An optimal test for strategic interaction in social and economic network formation between heterogeneous agents

Andrin Pelican and Bryan S. Graham *

August 31, 2020

*Pelican: e-mail: [pelicanandrin@gmail.com]. Graham: Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880 and National Bureau of Economic Research, e-mail: [bgraham@econ.berkeley.edu], web: http://bryangraham.github.io/econometrics/. We thank seminar participants at the University of California - Berkeley, the University of Cambridge, ITAM, Jinan University, Yale University, Microsoft Research and the World Congress of the Econometric Society for feedback and suggestions. We also thank Kaïla A. Munro, Daniele Ballinari, Gabriel Okasa, Enrico De Giorgi, Piotr Lukaszuk and Yi Zhang for their generous feedback and input. A special thanks to Winfried Hochstättler for continuously pointing out the connection to the discrete mathematics literature and to Michael Jannson for help and insight into the nature of our testing problem. All the usual disclaimers apply. Financial support from NSF grant SES #1357499 is gratefully acknowledged by the second author. This paper is a revised version of the co-authored chapter "Testing Strategic Interaction in Networks" which appeared in Pelican's 2019 doctoral dissertation at the University of St. Gallen.

Abstract

We introduce a test for whether agents' preferences over network structure are *interdependent*. Interdependent preferences induce strategic behavior since the optimal set of links directed by agent i will vary with the configuration of links directed by other agents.

Our model also incorporates agent-specific in- and out-degree heterogeneity and homophily on observable agent attributes. This introduces $2N + K^2$ nuisance parameters (N is number of agents in the network and K the number of possible agent attribute configurations).

Under the null equilibrium is unique, but our hypothesis is nevertheless a composite one as the degree heterogeneity and homophily nuisance parameters may range freely across their parameter space. Under the alternative our model is incomplete; there may be multiple equilibrium network configurations and our test is agnostic about which one is selected.

Motivated by size control, and exploiting the exponential family structure of our model *under the null*, we restrict ourselves to conditional tests. We characterize the exact null distribution of a family of conditional tests and introduce a novel Markov Chain Monte Carlo (MCMC) algorithm for simulating this distribution.

We also characterize the locally best test. The form of this test depends upon the gradient of the likelihood with respect to the strategic interaction parameter in the neighborhood of the null. Remarkably, this gradient, and consequently the form of the locally best test statistic, does not depend on how an equilibrium is selected. Exploiting this lack of dependence, we outline a feasible version of the locally best test.

We present two illustrative applications. First, we test for whether nations behave strategically when choosing locations for overseas diplomatic missions. Second, we test for whether firms prefer to sell to firms with richer customer bases (i.e., whether firms value "indirect customers"). Some Monte Carlo experiments explore the size and power properties of our test in practice.

JEL Codes: C31

Keywords: *Network formation, Locally Best Tests, Similar Tests, Exponential Family, Incomplete Models, Degree Heterogeneity, Homophily, Binary Matrix Simulation, Edge Switching Algorithms*

Network data feature in many areas of economic research. Examples include buyer-supplier networks or supply-chains (e.g., Atalay et al., 2011), research and development (R&D) and other types of strategic partnerships across firms (e.g., König et al., 2019), patterns of trade among nations (e.g., Tinbergen, 1962), the structure of friendships between adolescents (e.g., Calvó-Armengol et al., 2009), and interbank lending and borrowing (e.g., Boss et al., 2004). Jackson et al. (2017) present many other examples. Such data abound in the other social sciences as well (e.g., Apicella et al., 2012).

One approach to modelling networks proceeds pairwise, or dyad-by-dyad. In this approach (the realization of) each possible link in a network is independent of all others. Importantly this independence may only hold conditional on latent agent-specific attributes; such latent attributes may induce dependence across links unconditionally. Gravity models of trade, with exporter and importer fixed effects, provide a familiar illustration (Anderson, 2011). Stochastic block models (SBMs), widely studied in statistics, also fall into this category (Airoldi et al., 2008; Bickel et al., 2013; Gao et al., 2015).

A second approach views a network as an equilibrium outcome of a large N-player game. In this approach agents' preferences over links may vary with the presence or absence of links elsewhere in the network. For example agents' may prefer reciprocated to unreciprocated links. Alternatively they may attach extra utility to links which induce transitive closure (Granovetter, 1973). In such settings small, local, re-wirings of a network may induce a cascade of additional link updates which can, at least in principle, change the global topology of a network. Multiple equilibria may also arise. In strategic models, stable networks need not be efficient as agents fail to account for the costs and benefits of links they form on others. The two classes of network formation models, in addition to being scientifically distinct, generate different policy implications (Goyal, 2009).

Graham (2017) and de Paula et al. (2018) represent two recent attempts to actualize, respectively the dyad-by-dyad and strategic approaches, into workable econometric models.[1] In this paper we take a first step toward integrating these two econometric modelling approaches. We study a model of network formation which simultaneously incorporates rich agent-level unobserved heterogeneity, homophily, *and* interdependent preferences. We are aware of no prior attempt to incorporate these three features into a single econometric model. Incorporating heterogeneity and homophily into the null model is important because these factors provide alternative explanations for the types of network microstructure often associated with strategic behavior.

In our model the importance of preference interdependencies is indexed by a parameter (or vector of parameters). Our goal is to test whether this parameter equals zero. Testing the

[1]Graham (2020) surveys the larger econometric literature on network formation.

null of no strategic interaction in an interesting baseline model is a natural point of departure for empirical work. In some settings (e.g., international trade, friendship formation), strategic behavior may be plausibly second order. In such situations our test provides researchers with a useful specification check. When strategic interaction is suspected to be central, a rejection of our test would confirm these priors.

Our null model can match any network in- and out-degree sequence, as well as rich patterns of homophilous linking. Heavy-tailed degree distributions characterize many real world networks, as does homophily (e.g., Barabási, 2016; McPherson et al., 2001). Degree heterogeneity and homophily also generate local network structures often associated with strategic behavior, such as reciprocity, transitivity and degree assortativity. Our test is able to discriminate between, for example, clustering due to homophilous sorting and that due to a structural taste for transitivity in relationships. Relatedly our null model is not a universal "strawman"; our test will not "always reject".

Admittedly, we are unable to offer any direct guidance as to how to proceed in the event of a rejection. While a small collection of papers outline methods of set identification for parameters in strategic network formation models (e.g., Sheng, 2014; Menzel, 2016; de Paula et al., 2018), no existing approach allows for unobserved agent-specific effects in addition to strategic behavior. It is possible that some of the ideas which appear below could be re-purposed for set identification; but such re-purposing is not direct. We leave this for future research.

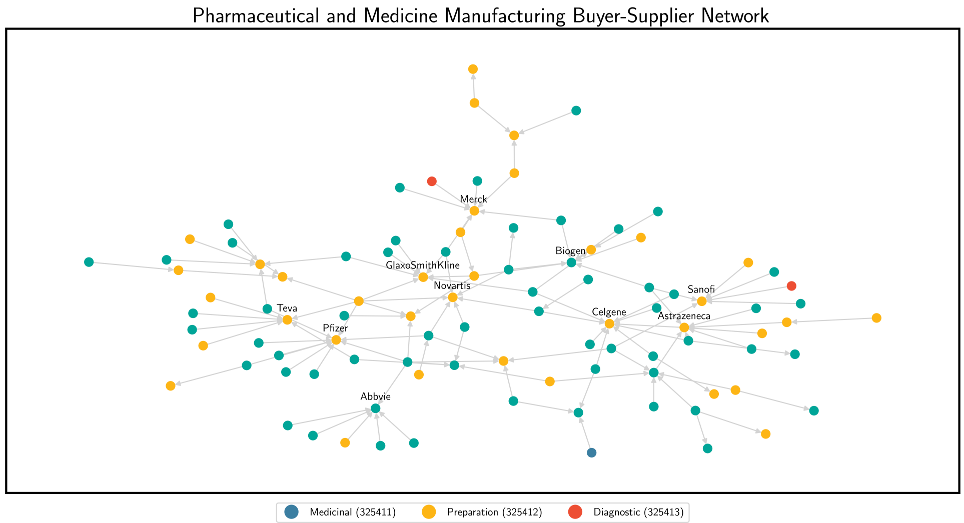
In constructing our test we faced several challenges.

1. Our nuisance parameter is high dimensional. Our null model includes $2N$ agent-specific (incidental) degree heterogeneity parameters as well as K^2 homophily coefficients (where K equals the number of observed agent types). Because these nuisance parameters can range freely across their parameter space under the null, avoiding size distortion is difficult. This problem famously arises in instrumental variables models, where the size properties of common tests may vary with instrument strength (c.f., Moreira, 2009; Andrews et al., 2019).

2. Our model is incomplete under the alternative (cf., de Paula, 2013). When preferences are inter-dependent multiple equilibrium networks may occur. We leave the mechanism which selects the observed equilibrium unspecified. Because the alternative is incomplete it is not obvious how to choose a test statistic with good power. A likelihood ratio test, for example, would require a complete specification of the equilibrium selection under the alternative.

3. We characterize the exact distribution of our test statistic under the null. Practical

Figure 1: Pharmaceutical Buyer-Supplier Network, 2015



Source: Compustat and authors' calculations.
Notes: 2015 buyer-supplier relationships among publicly traded firms in NAICS industry 3254 (pharmaceuticals). The head of each arc denotes the buying firm. Firms in each of the six-digit sub-sectors are shaded differently (see the legend). The largest weakly-connected component is shown.

application of this result, however, requires a feasible simulation algorithm.

Section 1 presents our model of strategic network formation. We begin by defining agent preferences and characterizing equilibrium networks. With this foundation we are able to write down a likelihood function for the network. Since there may exist multiple equilibrium networks, this likelihood depends on an unknown (and unmodelled) equilibrium selection mechanism. Although well-defined (see Theorem 1.1 below), our likelihood function cannot be numerically evaluated in practice.

Section 2 outlines our approach to testing and derives the form of the locally best test statistic. We characterize the exact distribution of our test statistic under the null. However, for reasons of practically, we approximate the exact null distribution by simulation. Section 3 outlines a new Markov Chain Monte Carlo (MCMC) algorithm for generating random draws from the required null distribution. Our algorithm may be of independent interest to those familiar with binary matrix simulation and counting problems arising in machine learning, ecology and other fields (e.g., Sinclair, 1993).

Section 4 presents two small applications of our test. First we test for whether nations behave strategically when choosing locations for their diplomatic missions. In particular, we focus on whether nations value transitivity in diplomatic ties. We might posit, for example, that the value of a diplomatic mission in the People's Republic China (PRC) increased for many countries after President Carter's decision to formally recognize the PRC in 1978 (cf., Kinne, 2014). If prior to 1978 many countries had diplomatic relations with the United States, but not the PRC, directing an arc to the PRC after the US did so would generate a transitive triad.

In a second application, we test for whether firms value indirect customers (i.e., do they prefer to sell to firms which themselves sell to many other firms). For this illustration we use three Buyer-Supplier networks that we constructed from Compustat data.[2] Specifically we look at the vehicle, computer and pharmaceutical manufacturing industries. Figure 1 plots the pharmaceutical and medicine buyer-supplier network.

