. and B. W. Rogers (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review* 97(3), 890–915.
- Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. *Journal of Economic Literature* 55(1), 49–95.
- 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.
- Michell, L. and P. West (1996). Peer pressure to smoke: the meaning depends on the method. *Health education research* 11(1), 39–49.
- Nielsen, A. M. and D. Witten (2018). The multiple random dot product graph model. *arXiv preprint arXiv:1811.12172*.
- Pelican, A. and B. S. Graham (2020). An optimal test for strategic interaction in social and economic network formation between heterogeneous agents. Technical report, National Bureau of Economic Research.
- Rose, A. K. (2004). Do we really know that the wto increases trade? *American economic review* 94(1), 98–114.
- 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.

A Proofs

A.1 Theorem 1

Lemma 1: Let $\{X_s\}_{s \in [S]}$ be an arbitrary finite collection of $N \times N$ random symmetric matrices with independent and mean-zero entries absolutely bounded by 1 above the diagonal and zeros on the main diagonal. Then for any fixed $\alpha \in [0, 1]$ and $\gamma \in [0, 1/2]$

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

with probability at least $1 - \alpha$. \square

Proof of Lemma 1: The lower bound holds for any collection of matrices. It 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 $N \times N$ 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 ij th 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, for any $\varepsilon > 0$, Talagrand's inequality implies

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

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

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

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

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

with probability less than $S \exp(-\varepsilon^2/2)$ by the union bound. The claim follows by setting $\varepsilon = \sqrt{-2 \ln(\frac{\alpha}{S})}$. \square

Proof of Theorem 1: The main part of the proof is to show that for any $r \in [R]$,

$T_{2 \rightarrow 2}(D_1, D_2) \geq T_{2 \rightarrow 2}(D_1^r, D_2^r)$ with high probability. The first step is to bound

$T_{2 \rightarrow 2}(D_1^r, D_2^r)$ from above. Lemma 1 implies that for any $\beta \in [0, 1]$ and $\gamma \in [0, 1/2]$

$$T_{2 \rightarrow 2}(D_1^r, D_2^r) \leq \sqrt{-2 \ln \left(\frac{\beta}{N(N-1)} \right)} + (1 + \gamma) 2 \max_{s \in [R]} \max_{j \in [N]} \sqrt{\sum_{i \in [N]} \nu_{ij}(s)} + \frac{6(1 + \gamma)}{\sqrt{\ln(1 + \gamma)}} \sqrt{\ln(N)}$$

with probability greater than $1 - \beta$. Since $-\ln(\alpha)/\ln(N) = O(1)$ by assumption in Section 4.2.1, it is possible to choose β such that $-\ln(\beta)/\ln(N) = O(1)$ and $\beta = o(\alpha)$. Along with the rate condition $\tau/\sqrt{\ln(N)} \rightarrow \infty$, this implies $\left(\sqrt{-\ln(\beta) + \ln(N)} \right) / \tau \rightarrow 0$ and so $T_{2 \rightarrow 2}(D_1^r, D_2^r)/\tau \leq 3$ with probability greater than $1 - \beta$ eventually.

The second step is to bound $T_{2 \rightarrow 2}(D_1, D_2)$ from below. By the triangle inequality,

$$T_{2 \rightarrow 2}(D_1, D_2) \geq T_{2 \rightarrow 2}(F_1, F_2) - |T_{2 \rightarrow 2}(D_1, F_1) + T_{2 \rightarrow 2}(D_2, F_2)|$$

where $T_{2 \rightarrow 2}(D_1, F_1) := \max_{s \in \mathbb{R}} \max_{\varphi \in \mathcal{S}^N} \sqrt{\sum_{i \in [N]} \left(\sum_{j \in [N]} (\mathbb{1}_{D_{ij,1} \leq s} - F_{ij,1}(s)) \varphi_j \right)^2}$. Lemma 1 implies $T_{2 \rightarrow 2}(D_t, F_t)/\tau \leq 3$ for $t \in [2]$ with probability greater than $(1 - 2\beta)$ eventually.