Section 4 also reports on a small number of Monte Carlo experiments we conducted to verify the theoretical size and power properties of our test. Section 5 finishes with a short discussion of some possible areas for additional research.

While our focus is on strategic interaction in the context of a single network (with many agents), our results are also applicable to settings where the econometrician observes many independent games, each with a small number of players (e.g., market entry decisions by rival firms across many markets). Chen et al. (2018) and Kaido and Zhang (2019) are two recent examples of attempts to extend likelihood-ratio ideas to this type of setting. The test we introduce below is an analogous to a score-type test, complementing these likelihood-based approaches.

1 Model of network formation

Here we outline a model of strategic network formation. In this model N heterogenous agents form a directed network (or digraph). We begin by establishing some basic notation. We then introduce agent preferences over the form of the network, discuss equilibrium networks and, finally, develop a likelihood function for the observed network.

1.1 Notation

A directed graph $G(\mathcal{N}, \mathcal{A})$ consists of a set of nodes (agents) $\mathcal{N} = \{1, \ldots, N\}$ and a set of ordered pairs of nodes $\mathcal{A} = \{(i,j), (k,l), \ldots\}$ for $i \neq j$, $k \neq l$, and $i, j, k, l \in \mathcal{N}$. The elements of $\mathcal{A}$ correspond to those arcs, or directed links, present in $G(\mathcal{N}, \mathcal{A})$.

In what follows we typically work with the adjacency matrix $\mathbf{D} = [D_{ij}]$ where

$$D_{ij} = \begin{cases} 1 & \text{if } ij \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

[2] Atalay et al. (2011) constructed a similar network also using data from Compustat.

Since we rule out self-links, as might arise in, for example, citation networks, the diagonal of $\mathbf{D}$ consists of structural zeros.

Let $G-ij$ denote the network obtained by deleting link ij from G (if present), and $G+ij$ the network one gets after adding this link (if absent). Let $\mathbf{D}\pm ij$ denote the adjacency matrix associated with the network obtained by adding/deleting link ij from G. Let $\mathbb{D}_N$ denote the set of all $2^{N(N-1)}$ possible adjacency matrices and $\mathbb{I}_J$ the set of all possible J-dimensional binary vectors.

Associated with each agent in the network is the triple $(A_i, B_i, X_i')'$. Here A_i and B_i are, as explained further below, agent-level out- and in-degree heterogeneity terms *unobserved* by the econometrician. In contrast X_i is a $K \times 1$ vector of group membership indicators that is *observed* by the econometrician. These indicators might reflect the industrial classification of a firm, the gender or race of an individual, or the broad geographic location of a nation. More generally X_i enumerates the support points of a collection of (observed) discrete regressors.

We leave the joint distribution of $(A_i, B_i, X_i')'$ unrestricted.[3] This implies, for example, that the unobserved degree heterogeneity $(A_i, B_i)'$ may be correlated with the observed covariates X_i, as in fixed effects panel data analyses (see also Graham (2017), Dzemski (2018), Jochmans (2018) and Yan et al. (2018)).

1.2 Preferences

We assume that agents care about the shape of the network. Let $\mathbf{d} \in \mathbb{D}_N$ be a feasible N-player network. Agent utility varies with the configuration of this network. The utility agent i gets from some feasible network wiring $\mathbf{d}$ is assumed equal to

$$\nu_i(\mathbf{d}_i, \mathbf{d}_{-i}; \mathbf{U}) = \sum_j d_{ij}[A_i + B_j + X_i'\Lambda_0 X_i + \gamma_0 s_{ij}(\mathbf{d}) - U_{ij}], \tag{2}$$

Here $\mathbf{d}_i = (d_{i1}, \ldots, d_{ii-1}, d_{ii+1}, \ldots, d_{iN})'$ corresponds the set of links that agent i chooses to form (or not), while $\mathbf{d}_{-i}$ equals the links that the other $N-1$ agents in the network choose to form (or not). In the language of game theory, $\mathbf{d}_i$ corresponds to a pure strategy.

Agent i's utility varies with number and nature of those links she chooses to send, or direct, towards others. The utility associated with i directing a link to j is increasing in the heterogeneity terms A_i and B_j. Agents with high values of out-degree heterogeneity A_i get a large amount of baseline utility from any link they send. In a social network context high A_i agents are "extroverts". High B_j agents, in contrast, are especially attractive targets for links sent by others. In a social network high B_j agents are "prestigious".

[3] This distribution does have implications for test power, as will become apparent below.

In a buyer-supplier context high A_i firms might especially value a diverse customer base or supply a "critical" input used in the production processes of many other firms. High B_j firms correspond to especially attractive customers. For example, national big box retail chains with many retail locations, like Walmart and Target, may have high B_j values since their purchases are less sensitive to local economic shocks.

The $X_i'\Lambda_0 X_j \overset{def}{\equiv} W_{ij}'\lambda_0$ term allows for assortative matching on agent attributes.[4] The elements of the $K \times K$ matrix $\Lambda = [\lambda_{kl}]$ parameterize the systematic utility generated by links, say, from group k to group l. This allows, for example, the utility generated by links across agents belonging to different groups to systematically differ from that generated by within-group links. In the buyer-supply context, arcs between firms with particular industrial classifications may generate greater surplus. In a social network girls might, all things equal, prefer other girls as friends. The Λ_0 matrix parameterizes homophily (or heterophily) of these types.

Network generating processes where link utility varies with agent-level degree heterogeneity and observable dyad attributes – the first three terms in (2) – can successfully match many features of real world networks (Graham, 2017). The third term in (2) – $s_{ij}(\mathbf{d})$ – enriches this baseline model to allow agent preferences over links to vary with the presence or absence of links elsewhere in the network. de Paula et al. (2018) call preferences of this type "interdependent". It is the dependence of utility on $s_{ij}(\mathbf{d})$ that makes the model "strategic": agent i's optimal action may vary with the configuration of links directed by others.

For now the only restriction we place on $s_{ij}(\mathbf{d})$ is that

$$s_{ij}(\mathbf{d}) = s_{ij}(\mathbf{d} - ij) = s_{ij}(\mathbf{d} + ij). \tag{3}$$

If existence of a *pure* strategy equilibrium is additionally desired, then additional restrictions on $s_{ij}(\mathbf{d})$ may be needed. Although we emphasize pure strategy equilibria in our discussion and examples, all of our results allow for mixed strategy equilibria as well. Consequently, in practice, $s_{ij}(\mathbf{d})$ may be specified quite freely, although our test may have low power for some choices.

One feature of $s_{ij}(\mathbf{d})$, which will prove central to our analysis, is that is has finite range. To see this observe that since the set of all networks $\mathbb{D}_N$ is finite, the strategic interaction term $s_{ij}(\mathbf{d})$ also takes only a finite number of values. Let $\mathbb{S} = \{\underline{s}, s_1, \ldots, s_M, \overline{s}\}$ be the set of possible values for $s_{ij}(\mathbf{d})$, ordered from smallest to largest.

[4] We define $W_{ij} = (X_i \otimes X_j)$ and $\lambda = \mathrm{vec}(\Lambda')$.

An example illustrates. Let $s_{ij}(\mathbf{d})$ equal

$$s_{ij}(\mathbf{d}) = d_{ji}, \tag{4}$$

as would be appropriate when agents have a taste for reciprocated links. In this case $\mathbb{S} = \{0, 1\}$. If agents prefer transitive links (i.e., they prefer to direct friendships to "friends of friends"), then we might set

$$s_{ij}(\mathbf{d}) = \sum_k d_{ik} d_{kj}, \tag{5}$$

which implies that $\mathbb{S} = \{0, 1, \ldots, N-1\}$. Finiteness of the cardinality of $\mathbb{S}$ (for a given N) plays an important role in our analysis, as will become apparent below.

The final component of agent utility is idiosyncratic; we assume that the $\{U_{ij}\}_{i \neq j}$ are independent and identically distributed (iid) logistic random variables. The logistic assumption is also important: it generates exponential family structure which we exploit when forming our test.

1.3 Equilibrium networks

Throughout we assume that the observed network $\mathbf{D}$ coincides with an equilibrium outcome of an N-player complete information game. Each agent (i) observes $\{(A_i, B_i, X_i')\}_{i=1}^N$ and $\{U_{ij}\}_{i \neq j}$ and then (ii) decides which, out of $N-1$ other agents, to send links to. Agents may play mixed strategies.

Any feasible network configuration $\mathbf{d} \in \mathbb{D}_N$ coincides with a pure strategy combination. We assume that the observed network corresponds to a pure strategy contained in a (possibly mixed strategy) Nash equilibrium (NE). In practice most (common) choices of $s_{ij}(\mathbf{d})$ are monotonic, which ensures (by Tarski fixed point theorem), the existence of an equilibrium in pure strategies. We emphasize this special case in most of what follows, but nothing essential hinges upon it and our results apply to equilibria in mixed strategies as well.

In the analysis of *undirected* networks, the pairwise stability equilibrium concept introduced by Jackson and Wolinsky (1996) plays a prominent role. The use of NE, however, is standard in the context of *directed* networks. For example, Bala and Goyal (2000) and Dutta and Jackson (2000) study the efficiency properties of pure strategy NE directed networks.

Pure strategy equilibria

A pure strategy NE corresponds to a pure strategy combination $\mathbf{d}^*$ where, for $\mathbf{U} = \mathbf{u}$ (with $\mathbf{U} \stackrel{def}{\equiv} [U_{ij}]$) and all $i = 1, \ldots, N$,

$$\nu_i \left(\mathbf{d}_i^*, \mathbf{d}_{-i}^*, \mathbf{u}\right) \geq \nu_i \left(\mathbf{d}_i, \mathbf{d}_{-i}^*, \mathbf{u}\right) \tag{6}$$

for all possible pure strategies $\mathbf{d}_i \in \mathbb{I}_{N-1}$.

To further understand the structure of a pure strategy equilibrium it is helpful to introduce a notion of marginal utility. The marginal utility i receives from sending a link to j equals:

$$MU_{ij} \left(\mathbf{d}_i, \mathbf{d}_{-i}; \mathbf{U}\right) = \begin{cases} \nu_i \left(\mathbf{d}\right) - \nu_i \left(\mathbf{d} - ij\right) & \text{if } d_{ij} = 1 \\ \nu_i \left(\mathbf{d} + ij\right) - \nu_i \left(\mathbf{d}\right) & \text{if } d_{ij} = 0 \end{cases}. \tag{7}$$

Under preferences (2) the marginal utility of the ij link is therefore

$$MU_{ij} \left(\mathbf{d}_i, \mathbf{d}_{-i}; \mathbf{U}\right) = A_i + B_j + W_{ij}' \lambda_0 + \gamma_0 s_{ij} \left(\mathbf{d}\right) - U_{ij}. \tag{8}$$

With this notation any adjacency matrix which simultaneously satisfies the $N(N-1)$ non-linear equations:

$$D_{ij} = \mathbf{1}(A_i + B_j + W_{ij}' \lambda + \gamma_0 s_{ij} \left(\mathbf{D}\right) \geq U_{ij}) \tag{9}$$

for $i = 1, \ldots, N$ and $j \neq i$ is a pure strategy NE.

Similar to Miyauchi (2016), consider the mapping $\varphi(\mathbf{D}) : \mathbb{D}_N \to \mathbb{I}_{N(N-1)}$:

$$\underset{N(N-1)\times 1}{\varphi(\mathbf{D})} \equiv \begin{bmatrix} \mathbf{1}(MU_{12}(\mathbf{D}) \geq 0) \\ \mathbf{1}(MU_{13}(\mathbf{D}) \geq 0) \\ \vdots \\ \mathbf{1}(MU_{NN-1}(\mathbf{D}) \geq 0) \end{bmatrix}. \tag{10}$$

Next let $\text{vec}_*(\mathbf{A})$ be a modification of the matrix vectorization operator which drops the diagonal elements of the square matrix $\mathbf{A}$. Define its inverse operator as reconstituting $\mathbf{A}$, but now with zeros on its main diagonal. With this notation it easy to see that any pure strategy NE equilibrium network, $\mathbf{d}^*$, including possibly the observed one, $\mathbf{D}$, corresponds to a fixed point:

$$\mathbf{d}^* = \text{vec}_*^{-1}(\varphi(\mathbf{d}^*)). \tag{11}$$

One advantage of the fixed point representation (11) is that is allows for the application of Tarski's (1955) fixed point theorem. For $\gamma \geq 0$ and $s_{ij}(\mathbf{d})$ weakly increasing in $\mathbf{d}$ for all dyads, Tarski's theorem guarantees (i) the existence of an equilibrium and (ii) that the set of all equilibria constitutes a non-empty complete lattice (cf., Miyauchi, 2016, Proposition 1). This characterization applies when $s_{ij}(\mathbf{d})$ takes either of the two example forms introduced above. Of course, as shown by Nash (1950), an equilibrium in mixed strategies will always exist. In practice this allows for substantial flexibility in the form of $s_{ij}(\mathbf{d})$.

1.4 Likelihood

While we remain agnostic about equilibrium selection in the presence of multiplicity, it is nevertheless useful to develop an abstract notation for the unknown equilibrium selection rule. This notation allows us to write down a likelihood for the network. Of course this likelihood function could not be evaluated numerically without first replacing our abstract selection mechanism with something more concrete.

Let $n \overset{def}{\equiv} N(N-1)$ equal the number of (ordered) dyads in the network. Further let $\mathbf{A} \overset{def}{\equiv} [A_i]$ and $\mathbf{B} \overset{def}{\equiv} [B_i]$ be the $N \times 1$ vectors of agent-specific out- and in-degree heterogeneity parameters. The non-strategic (nuisance) parameters of our model are collected in $\delta = (\lambda', \mathbf{A}', \mathbf{B}')'$, recalling that λ is the $K^2 \times 1$ vector which parameterizes homophily on X_i. Adding our strategic interaction parameter, γ we get a full parameter vector of $\theta = (\gamma, \delta')'$.

Let $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ be a function which assigns, for $\mathbf{U} = \mathbf{u}$, a probability weight to network or, equivalently, pure strategy combination $\mathbf{d}$:

$$\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) : \mathbb{D}_N \times \mathbb{R}^n \to [0, 1] \tag{12}$$

We assume that the selection mechanism (12) is such that:

1. if $\mathbf{d}$ is the only network which satisfies (6) when $\mathbf{U} = \mathbf{u}$ (i.e., is the unique NE), then $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) = 1$;

2. if $\mathbf{d}$ is not a NE when $\mathbf{U} = \mathbf{u}$, then $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) = 0$;

3. if there are multiple pure strategy NE, then $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) \geq 0$ for any $\mathbf{d}$ which is a NE and zero otherwise (subject to the adding-up constraint $\sum_{\mathbf{d} \in \mathbb{D}_N} \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) = 1$);

4. if there is a unique *mixed* strategy NE when $\mathbf{U} = \mathbf{u}$, then $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) \geq 0$ equals the probability assigned to pure strategy combination $\mathbf{d}$ (contained in the mixed strategy NE). If there are multiple mixed strategy NE when $\mathbf{U} = \mathbf{u}$, then $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) \geq 0$ will additionally reflect the probabilities attached to different equilibria, etc.

With $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ defined, we can write the likelihood of observing network $\mathbf{D} = \mathbf{d}$ as

$$P(\mathbf{d}; \theta, \mathcal{N}) = \int_{\mathbf{u} \in \mathbb{R}^n} \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) f_{\mathbf{u}}(\mathbf{u}) \mathrm{d}\mathbf{u}, \tag{13}$$

where $f_{\mathbf{u}}(\mathbf{u}) = \prod_{i \neq j} f_U(u_{ij})$ with $f_U(u) = e^u/[1+e^u]^2$

Theorem 1.1. *For any network* $\mathbf{d} \in \mathbb{D}_N$ *there exists a measurable function* $\mathcal{N}(\mathbf{d}, \cdot; \theta) : \mathbb{R}^n \to [0, 1]$*, which assigns to* $\mathbf{u} \in \mathbb{R}^n$ *the NE weight on the pure strategy combination corresponding to* $\mathbf{d}$.

The proof of Theorem 1.1 can be found in Appendix A.1. Although we do not explicitly define $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$, only stating its key properties, Theorem 1.1 shows that $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ exists and is measurable. An implication of this result is that the likelihood (13) is well-defined.

To understand the likelihood (13) it is helpful to consider a (relatively) simple example. This example will also help in understanding our derivation of the optimal test statistic below. Assume that $s_{ij}(\mathbf{d}) = d_{ji}$ such that agents prefer reciprocated links when $\gamma \geq 0$. In this example $s_{ij}(\mathbf{d})$ equals either zero (j does not reciprocate) or one (j does reciprocate). We can use the two elements of $\mathbb{S}$ to partition the real line into what we will call *buckets*:

$$\mathbb{R} = (-\infty, \mu_{ij}] \cup (\mu_{ij}, \mu_{ij} + \gamma] \cup (\mu_{ij} + \gamma, \infty). \tag{14}$$

Here $\mu_{ij} = A_i + B_j + X_i' \Lambda_0 X_i$ equals the systematic, non-strategic, component of utility generated by arc ij. Next consider the realization of U_{ij}, the idiosyncratic utility agent i gets when she directs a link to j. If U_{ij} falls into the first bucket in (14), then agent i will always direct a link to j; irrespective of whether j chooses to direct a link to i or not. If U_{ij} falls into the middle or *inner* bucket, however, then i will direct a link to j only if j reciprocates. Finally, if U_{ij} falls into the last bucket, then i will never direct a link to j regardless of whether j directs a link to i or not. We will call the first and last buckets in (14) *outer* buckets.

If both U_{ij} and U_{ji} fall in their respective *inner* buckets, then the $\{i, j\}$ dyad can either take the empty ($D_{ij} = D_{ji} = 0$) or reciprocated ($D_{ij} = D_{ji} = 1$) configuration in equilibrium. In contrast, if *either* U_{ij} or U_{ji} falls into an outer bucket, then the $\{i, j\}$ dyad's wiring is uniquely determined. For example if U_{ij} is in the first outer bucket and U_{ji} is in the inner bucket, then the $\{i, j\}$ dyad will take the reciprocated form with probability one. It is a strictly dominant strategy for i to direct an link to j in this case and a best response for j to reciprocate.

For $\mathbf{U} = \mathbf{u}$, let $J(\mathbf{u}) \leq \binom{N}{2}$ equal the number of dyads $\{i, j\}$, where both u_{ij} and u_{ji} fall into their inner bucket. For each of these dyads both the empty and reciprocated configuration is an equilibrium outcome. There are therefore $2^{J(\mathbf{u})}$ equilibrium networks in this case; the $\mathcal{N}(\mathbf{d},\mathbf{u};\theta)$ function would assign some probability between zero and one to each of these $2^{J(\mathbf{u})}$ networks (summing to one in total).[5]

Let $\mathbb{D}_N^{\mathrm{NE}}(\mathbf{u})$ be the set of $2^{J(\mathbf{u})}$ equilibrium networks when $\mathbf{U}=\mathbf{u}$. One equilibrium selection rule would assign equal probability to all NE. In this case we could write the likelihood as

$$P(\mathbf{d};\theta,\mathcal{N})=\int_{\mathbf{u}\in\mathbb{R}^n}\frac{\mathbf{1}\left(\mathbf{d}\in\mathbb{D}_N^{\mathrm{NE}}(\mathbf{u})\right)}{\left|\mathbb{D}_N^{\mathrm{NE}}(\mathbf{u})\right|}f_{\mathbf{u}}(\mathbf{u})\mathrm{d}\mathbf{u}, \tag{15}$$

such that $\mathcal{N}(\mathbf{d},\mathbf{u};\theta)=\frac{\mathbf{1}\left(\mathbf{d}\in\mathbb{D}_N^{\mathrm{NE}}(\mathbf{u})\right)}{\left|\mathbb{D}_N^{\mathrm{NE}}(\mathbf{u})\right|}$. This example illustrates that (13), while well-defined, is generally intractable; even when the equilibrium selection mechanism is fully-specified.

2 Testing for strategic interaction

Our goal is to construct a powerful test for the presence of strategic interaction in network formation with good size properties. Importantly we wish to remain agnostic about any degree heterogeneity and homophily. Let Δ denote a subset of the K^2+2N dimensional Euclidean space in which $\delta_0=(\lambda_0,\mathbf{A}_0,\mathbf{B}_0)$ is, a priori, known to lie, and

$$\Theta_0=\left\{(\gamma,\delta')\,:\,\gamma=0,\delta\in\Delta\right\}. \tag{16}$$

Our null hypothesis is the *composite* one:

$$H_0\,:\,\theta\in\Theta_0 \tag{17}$$

since δ may range freely over $\Delta\subset\mathbb{R}^{K^2+2N}$ under the null.

Under the null the likelihood is $P_0(\mathbf{d};\delta)\overset{def}{\equiv}P(\mathbf{d};(0,\delta')',\mathcal{N}_0)$ with

$$\begin{aligned}\mathcal{N}_0(\mathbf{d},\mathbf{u};\theta)=&\prod_i\prod_j\mathbf{1}\left(A_i+B_j+W_{ij}'\lambda\geq u_{ij}\right)^{d_{ij}}\\&\times\mathbf{1}\left(A_i+B_j+W_{ij}'\lambda<u_{ij}\right)^{1-d_{ij}}.\end{aligned}$$

Under the null the unique "equilibrium" network is the one where all links with positive marginal utility are present and those with negative marginal utility are not; $\mathcal{N}_0(\mathbf{d},\mathbf{u};\theta)$ places a probability of 1 on this network. Evaluating the integral (13) under the null yields

[5] For simplicity we ignore mixed strategy equilibria in this example.

$$P_0(\mathbf{d};\delta) = \prod_{i=1}^{N}\prod_{j\neq i}\left[\frac{\exp\left(W_{ij}'\lambda + R_i'\mathbf{A} + R_j'\mathbf{B}\right)}{1+\exp\left(W_{ij}'\lambda + R_i'\mathbf{A} + R_j'\mathbf{B}\right)}\right]^{d_{ij}}$$
$$\times\left[\frac{1}{1+\exp\left(W_{ij}'\lambda + R_i'\mathbf{A} + R_j'\mathbf{B}\right)}\right]^{1-d_{ij}}$$

where R_i is the $N \times 1$ vector with a 1 in its i^{th} element and zeros elsewhere. Variants of this likelihood are analyzed by Chatterjee et al. (2011), Charbonneau (2017), Graham (2017), Jochmans (2018), Dzemski (2018) and Yan et al. (2018).