The rate condition $T_{2 \rightarrow 2}(F_1, F_2)/\tau \rightarrow \infty$ then implies $T_{2 \rightarrow 2}(F_1, F_2)/\tau \rightarrow \infty$, and so $T_{2 \rightarrow 2}(D_1, D_2)/\tau \rightarrow \infty$. This demonstrates that $T_{2 \rightarrow 2}(D_1, D_2) \geq T_{2 \rightarrow 2}(D_1^r, D_2^r)$ with probability greater than $1 - 3\beta$ for any $r \in [R]$.

To finish the proof, write

$$\begin{aligned} & P \left((R+1)^{-1} \left(1 + \sum_{r \in [R]} \mathbb{1}\{T_{2 \rightarrow 2}(D_1^r, D_2^r) \geq T_{2 \rightarrow 2}(D_1, D_2)\} \right) \leq \alpha \right) \\ &= P \left(R^{-1} \sum_{r \in [R]} \mathbb{1}\{T_{2 \rightarrow 2}(D_1^r, D_2^r) \geq T_{2 \rightarrow 2}(D_1, D_2)\} \leq \frac{\alpha(R+1) - 1}{R} \right). \end{aligned}$$

On the right-hand side, $\frac{\alpha(R+1)-1}{R} \geq \alpha/2$ since $\alpha R \geq 2$ by assumption in Section 4.2.1. On the left-hand side, the entries of $\{\mathbb{1}\{T_{2 \rightarrow 2}(D_1^r, D_2^r) \geq T_{2 \rightarrow 2}(D_1, D_2)\}\}_{r \in [R]}$ are independent and $o_p(\alpha)$ by choice of β , and so their average is $o_p(\alpha)$. The claim follows. \square

A.2 Theorem 2

The proof of Theorem 2 relies on the following inequality due to Grothendieck and Krivine (1979), see generally Alon and Naor (2006).

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

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

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

an arbitrary Hilbert space and $\|\cdot\|_{\mathcal{H}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ are the associated norm and inner product operators. \square

Bounds from Section 4.1: $T_{\infty \rightarrow 1} \leq S_{\infty \rightarrow 1} \leq \frac{\pi}{2 \ln(1+\sqrt{2})} T_{\infty \rightarrow 1} \leq 1.783 T_{\infty \rightarrow 1}$.

Proof of claim: The first inequality follows from

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

where \otimes refers to the vector outer product operator, the first equality follows from taking $\psi = \text{sign}(\Delta(s)\varphi)$, 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 directly implies the second inequality. \square

Lemma 2: Let $\{X_s\}_{s \in [S]}$ be an arbitrary finite collection of $N \times N$ random symmetric matrices with independent and mean-zero entries absolutely bounded by 1 above the diagonal and zeros on the main diagonal. Then for any fixed $\alpha \in [0, 1]$

$$\frac{1}{K} \max_{s \in [S]} \sum_{i \in [N]} \sqrt{\sum_{j \in [N]} X_{ij,s}^2} \leq \max_{s \in [S]} \max_{\varphi \in \mathcal{C}^N} \sum_{i \in [N]} \left| \sum_{j \in [N]} X_{ij,s} \varphi_j \right| \leq \sqrt{-2 \ln \left(\frac{\alpha}{S} \right)} + 4 \max_{s \in [S]} \sum_{i \in [N]} \sqrt{\sum_{j \in [N]} E[X_{ij,s}^2]}$$

with probability at least $1 - \alpha$ where $K = 1.783$ is Grothendieck's constant. \square

Proof of Lemma 2: The lower bound holds for any collection of matrices. It 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 $N \times N$ matrices indexed by a set S then

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

is the ij th 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 $N \times N$ 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

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

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

$$E \left[\max_{\varphi \in \mathcal{C}^N} \sum_{i \in [N]} \left| \sum_{j \in [N]} X_{ij} \varphi_j \right| \right] \leq 4 \sum_{i \in [N]} \sqrt{\sum_{j \in [N]} E[X_{ij}^2]}.$$

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

$$\max_{s \in [S]} \max_{\varphi \in \mathcal{C}^N} \sum_{i \in [N]} \left| \sum_{j \in [N]} X_{ij,s} \varphi_j \right| > \varepsilon + 4 \max_{s \in [S]} \sum_{i \in [N]} \sqrt{\sum_{j \in [N]} E[X_{ij,s}^2]}$$

with probability less than $S \exp(-\varepsilon^2/2)$ by the union bound. The claim follows by setting $\varepsilon = \sqrt{-2 \ln(\frac{\alpha}{S})}$. \square