2.1 Exponential family structure under the null

Under the null our likelihood, $P_0(\mathbf{d};\delta)$, is a member of the exponential family. To see this it is helpful to establish some additional notation. The out- and in-degree sequences equal:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{\text{out}} \\ \mathbf{S}_{\text{int}} \end{pmatrix}' = \begin{pmatrix} D_{1+},\ldots,D_{N+} \\ D_{+1},\ldots,D_{+N} \end{pmatrix}. \tag{18}$$

Here $D_{+i} = \sum_j D_{ji}$ and $D_{i+} = \sum_j D_{ij}$ equal the in- and out-degree of agents $i = 1,\ldots,N$.

The $K \times K$ *cross-link matrix* equals

$$\mathbf{M} = \sum_i \sum_j D_{ij} X_i X_j'. \tag{19}$$

This matrix summarizes the inter-group link structure in the network (homophily). The kl^{th} element of $\mathbf{M}$ records the number of links sent by type k agents (e.g., semiconductor manufacturers) to type l agents (e.g., computer manufacturers).

Let $\mathbf{S},\mathbf{M}$ be a degree sequence and cross-link matrix. We say $\mathbf{S},\mathbf{M}$ is *graphical* if there exists at least one arc set $\mathcal{A}$ such that $G(\mathcal{V},\mathcal{A})$ is a simple directed graph with degree sequence $\mathbf{S}$ and cross link matrix $\mathbf{M}$. We call any such network a *realization* of $\mathbf{S},\mathbf{M}$. The set of all possible realizations of $\mathbf{S},\mathbf{M}$ is denoted by $\mathbb{G}_{\mathbf{S},\mathbf{M}}$ ($\mathbb{D}_{\mathbf{S},\mathbf{M}}$ denotes the associated set of adjacency matrices).

With this notation it is easy to verify that the null model belongs to the exponential family:

$$P_0(\mathbf{d};\delta) = c(\delta)\exp\left(\mathbf{t}'\delta\right) \tag{20}$$

with a (minimally) sufficient statistic for δ of $\mathbf{t} = \left(\operatorname{vec}\left(\mathbf{m}'\right)', \mathbf{s}_{\text{out}}', \mathbf{s}_{\text{in}}'\right)'$. In words, the $K^2 + N + N$ sufficient statistics are (i) the cross link matrix, (ii) the out-degree sequence and (iii) the in-degree sequence.

Under H_0 the conditional likelihood of the event $\mathbf{D} = \mathbf{d}$ is

$$P_0\left(\mathbf{d}\middle|\, \mathbf{T} = \mathbf{t}\right) = \frac{1}{|\mathbb{D}_{\mathbf{s},\mathbf{m}}|} \tag{21}$$

if $\mathbf{d} \in \mathbb{D}_{\mathbf{s},\mathbf{m}}$ and zero otherwise. Under the null of no strategic interaction all networks with the same in- and out-degree sequences and cross link structure are equally likely. Importantly this conditional likelihood is invariant to actual value of the nuisance parameter δ.

2.2 Locally best similar test

In our setting a test with critical function $\phi(\mathbf{D})$ will have size α if its null rejection probability (NRP) is less than or equal to α for *all* values of the nuisance parameter:

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta \left[\phi(\mathbf{D})\right] = \sup_{\gamma = \gamma_0, \delta \in \triangle} \mathbb{E}_\theta \left[\phi(\mathbf{D})\right] = \alpha. \tag{22}$$

Since the nuisance parameter δ is very high dimensional, size control is non-trivial. For some intuition as to why consider, as an example, the case where $s_{ij}(\mathbf{d}) = \sum_k d_{ik} d_{kj}$, such that agents' have a taste for transitivity when $\gamma_0 > 0$. A natural test statistic in this case would be some function of $\mathbf{D}$ that is increasing in the number of transitive triads, $\triangle$, in the network. The researcher would then reject the null of $\gamma_0 = 0$ when this statistic is large enough. Unfortunately, the expected number of transitive triads varies dramatically under the null depending on the value of δ. Certain configurations of $\mathbf{A}$, $\mathbf{B}$ and/or λ may result in a network with large numbers of transitive triads even when agents' have no taste for transitivity per se (i.e., under the null). If we choose a single critical value for rejection then, depending on the values of $\mathbf{A}$, $\mathbf{B}$ and/or λ, size may be very poor.

To avoid any size distortion induced by variation in δ over $\Delta \subset \mathbb{R}^{K^2+2N}$ we exploit the exponential family structure of our model (under the null). Let $\mathbb{T} = \{(\mathbf{s}, \mathbf{m}) : \mathbf{s}, \mathbf{m} \text{ is graphical}\}$ be the set of possible sufficient statistics $\mathbf{T}$. We proceed conditional on $\mathbf{T}$; that is, instead of choosing a single critical value, which may result in under- or over-rejection, depending on the value of δ, we proceed conditionally on $\mathbf{T}$ (the minimally sufficient statistic for δ). Our chosen critical value varies with $\mathbf{T}$. In this way we ensure good size control.

Formally, for each $\mathbf{t} \in \mathbb{T}$ we form a test with the property that, for all $\theta \in \Theta_0$,

$$\mathbb{E}_\theta \left[\phi(\mathbf{D})\middle|\, \mathbf{T} = \mathbf{t}\right] = \alpha. \tag{23}$$

Such an approach ensures *similarity* of our test since, by iterated expectations,

$$\mathbb{E}_{\theta}\left[\phi\left(\mathbf{D}\right)\right] = \mathbb{E}_{\theta}\left[\mathbb{E}_{\theta}\left[\phi\left(\mathbf{D}\right)\middle|\,\mathbf{T}\right]\right] = \alpha \tag{24}$$

for any $\theta \in \Theta_0$ (Ferguson, 1967). By proceeding conditionally we ensure that the NRP is unaffected by the value of δ.

By Ferguson (1967, Lemma 1, Section 3.6) $\mathbf{T}$ is a boundedly complete sufficient statistic for θ under the null. By Ferguson (1967, Theorem 2, Section 5.4) every similar test will thus take the form

$$\mathbb{E}_{\theta}\left[\phi\left(\mathbf{D}\right)\middle|\,\mathbf{T} = \mathbf{t}\right] = \alpha \tag{25}$$

for $\mathbf{t} \in \mathbb{T}$. Therefore, if we desire similarity of our test we must take the conditional approach.

A conditional test

We have shown that proceeding conditionally results in a similar test. Here we outline, concretely, how to construct an *exact* similar test. There will be two limitations associated with this test. First, since the test statistic is chosen heuristically, it may not have good power in the direction of the alternative of primary interest. Second, it is generally not computationally feasible to compute the exact test.

In subsequent sections we address both of these limitations. Specifically we derive the form of the optimal test statistic and outline an MCMC algorithm for approximating its null distribution.

Let $R(\mathbf{D})$ be some statistic of the adjacency matrix. For example $R(\mathbf{D})$ might be the network reciprocity index (Newman, 2010):

$$R(\mathbf{D}) = \frac{2\hat{P}\left(\leftrightarrow\right)}{2\hat{P}\left(\leftrightarrow\right) + \hat{P}\left(\rightarrow\right)}, \tag{26}$$

where

$$\hat{P}\left(\rightarrow\right) = \frac{2}{N\left(N-1\right)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[D_{ij}\left(1 - D_{ji}\right) + \left(1 - D_{ij}\right) D_{ji}\right] \tag{27}$$

equals the fraction of dyads which take an unreciprocated or "asymmetric" $\rightarrow$ configuration and

$$\hat{P}\left(\leftrightarrow\right) = \frac{2}{N\left(N-1\right)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} D_{ij} D_{ji} \tag{28}$$

the fraction which take a reciprocated or "mutual" configuration $\leftrightarrow$. Heuristically, this choice of $R(\mathbf{D})$ might be useful for detecting whether agents have a taste for reciprocated links.

A conditional test based upon $R(\mathbf{D})$ will have a critical function of

$$\phi(\mathbf{d}) = \begin{cases} 1 & R(\mathbf{d}) > c_\alpha(\mathbf{t}) \\ g_\alpha(\mathbf{t}) & R(\mathbf{d}) = c_\alpha(\mathbf{t}) \\ 0 & R(\mathbf{d}) < c_\alpha(\mathbf{t}) \end{cases} \tag{29}$$

where the values of $c_\alpha(\mathbf{t})$ and $g_\alpha(\mathbf{t}) \in [0,1]$ are chosen to satisfy the requirement that $\mathbb{E}_\theta[\phi(\mathbf{D})|\, \mathbf{T} = \mathbf{t}] = \alpha$.

Under the null all adjacency matrices with the $\mathbf{S} = \mathbf{s}$ and $\mathbf{M} = \mathbf{m}$ are equally probable. Therefore the null distribution of $R(\mathbf{D})$ coincides with the one induced by a discrete uniform distribution on $\mathbb{D}_{\mathbf{s},\mathbf{m}}$. By enumerating all adjacency matrices in $\mathbb{D}_{\mathbf{s},\mathbf{m}}$ we could exactly compute this distribution and calculate the critical values $c_\alpha(\mathbf{t})$ and $g_\alpha(\mathbf{t})$. In general such a brute force approach will be infeasible.[6] Therefore a method of approximating the exact null distribution is required.

The intuition behind this test is straightforward. If the network in hand has an "unusually" large value of $R(\mathbf{D})$ relative to the set of all networks with same in- and out-degree sequences and cross-link matrix, then we reject our null.

The locally best conditional test

While choosing a statistic of adjacency matrix heuristically may lead to a test with good power in practice, there is no guarantee that it will.[7] In this section we derive, for any interdependent preference structure, a test statistic with good power to detect small deviations from the no strategic interaction null.

Under the alternative of strategic interaction the conditional likelihood is

$$P(\mathbf{d}|\, \mathbf{T} = \mathbf{t}; \theta, \mathcal{N}) = \frac{P(\mathbf{d}; \theta, \mathcal{N})}{\sum_{\mathbf{v} \in \mathbb{D}_{\mathbf{s},\mathbf{m}}} P(\mathbf{v}; \theta, \mathcal{N})}. \tag{30}$$

Two features of this likelihood make it impractical and/or unattractive for use in testing. First, it is complicated and (logically) cannot be evaluated without specifying an explicit equilibrium selection mechanism, $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$. We wish to develop inference methods which do not depend upon details of equilibrium selection. Even if the researcher were able to specify $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$, numerical evaluation of the likelihood may be impractical. Second, equation (30)

[6] In fact very little is known about the set $\mathbb{D}_{\mathbf{s},\mathbf{m}}$; for example we are aware of no method for checking whether a given $\mathbf{s}, \mathbf{m}$ pair is graphic. From related settings we believe that the cardinality of $\mathbb{D}_{\mathbf{s},\mathbf{m}}$ will typically be intractably huge even for modestly-sized networks. See Blitzstein and Diaconis (2011) for discussion of this point and examples from a related setting.

[7] Of course, there is longstanding tradition of choosing test statistics heuristically. See Cox (2006) for interesting discussion.

depends on the value of the nuisance parameter δ.

Assume that both δ_0 and the equilibrium selection mechanism $\mathcal{N}(\mathbf{d},\mathbf{u};\theta)$ are known (the latter up to knowledge of γ_0 of course), then, by the Neyman-Pearson Lemma, the most powerful test for the simple hypothesis $H_0:\gamma_0=0$ versus $H_1:\gamma_0=\gamma_a$ would be based upon the likelihood ratio (LR)

$$\frac{P\left(\mathbf{d}|\,\mathbf{T}=\mathbf{t};(\gamma_a,\delta_0')',\mathcal{N}\right)}{\left|\mathbb{D}_{\mathbf{s},\mathbf{m}}\right|^{-1}}.$$

If the likelihood of the network in hand ($\mathbf{D}=\mathbf{d}$) is high under the alternative relative to the discrete uniform reference distribution, then we reject. Unfortunately, as noted above, forming a LR requires specifying an equilibrium selection mechanism, $\mathcal{N}$. This is not straightforward to do, and we would prefer to avoid doing so in any case.

As an alternative to a LR test, we instead choose, for each $\mathbf{t}\in\mathbb{T}$, the critical function, $\phi(\mathbf{D})$ to maximize the *derivative* of the (conditional) power function $\beta(\gamma,\mathbf{t})=\mathbb{E}\left[\phi(\mathbf{D})|\,\mathbf{T}=\mathbf{t}\right]$ evaluated at $\gamma=0$ subject to the (conditional) size constraint $\mathbb{E}_\theta\left[\phi(\mathbf{D})|\,\mathbf{T}=\mathbf{t}\right]=\alpha$. Such a $\phi(\mathbf{D})$ is *locally best* (Ferguson, 1967, Lemma 1, Section 5.5). Remarkably we show that the locally best test doesn't not depend upon the form of the equilibrium selection mechanism $\mathcal{N}(\mathbf{d},\mathbf{u};\theta)$.

Differentiating the power function we get

$$\left.\frac{\partial\beta(\gamma,\mathbf{t})}{\partial\gamma}\right|_{\gamma=0}=\mathbb{E}\left[\phi(\mathbf{D})\,\mathbb{S}_\gamma(\mathbf{D}|\,\mathbf{T};\theta)|\,\mathbf{T}=\mathbf{t}\right]\tag{31}$$

with $\mathbb{S}_\gamma(\mathbf{d}|\,\mathbf{t};\theta)$ denoting the conditional score function

$$\begin{aligned}\mathbb{S}_\gamma(\mathbf{d}|\,\mathbf{t};\theta)&=\frac{1}{P_0(\mathbf{d};\delta)}\left.\frac{\partial P(\mathbf{d};\theta)}{\partial\gamma}\right|_{\gamma=0}-\sum_{\mathbf{v}\in\mathbb{D}_{\mathbf{s},\mathbf{m}}}\left.\frac{\partial P(\mathbf{v};\theta)}{\partial\gamma}\right|_{\gamma=0}\\&=\frac{1}{P_0(\mathbf{d};\delta)}\left.\frac{\partial P(\mathbf{d};\theta)}{\partial\gamma}\right|_{\gamma=0}+k(\mathbf{t})\end{aligned}$$

and $k(\mathbf{t})$ only depending on the data through $\mathbf{T}=\mathbf{t}$.

By the Neyman-Pearson lemma the test with critical function

$$\phi(\mathbf{d})=\begin{cases}1 & \frac{1}{P_0(\mathbf{d};\delta)}\left.\frac{\partial P(\mathbf{d};\theta)}{\partial\gamma}\right|_{\gamma=0}>c_\alpha(\mathbf{t})\\ g_\alpha(\mathbf{t}) & \frac{1}{P_0(\mathbf{d};\delta)}\left.\frac{\partial P(\mathbf{d};\theta)}{\partial\gamma}\right|_{\gamma=0}=c_\alpha(\mathbf{t})\\ 0 & \frac{1}{P_0(\mathbf{d};\delta)}\left.\frac{\partial P(\mathbf{d};\theta)}{\partial\gamma}\right|_{\gamma=0}<c_\alpha(\mathbf{t})\end{cases}\tag{32}$$

where the values of $c_\alpha(\mathbf{t})$ and $g_\alpha(\mathbf{t})\in[0,1]$ are chosen to satisfy (23), will be locally best.

The idea behind the locally best test is as follows. If the likelihood increases sharply as we move away from the null *in the direction of the alternative of interest*, then we take this as evidence against the null. Intuitively if the likelihood gradient in the neighborhood of the null is large, then the likelihood ratio will also be large for simple alternatives close to the null (i.e., when $\gamma_a \in (-\epsilon, \epsilon)$).

Constructing (32) requires calculating $\frac{1}{P_0(\mathbf{d};\delta)} \left.\frac{\partial P(\mathbf{d};\theta)}{\partial \gamma}\right|_{\gamma=0}$. This is not straightforward since it depends on properties of the likelihood under the alternative. Surprisingly we are able to derive the form of this derivative.

Theorem 2.1. *$P(\mathbf{d};\theta,\mathcal{N})$ is twice differentiable with respect to γ at $\gamma = 0$. Its first derivative at $\gamma = 0$ is*

$$\left.\frac{\partial P(\mathbf{d};\theta,\mathcal{N})}{\partial \gamma}\right|_{\gamma=0} = P_0(\mathbf{d};\delta) \times \left[\sum_{i\neq j} s_{ij}(\mathbf{d}) \left\{ d_{ij} \frac{f_U(\mu_{ij})}{\int_{-\infty}^{v_{ij}} f_U(u)\,\mathrm{d}u} - (1-d_{ij}) \frac{f_U(\mu_{ij})}{\int_{v_{ij}}^{\infty} f_U(u)\,\mathrm{d}u} \right\}\right]. \quad (33)$$

recalling that $\mu_{ij} = A_i + B_j + X_i'\Lambda_0 X_i$ equals the systematic, non-strategic, component of utility generated by arc ij and that f_U is the logistic density.

The proof of Theorem 2.1 can be found in Section A.2 of the Appendix. Here, because it is one of our main results, and also insightful to do so, we provide a high level overview of its derivation. Although $\left.\frac{\partial P(\mathbf{d};\theta,\mathcal{N})}{\partial \gamma}\right|_{\gamma=0}$ does vary with δ_0, it does *not* depend upon $\mathcal{N}$. Below we provide some intuition for this result.

Recall that $\mathbb{S} = \{\underline{s}, s_1, \ldots, s_M, \overline{s}\}$ equals the possible values of $s_{ij}(\mathbf{d})$, arranged from smallest to largest. We can use these support points to partition $\mathbb{R}$ into a set of intervals $\mathbb{B}$:

$$\mathbb{R} = (-\infty, \mu_{ij} + \gamma\underline{s}] \cup (\mu_{ij} + \gamma\underline{s}, \mu_{ij} + \gamma s_1] \cup \cdots \cup (\mu_{ij} + \gamma s_M, \mu_{ij} + \gamma\overline{s}] \cup (\mu_{ij} + \gamma\overline{s}, \infty). \quad (34)$$

The elements of $\mathbb{B}$, called *buckets*, correspond to the intervals listed in (34). In principle we should write $\mathbb{B}_{ij}$ instead of $\mathbb{B}$, reflecting the dependence of the bucket definitions on the value of μ_{ij}, the systematic non-strategic utility associated with an i-to-j link. However, since this dependence is not essential to any of the arguments that follow we leave it implicit. Note that the cardinality of $\mathbb{B}$ does not depend on μ_{ij}, but instead equals $|\mathbb{S}| + 1$.

Agent i's linking behavior vis-a-vis j depends on which bucket U_{ij} falls into. For $B \in \mathbb{B}$, if $U_{ij} \in B$, then we say U_{ij} is in, or falls into, bucket B. The first and last buckets, respectively

$(-\infty, \mu_{ij} + \gamma \underline{s}]$ and $(\mu_{ij} + \gamma \overline{s}, \infty)$, play an important role in our argument. We call these two buckets *outer buckets*. The rest of the buckets we call *inner buckets*.

If U_{ij} falls into one of these outer buckets then player i has a pure strategy for d_{ij} which is strictly dominating. Specifically if U_{ij} falls into the lowest bucket, then i will direct an link to j regardless of what actions are taken by the other agents in the network. The marginal utility generated by link ij is so large that it remains positive across all possible configurations of the rest of the network; hence i always chooses to direct an link to j.

If, instead, U_{ij} falls into the highest bucket, then i will never direct an link to j. In this case the marginal utility associated with link ij is so low that it remains negative across all possible configurations of the rest of the network; hence i never chooses to direct a link to j.

Finally, if U_{ij} falls into an inner bucket, say $(\mu_{ij} + \gamma s_m, \mu_{ij} + \gamma s_{m+1}]$, then agent i's optimal choice for d_{ij} is contingent upon the linking behavior of other agents. If other agents' link actions are such that $s_{ij}(\mathbf{d}) \geq s_m$, then it is a best response for i to link with j, but not otherwise.

The vector of idiosyncratic taste shocks, $\mathbf{U}$ contains $n = N(N-1)$ elements; one for each possible arc. Let the boldface subscripts $\mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots$ index these potential arcs in arbitrary order (e.g., $\mathbf{i}$ maps to some ij and vice-versa). Let $\mathbf{b} \in \mathbb{B}^n \overset{def}{\equiv} \mathbb{B} \times \cdots \times \mathbb{B}$ and $\mathbf{U} = (U_{\mathbf{1}}, \ldots, U_{\mathbf{n}})'$; we have that $\mathbf{U} \in \mathbf{b}$ for $\mathbf{b} \in \mathbb{B}^n$ so that each element of $\mathbf{u}$ falls into a bucket.

With the above notation established we can rewrite the likelihood (13) as:

$$P(\mathbf{d}; \theta, \mathcal{N}) = \sum_{\mathbf{b} \in \mathbb{B}^n} \int_{\mathbf{u} \in \mathbf{b}} \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) f_{\mathbf{U}}(\mathbf{u}) \, d\mathbf{u} \tag{35}$$

Expression (35) suggests a derivation by cases approach to finding $\left.\frac{\partial P(\mathbf{d};\theta,\mathcal{N})}{\partial \gamma}\right|_{\gamma=0}$. Fortunately a brute force exhaustive approach is not required because it is possible to show that most of the summands in (35) do not influence the derivative at $\gamma = 0$.