Proof of Theorem 2: The main part of the proof is to show that for any $r \in [R]$, $S_{\infty \rightarrow 1}(D_1, D_2) \geq S_{\infty \rightarrow 1}(D_1^r, D_2^r)$ with high probability. The first step is to bound $S_{\infty \rightarrow 1}(D_1^r, D_2^r) \leq KT_{\infty \rightarrow 1}(D_1^r, D_2^r)$ from above. Lemma 2 implies that for any $\beta \in [0, 1]$

$$T_{\infty \rightarrow 1}(D_1^r, D_2^r) \leq \sqrt{-2 \ln \left(\frac{\beta}{N(N-1)} \right)} + 4 \max_{s \in \mathbb{R}} \sum_{i \in [N]} \sqrt{\sum_{j \in [N]} \nu_{ij}(s)}$$

with probability greater than $1 - \beta$. Since $-\ln(\alpha)/\ln(N) = O(1)$ by assumption in Section 4.2.1, it is possible to choose β such that $-\ln(\beta)/\ln(N) = O(1)$ and $\beta = o(\alpha)$. Along with the rate condition $\sigma/\sqrt{\ln(N)} \rightarrow \infty$, this implies $\left(\sqrt{-\ln(\beta) + \ln(N)}\right)/\sigma \rightarrow 0$ and so $T_{\infty \rightarrow 1}(D_1^r, D_2^r)/\sigma \leq 3$ with probability greater than $1 - \beta$ eventually.

The second step is to bound $S_{\infty \rightarrow 1}(D_1, D_2) \geq T_{\infty \rightarrow 1}(D_1, D_2)$ from below. By the triangle inequality,

$$T_{\infty \rightarrow 1}(D_1, D_2) \geq T_{\infty \rightarrow 1}(F_1, F_2) - |T_{\infty \rightarrow 1}(D_1, F_1) + T_{\infty \rightarrow 1}(D_2, F_2)|$$

where $T_{\infty \rightarrow 1}(D_1, F_1) := \max_{s \in \mathbb{R}} \max_{\varphi \in \mathcal{C}^N} \sum_{i \in [N]} \left| \sum_{j \in [N]} (\mathbb{1}_{D_{ij,1} \leq s} - F_{ij,1}(s)) \varphi_j \right|$. Lemma 2 implies $T_{\infty \rightarrow 1}(D_t, F_t)/\sigma \leq 3$ for $t \in [2]$ and with probability greater than $(1 - 2\beta)$ eventually. The rate condition $T_{\infty \rightarrow 1}(F_1, F_2)/\sigma \rightarrow \infty$ implies $T_{\infty \rightarrow 1}(F_1, F_2)/\sigma \rightarrow \infty$ and so $S_{\infty \rightarrow 1}(D_1, D_2)/\sigma \rightarrow \infty$. This demonstrates that $S_{\infty \rightarrow 1}(D_1, D_2) \geq S_{\infty \rightarrow 1}(D_1^r, D_2^r)$ with probability greater than $1 - 3\beta$ eventually for any $r \in [R]$.

To finish the proof, write

$$\begin{aligned} & P \left((R+1)^{-1} \left(1 + \sum_{r \in [R]} \mathbb{1}\{S_{\infty \rightarrow 1}(D_1^r, D_2^r) \geq S_{\infty \rightarrow 1}(D_1, D_2)\} \right) \leq \alpha \right) \\ &= P \left(R^{-1} \sum_{r \in [R]} \mathbb{1}\{S_{\infty \rightarrow 1}(D_1^r, D_2^r) \geq S_{\infty \rightarrow 1}(D_1, D_2)\} \leq \frac{\alpha(R+1) - 1}{R} \right). \end{aligned}$$

On the right-hand side $\frac{\alpha(R+1)-1}{R} \geq \alpha/2$ by assumption in Section 4.2.1. On the left-hand side, the entries of $\{\mathbb{1}\{S_{\infty \rightarrow 1}(D_1^r, D_2^r) \geq S_{\infty \rightarrow 1}(D_1, D_2)\}\}_{r \in [R]}$ are independent and $o_p(\alpha)$ by choice of β , and so their average is $o_p(\alpha)$. The claim follows. \square