Let $\tilde{\mathbb{B}}^n$ be the set of bucket configurations with at least two inner buckets. If at least two elements of $\mathbf{U}$ fall in inner buckets, then we have that $\mathbf{U} \in \mathbf{b}$ with $\mathbf{b} \in \tilde{\mathbb{B}}^n$. If, instead, at most one element of $\mathbf{U}$ falls in an inner bucket, then we have that $\mathbf{U} \in \mathbf{b}$ with $\mathbf{b} \in \mathbb{B}^n \setminus \tilde{\mathbb{B}}^n$. This set-up gives the likelihood decomposition:

$$P(\mathbf{d}; \theta, \mathcal{N}) = \tilde{P}(\mathbf{d}; \theta, \mathcal{N}) + Q(\mathbf{d}; \theta, \mathcal{N}), \tag{36}$$

with

$$\tilde{P}(\mathbf{d};\theta,\mathcal{N}) = \sum_{\mathbf{b}\in\mathbb{B}^n\setminus\tilde{\mathbb{B}}^n} \int_{\mathbf{u}\in\mathbf{b}} \mathcal{N}(\mathbf{d},\mathbf{u};\theta)\, f_{\mathbf{U}}(\mathbf{u})\,\mathrm{d}\mathbf{u} \tag{37}$$

$$Q(\mathbf{d};\theta,\mathcal{N}) = \sum_{\mathbf{b}\in\tilde{\mathbb{B}}^n} \int_{\mathbf{u}\in\mathbf{b}} \mathcal{N}(\mathbf{d},\mathbf{u};\theta)\, f_{\mathbf{U}}(\mathbf{u})\,\mathrm{d}\mathbf{u}. \tag{38}$$

To proove Theorem 2.1 we show that for $\gamma \to 0$

$$P(\mathbf{d};\theta,\mathcal{N}) = \tilde{P}(\mathbf{d};\theta,\mathcal{N}) + \mathcal{O}\left(\gamma^2\right). \tag{39}$$

Intuitively, this follows from the fact that the chance that two or more elements of $\mathbf{U}$ fall in inner buckets is negligible when γ is close to zero (because most of the probability mass for U_{ij} is contained in the two outer buckets when strategic interactions are small). Hence when calculating the optimal test statistic we are free to focus on the cases where either all, or all but one, of the elements of $\mathbf{U}$ fall in outer buckets. We can then show that

$$\left.\frac{\partial P(\mathbf{d};\theta,\mathcal{N})}{\partial\gamma}\right|_{\gamma=0} = \left.\frac{\partial \tilde{P}(\mathbf{d};\theta,\mathcal{N})}{\partial\gamma}\right|_{\gamma=0}. \tag{40}$$

Hence to derive the form of $\left.\frac{\partial P(\mathbf{d};\theta,\mathcal{N})}{\partial\gamma}\right|_{\gamma=0}$ we need only calculate $\left.\frac{\partial \tilde{P}(\mathbf{d};\theta,\mathcal{N})}{\partial\gamma}\right|_{\gamma=0}$. This calculation is non-trivial, but doable. Details of this calculation are provided in the proof.

It was not *ex ante* obvious that a useful expression for $\left.\frac{\partial P(\mathbf{d};\theta,\mathcal{N})}{\partial\gamma}\right|_{\gamma=0}$ would be available without any assumptions about the nature of equilibrium selection under the alternative. That such as expression is available follows from the fact that when γ is small most agents will have a strictly dominating pure strategy for how to link; hence the chance for multiple equilibria is low and the form of $\mathcal{N}$ can generally be deduced. Conversely, when γ is small the probability of a draw of $\mathbf{U}$ where many agents do not have a strictly dominating pure strategy, and hence the details of equilibrium selection matter, is very low.

Locally best vs. heuristic test statistics

With a little manipulation we can simplify:

$$\frac{1}{P_0(\mathbf{d};\delta)} \left.\frac{\partial P(\mathbf{d};\theta)}{\partial\gamma}\right|_{\gamma=0} = \sum_{i\neq j} \left[d_{ij} - F_U(\mu_{ij})\right] s_{ij}(\mathbf{d}) \tag{41}$$

where $F_U(u) = e^u / [1+e^u]$ is the logistic CDF. This form of the statistic provides insight into how our test accumulates evidence against the null in practice. Consider the case where

$s_{ij}(\mathbf{d}) = d_{ji}$, as would be true in agents' have a taste for reciprocated links. Observe that $F_U(\mu_{ij})$ corresponds to the probability of an ij edge under the null. Therefore the optimal test statistic is large if we observe that many ij links *with low probability under the null* are reciprocated. It is not many reciprocated links that drives rejection per se, but the presence of many "unexpected" reciprocated links.

Consider a network of boys and girls with agents exhibiting a strong taste for gender-based homophily. The optimal test statistic in this case is the *conditional* sample covariance of D_{ij} and D_{ji} given (A_i, B_i, X_i) and (A_j, B_j, X_j). The test based upon the reciprocity index is – essentially – based upon the *unconditional* covariance. The effect of conditioning is to, for example, given more weight to heterophilous reciprocated links than to homophilous ones. Similarly we give more weight to reciprocated links across low degree agents, than to those across high degree agents.

3 Simulation

Because a complete enumeration of $\mathbb{D}_{\mathbf{s},\mathbf{m}}$ is not feasible unless N is very small, making our test practical requires a method of constructing uniform random draws from this set. Such draws can be used to simulate the null distribution of any test statistic of interest.

The problem of simulating binary matrices with fixed marginals is well-studied (e.g., Sinclair, 1993); with many domain specific applications (e.g., species co-occurrence/interaction analysis). In practice one of two simulation approaches is used (see Kolaczyk (2009) for a textbook overview).

The first approach begins with an empty graph and randomly adds links. Links need to be added such that the end graph satisfies the degree sequence constraint. Blitzstein and Diaconis (2011) develop an algorithm along these lines. They cleverly use checks for graphicality of a degree sequence, available in the discrete math literature, to add links in a way which constrains the end graph to be in the target set. They further use importance sampling to ensure that averages of simulated network statistics are with respect to the target uniform distribution. See also Del Genio et al. (2010) and Kim et al. (2012). Graham and Pelican (2020) provide a textbook discussion of the Blitzstein and Diaconis (2011) algorithm.

The second approach, to which our new method belongs, uses MCMC. Specifically an initial graph, satisfying the target constraints, is randomly rewired many times to create a new graph from the target set. Key to this approach is ensuring that each rewiring is compatible with the target constraints (e.g., maintains the network's degree sequence). The algorithm also needs to be constructed carefully to ensure that the end graph is a *uniform* random draw from the target set. Sinclair (1993), Rao et al. (1996), McDonald et al. (2007), Berger and Müller-Hannemann (2009) and Tao (2016) all developed MCMC methods for simulating graphs (or digraphs) with given degree sequences.

We are aware of no extant method of generating adjacency matrix draws from $\mathbb{D}_{\mathbf{s},\mathbf{m}}$. The novelty of this problem, relative to the work described above, is the presence of the additional cross link matrix constraint, $\mathbf{M}$. In the discrete math literature the cross link matrix constraint corresponds to what is called a partition adjacency matrix (PAM) constraint. Czabarka et al. (2017) conjecture that the determining whether a given $\mathbf{s}, \mathbf{m}$ pair is graphical, the PAM realization problem, is NP-complete. If their conjecture is correct (and NP $\neq$ P), using a Blitzstein and Diaconis (2011) type algorithm to draw from $\mathbb{D}_{\mathbf{s},\mathbf{m}}$ is not feasible.

This leaves MCMC methods. Erdős et al. (2017) showed that naively incorporating a PAM constraint into existing MCMC algorithms destroys their correctness. In this section we introduce a new MCMC algorithm that *does* generate uniform random draws from $\mathbb{D}_{\mathbf{s},\mathbf{m}}$. This algorithm is of independent interest. Before describing the algorithm we introduce some additional definitions and notation.

3.1 Notation and definitions

We start by defining an alternating walk.

Definition 3.1. (Alternating Walk) An alternating walk H is sequence of (ordered) dyads of the form

$$H := (i_1, i_2), (i_3, i_2), (i_3, i_4), \ldots, (i_l, i_{l-1}) \tag{42}$$

or

$$H := (i_2, i_1), (i_2, i_3), (i_4, i_3), \ldots, (i_{l-1}, i_l) \tag{43}$$

with $i_k \in \mathcal{V}(G)$, $i_k \neq i_{k+1}$, $i_k \neq i_{k-1}$and
(i) if $(i_k, i_{k-1}), (i_k, i_{k+1})$ in H, $(i_k, i_{k-1}) \in \mathcal{A}(G)$, then $(i_k, i_{k+1}) \notin \mathcal{A}(G)$
(ii) if $(i_k, i_{k-1}), (i_k, i_{k+1})$ in H, $(i_k, i_{k-1}) \notin \mathcal{A}(G)$, then $(i_k, i_{k+1}) \in \mathcal{A}(G)$
(ii) if $(i_{k-1}, i_k), (i_{k+1}, i_k)$ in H, $(i_{k-1}, i_k) \in \mathcal{A}(G)$, then $(i_{k+1}, i_k) \notin \mathcal{A}(G)$
(iv) if $(i_{k-1}, i_k), (i_{k+1}, i_k)$ in H, $(i_{k-1}, i_k) \notin \mathcal{A}(G)$, then $(i_{k+1}, i_k) \in \mathcal{A}(G)$
for all $k = 2, \ldots, l-1$.

For brevity we will often refer to a walk simply by its node sequence, writing $H := i_1 i_2, \ldots, i_l$. To unpack Definition 3.1 it is easiest to consider an example.

Observe that for $H := i_1 i_2, \ldots, i_l$, the adjacency matrix entries $D_{i_1 i_2}, D_{i_3 i_2}, \ldots, D_{i_l i_{l-1}}$ alternate between ones and zeros (or zeros and ones). This observation suggests a method of constructing an alternating walk via a sequence of "hops" across the adjacency matrix: pick row i_1 of the adjacency matrix and move horizontally to column i_2, where i_2 corresponds to one of the agents to which i_1 directs a link, next move vertically to row i_3, where i_3 is an agent which does not direct a link to i_2, and so on.[8] We call the horizontal moves *active steps* and vertical moves *passive steps*. Figure 2 provides an example construction. The different cases in Definition 3.1 correspond to walks beginning/ending with passive/active steps.

The length of an alternating walk equals the number of ordered dyads used to define it. An important type of alternating walk, which following Tao (2016), we call an *alternating cycle*, is central to our algorithm.

Definition 3.2. (Alternating Cycle) The alternating walk C is an alternating cycle if $i_1 = i_l$ and C has even length.

The length of an alternating cycle is at least four. Let $D_{i_1 i_2}, D_{i_3 i_1}, \ldots, D_{i_l i_{l-1}}$ be the sequence of adjacency matrix entries associated with alternating cycle C in G. These entries necessarily form a sequence of zeros and ones (or ones and zeros).

Consider constructing an alternative digraph, say G', by replacing all the "ones" in C with "zeros" and all "zeros" with "ones". Rewiring G in this way is degree preserving: G' has the same in- and out-degree sequence as G. We refer to such operations as switching the cycle (since we switch the zeros and ones).

Figure 3 depicts two canonical alternating cycles (Rao et al., 1996). The first, $C := abcda$, is a so called *alternating rectangle*. In the configuration to the left a and d each have a single outlink and b and c a single inlink; this is also true in the configuration to the right, which corresponds to the the network generated by switching C. The second cycle, called a *compact alternating hexagon*, can be constructed from a $030C$ () triad.

Let $\mathbb{G}_{\mathbf{s}}$ denote the set of all digraphs with degree sequence $\mathbf{s}$. Rao et al. (1996) showed that, for G and G' distinct and belonging to $\mathbb{G}_{\mathbf{s}}$, it is possible to obtain G' from G by switching a sequence of alternating rectangles and compact alternating hexagons. In practice it may be useful, in the sense that one can move from G to G' with fewer cycle switches, if longer alternating cycles are used (e.g., McDonald et al., 2007; Tao, 2016).

Let K_N denote the complete graph on N vertices. We define the link marking function $m : \mathcal{A}(K_N) \rightarrow \{0, 1\}$. We say the link (i_1, i_2) is marked if $m((i_1, i_2)) = 1$ and unmarked if $m((i_1, i_2)) = 0$. The expression "mark a link" means the marking function is changed such that $m((i_1, i_2)) = 1$. We use the expression "unmark a link analogously".

Definition 3.3. (Schlaufe) An alternating walk $H := i_1 i_2 \ldots i_l$ is a *schlaufe* if either
(i) There is a node $i_k \in \{i_1 i_2 \ldots i_l\}$ with $k \neq l$ such that $i_k = i_l$ and $(k - l) \bmod 2 = 0$. Furthermore for any two nodes i_j and i_h in $\{i_1 i_2 \ldots i_{l-1}\}$ with $i_j = i_h$ and $j \neq h$ it holds that $(j - h) \bmod 2 = 1$.

[8]This description is essentially due to (Tao, 2016, p. 124).

Figure 2: Constructing an alternating walk

A: Alternating Walk

	a	b	c	d	e	f	g	h	i	j
a	0	1	0	0	1	0	0	0	0	0
b	0	0	0	0	0	0	0	0	0	0
c	0	0	0	1	0	0	0	1	0	0
d	0	0	0	0	0	0	0	0	0	0
e	0	0	1	0	0	0	0	0	0	0
f	0	0	0	0	0	0	0	1	1	0
g	0	0	0	0	0	0	0	0	0	0
h	0	0	1	0	0	0	1	0	0	0
i	0	0	0	0	0	0	0	0	0	0
j	0	0	0	0	0	0	1	0	1	0

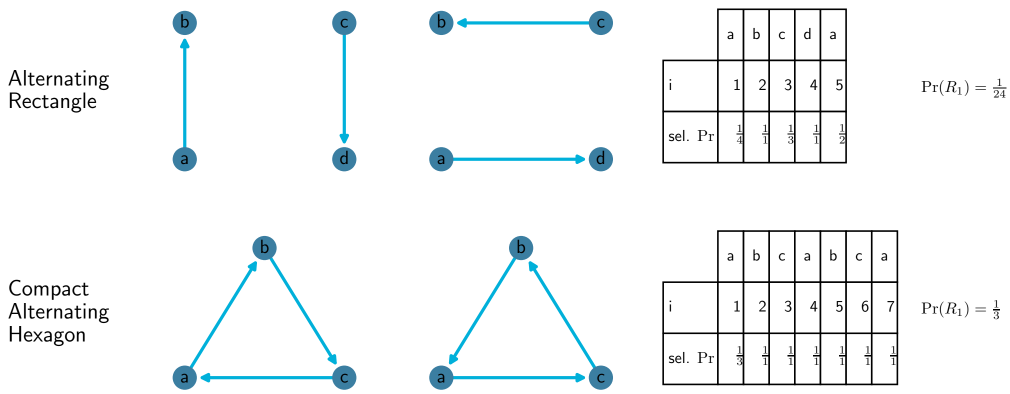
B: Degree Sequence

	Indegree	Outdegree
a	0	2
b	1	0
c	2	2
d	1	0
e	1	1
f	0	2
g	2	0
h	2	2
i	2	0
j	0	2

Source: Authors' calculations.

Notes: Panel A depicts an alternating walk constructed using the adjacency matrix. Agent labels are given in the first column and row of the table. To construct such a walk randomly we begin by choosing an agent at random. Here agent j is chosen, with an ex ante probability of $\frac{1}{10}$ since there are ten agents in the network. Next we take an active step where one of agent j's outlinks is chosen at random. Here we choose the outlink to agent g, an event with an ex ante probability of $\frac{1}{2}$ since agent j has just two outlinks. Following the active step comes a passive step. In a passive step we move vertically to the row of an agent which does not direct a link to the current agent. Here we choose a from the set $\{a, b, c, d, e, f, i\}$ uniformly at random (i.e., with an ex ante probability of $\frac{1}{7}$). We continue with active and passive steps until we choose to stop or can proceed no further. Panel B reports the indegree and outdegree of each agent in the network. Observe that in active steps the probability of any feasible choice equals the inverse of the outdegree of the current agent. In passive steps the probability of any feasible choice equals the inverse of the number of nodes minus the indegree of the node chosen in the prior step minus 1 (since $i_k \neq i_{k+1}$). We can also construct alternating walks by the above procedure, but instead starting with a passive step.

Figure 3: Examples of alternating cycles



Source: Authors' calculations.
Notes: The first row depicts an alternating rectangle before and after switching. The second row depicts a compact alternating hexagon before and after switching. The final column of the figure shows the probability with which each node is chosen, resulting in the total probability of the schlaufe.

(ii) At node i_l there is no other node i_{l+1} such that the alternating walk could be extended with the unmarked link (i_l, i_{l+1}).

In German schlaufe corresponds to "loop", "bow" or "ribbon" (its plural is schlaufen); the latter translation is evocative of our meaning here. In the first case the schlaufe will coincide with an alternating walk which includes exactly one alternating cycle.[9] Visually schlaufen of the first type, with the nodes appropriately placed, will look like loops and ribbons. In the second case the schlaufe does not include an alternating cycle.

Associated with a schlaufe, R, is a $K \times K$ violation matrix which records the number of extra links from group k to group l generated by switching the alternating cycle in R (if there is one). Consider an alternating rectangle consisting of two boys and two girls. If initially one boy directs a link to the other and one girl directs a link to the other, then after switching the cycle the violation matrix will equal:

Ego \Alter	Boy	Girl
Boy	-1	1
Girl	1	-1

[9]The requirement that $i_k = i_l$ and $(k-l) \bmod 2 = 0$ ensures that $C = i_k i_{k+1} \dots i_l$ is an alternating cycle (imposing even length). The "furthermore..." requirement ensures that if another node is visited multiple times it does not form an alternative cycle (imposing non-even length). See Figure 4 for an example.

After switching the cycle there are too few same gender links and too many mixed gender ones.

We call a sequence of schlaufen $\mathcal{R} = (R_1, \ldots, R_k)$ feasible if (i) the cycles of the schlaufen are link disjoint and (ii) the sum of their violation matrices is zero (and for $i < k$ the sum of their violation matrices is not zero).

Conventional MCMC adjacency matrix re-wiring algorithms work by switching short cycles (e.g., alternating rectangles and compact alternating hexagons). Switches of this type, while preserving the in- and out-degree sequence of the network will typically generate networks with the wrong inter-group link structure (i.e., non-zero link violation matrices). Our approach to solving this problem involves switching many alternating cycles simultaneously such that their individual link violation matrices sum to zero.

3.2 The MCMC algorithm

Let $\mathbf{S} = \mathbf{s}$ and $\mathbf{M} = \mathbf{m}$ be the degree sequence and cross link matrix of the network in hand. In order to a draw, say G', from $\mathbb{G}_{\mathbf{s},\mathbf{m}}$ we (i) start with a realization of $(\mathbf{s}, \mathbf{m})$, say G, (ii) randomly construct (link disjoint) schlaufen, and (iii) switch any alternating cycles in them. While switching cycles will preserve the degree sequence, it may – as discussed earlier – result in a graph without the appropriate cross link matrix. In order to ensure that G' has the appropriate cross link matrix, we construct schlaufen until either the sum of their violation matrices equals zero or we stop randomly. If the sum of the schlaufen violation matrices is zero we move to G' from G by switching the cycles, otherwise we set $G' = G$. Proceeding in this way ensures that G' is, in fact, a random draw from $\mathbb{G}_{\mathbf{s},\mathbf{m}}$. After sufficiently many iterations of this process we show that a graph constructed in this way corresponds to uniform random draw from $\mathbb{G}_{\mathbf{s},\mathbf{m}}$. A formal statement of the prococedure is provided by Algorithm 1.

Algorithm 1 uses a subroutine to find schlaufen. This subroutine, described in Algorithm 2, finds and marks a schlaufe in the graph.

To illustrate our method in more detail consider the network depicted in Panel A of Figure 4. This network consists of two types of agents: gold (light) and blue (dark). The cross link matrix for the graph is given in Panel D. In Panels B and C a sequence of three schlaufen is shown. The first schlaufen is $R_1 = jgabcdeca$. It is constructed through a sequence of active and passive steps as described earlier (see also the notes to Figure 2 above). We begin by choosing agent j randomly with a probability of $\frac{1}{10}$ (since there are ten agents in the network). We then take an active step, randomly choosing one of the two agents to which j directs a link (i.e., either agent g or i). Here we choose agent g. Next we take a passive step. Specifically we choose an agent at random from the set of agents that *do not* direct a

Algorithm 1 MARKOV DRAW ALGORITHM

Inputs: An adjacency matrix $\mathbf{d} \in \mathbb{D}_{\mathbf{s},\mathbf{m}}$; a mixing time τ

Procedure:

1. Set $t = 0$.
2. With probability $1 - q$ go to step 3, with probability q go to step 4.
3. find and mark a schlaufe (see Algorithm 2):
   (a) **if** the sum of the schlaufen violation matrices is zero, then
      i. switch the cycles in the schlaufen (changing the adjacency matrix $\mathbf{d}$),
      ii. unmark all links,
      iii. go to step 4.
   (b) **else**
      i. with probability $\frac{1}{2}$, go to step 3 or
      ii. with probability $\frac{1}{2}$, unmark all links and go to step 4.
4. Set $t = t + 1$
   (a) **if** $t = \tau$ then return $\mathbf{d}$
   (b) **else** go to step 2

Output: A uniform random draw $\mathbf{d}$ from $\mathbb{D}_{\mathbf{s},\mathbf{m}}$

Algorithm 2 SCHLAUFE DETECTION ALGORITHM

Inputs: An adjacency matrix $\mathbf{d} \in \mathbb{D}_{\mathbf{s},\mathbf{m}}$ (this network may have marked links in it)

Procedure:

1. Choose an agent/node, say i, at random.
2. Mark agent i as active and
   (a) **if** feasible, randomly choose one of $i's$ (unmarked) outlinks, say to j, and go to step 3;
   (b) **else** (i.e., no unmarked outlinks available) go to step 6.
3. Mark edge ij, chosen in step 2 and
   (a) **if** agent j is already marked passive, then go to step 6;
   (b) **else** go to step 4.
4. Mark agent j, chosen in step 3, as passive and
   (a) **if** feasible, randomly choose an agent, say k, from among those who *do not* direct links to j, and go to step 5,
   (b) **else** go to step 6.
5. Mark edge kj, with k the agent chosen in step 4, as passive and
   (a) **if** agent k is already marked active, then go to step 6;
   (b) **else** go to step 2.
6. return the (marked) adjacency matrix, the constructed schlaufe and its violation matrix.

Output: A schlaufe, its violation matrix and a marked adjacency matrix.

link to g (the agent chosen in the previous active step). The probability associated with our choice in this passive step is $\frac{1}{7}$; this corresponds to the reciprocal number of agents in the network (i.e., 10) minus the indegree the current agent (i.e., 2) minus one (since self-loops are not allowed). We continue taking active and passive steps in this way until we visit a for the second time. At this point we stop since our schlaufe now includes the alternating cycle $C_1 = abcdeca$. Note that c is also visited twice, but also that $cdec$ is not an alternating cycle since it is not of even length (see Definition 3.2).

As seen in the example we can calculate the probability of a schlaufe R as we go through the algorithm (see Panel E). In Step 1 of Algorithm 2 an agent is chosen with probability $\frac{1}{N}$. Next let $r_G^a(i)$ be the cardinality of the set of feasible out links in an active step. This set consists of all the out links of node i, which are not already marked in $\mathbf{D}$. Similarly, let $r_G^p(i)$ be the cardinality of the set of feasible outlinks in an passive step. That set consists of all the links ij for which ji is not in $\mathcal{A}(G)$ and which are not already marked. The probability of $R = (i_1, .., i_l)$ can now be written as

$$p_G(R) = \frac{1}{N} \prod_{k=1}^{l-1} \left(\frac{1}{r_G^a(i_k)} \left[k \bmod 2\right] + \frac{1}{r_G^p(i_k)} \left[(k-1) \bmod 2\right] \right) \tag{44}$$

In step 2 of Algorithm 1 we attempt to find a sequence of schlaufen with probability $1 - q$ and do not change the adjacency matrix otherwise. In step 3, a schlaufen sequence $\mathcal{R} = (R_1, .., R_h)$ is constructed/found. After each detected schlaufe in this sequence, say R_k, any cycle in it is marked. Let G_k be the graph with the cycles of $R_1, .., R_{k-1}$ marked. After each schlaufe added the construction is stopped with probability $\frac{1}{2}$. The probability of finding a cycle R_k is $p_{G_k}(R_k)$ as given in equation (44) above. The total probability of a feasible schlaufen sequence $\mathcal{R}$ is therefore

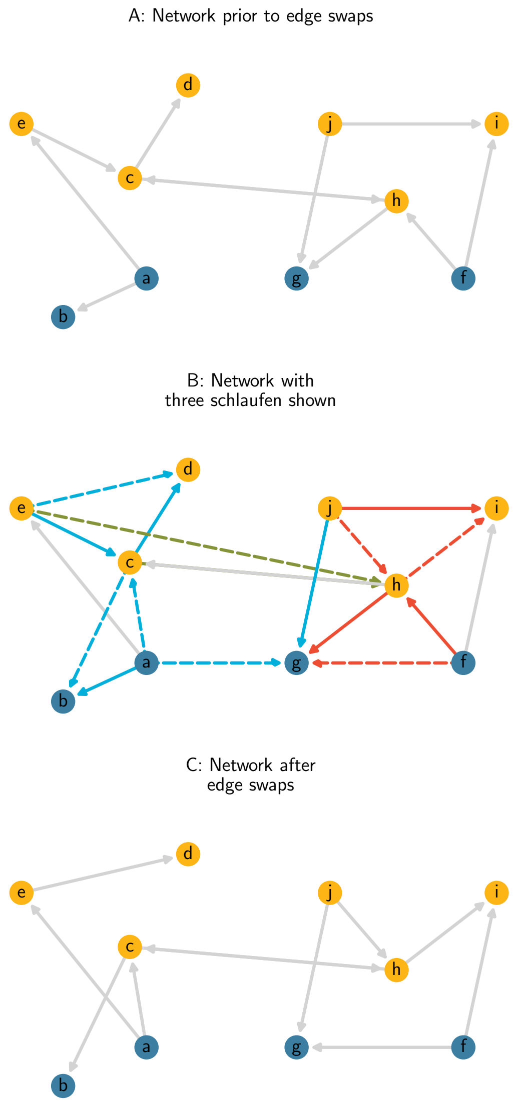
$$p_G(\mathcal{R}) = (1-q)\frac{1}{2^{(h-1)}} \prod_{i=1}^{h} p_{G_k}(R_k). \tag{45}$$

3.3 Correctness

To show that our algorithm does indeed generate a uniform random draw from the set $\mathbb{D}_{\mathbf{s},\mathbf{m}}$ we use standard Markov chain theory (e.g., Chapters 7 and 10 of Mitzenmacher and Upfal (2005)).

The random rewiring of the network implemented by Algorithm 1 can be described as a Markov chain. To show that, for τ large enough, it returns a uniform random draw from $\mathbb{D}_{\mathbf{s},\mathbf{m}}$ we prove that the stationary distribution of the Markov chain generated by Algorithm 1 is uniform on $\mathbb{D}_{\mathbf{s},\mathbf{m}}$. To show this it is helpful to develop a graphical representation of the

Figure 4: A feasible schlaufen sequence



D: Cross-link (PAM) matrix

	Blue	Gold
Blue	1	3
Gold	2	5

E: Three schlaufen

	j	g	a	b	c	d	e	c	a
i	1	2	3	4	5	6	7	8	9
sel. Pr	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{1}{7}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{1}$	$\frac{1}{8}$

	c	h	e
i	1	2	3
sel. Pr	$\frac{1}{10}$	$\frac{1}{1}$	$\frac{1}{7}$

	f	h	j	i	h	g	f
i	1	2	3	4	5	6	7
sel. Pr	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{1}$	$\frac{1}{7}$	$\frac{1}{1}$	$\frac{1}{6}$

F: Violation matrices for the three schlaufen

	Blue	Gold
Blue	-1	+1
Gold	+1	-1

	Blue	Gold
Blue	0	0
Gold	0	0

	Blue	Gold
Blue	+1	-1
Gold	-1	+1

Source: Authors' calculations.
Notes: See the discussion in the main text. The figure depicts three link disjoint schlaufen with violation matrices which sum to zero. Panel E reports the (ex ante) probability that a given node was selected as the schlaufe was constructed. See equation (44).

Markov chain.

We denote the state graph of the Markov chain by $\Phi = (\mathcal{V}_\phi, \mathcal{A}_\phi)$. Its underlying vertex set $\mathcal{V}_\phi$ is the set of all elements in $\mathbb{G}_{\mathbf{s},\mathbf{m}}$. That is each node in our state graph is a network with degree sequence $\mathbf{S} = \mathbf{s}$ and cross link matrix $\mathbf{M} = \mathbf{m}$. For G in $\mathbb{G}_{\mathbf{s},\mathbf{m}}$, we denote by v_G the corresponding vertex in $\mathcal{V}_\phi$. The arc set $\mathcal{A}_\phi$ is defined as follows.

1. For all vertices we add the self loop (v_G, v_G) with (probability) weight q (see Step 2 of Algorithm 1).

2. Let G and G' be two different networks in $\mathbb{G}_{\mathbf{s},\mathbf{m}}$. Let $G\Delta G'$ equal the union of the set of edges in G, but not in G' and the set of edges in G', but not in G. For each feasible schlaufen-sequence $\mathcal{R}$, with cycle edge set equal to $G\Delta G'$ we add the edge $(v_G, v_{G'})$ and assign to it probability weight $p_G(\mathcal{R})$.

3. Finally we add a directed loop (v_G, v_G) if the probability of all arrows leaving v_G, introduced in points 1 and 2 immediately above, do not sum to 1. The probability of this loop is 1 minus the sum of the probability of all other outward arrows.

The probability of any arc $a \in \mathcal{A}_\phi$ is denoted by $p(a)$. Note, by definition, the state graph can have parallel arcs and loops.

With these definitions in place we can prove correctness of the algorithm. First we show that the probability of the algorithm moving from graph G to G' coincides with the probability of moving in the reverse direction.

Lemma 3.1. *For any two vertexes v_G, v'_G the transition probability attached to (v_G, v'_G) equals that attached to (v'_G, v_G).*

Proof. See appendix A.3. □

Next we show the state graph is strongly connected. This means our Algorithm moves from any $G \in \mathbb{G}_{\mathbf{s},\mathbf{m}}$ to any other $G' \in \mathbb{G}_{\mathbf{s},\mathbf{m}}$ with positive probability.

Lemma 3.2. *The state graph Φ is strongly connected.*

Proof. See appendix A.3. □

With these two lemmata it is east to show that the stationary distribution is uniform on $\mathbb{G}_{\mathbf{s},\mathbf{m}}$. This gives us the main result of the section.

Theorem 3.3. *Algorithm 1 is a random walk on the state graph Φ which samples uniformly a network from $\mathbb{G}_{\mathbf{s},\mathbf{m}}$ for $\tau \to \infty$.*

Proof. See appendix A.3. □

We provide an easy to use python implementation of this algorithm as well as the optimal test to a given utility function.
Surce: `https://github.com/AndrinPelican/ugd`
Package: `https://pypi.org/project/ugd/`

4 Applications

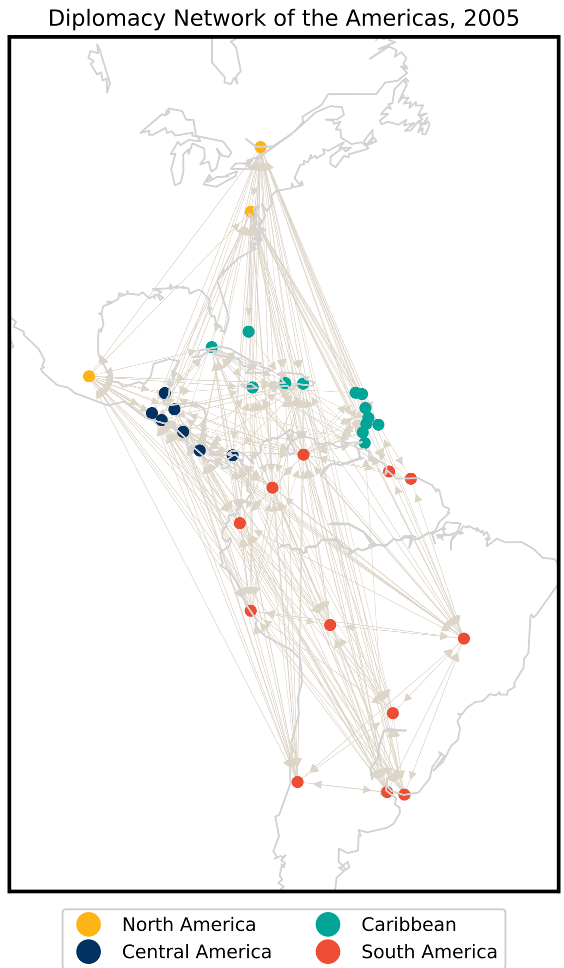
4.1 Diplomatic mission network: the strategic placement of ambassadors

Figure 5 shows the 2005 network of diplomatic exchanges in the Americas. We constructed this network using the Correlates of War Diplomatic exchange dataset (Bayer, 2006). An ij arc indicates that country i has sent an ambassador to country j. As emphasized by Kinne (2014), diplomatic recognition is a core tool of statecraft. Consequently, the decision to establish a diplomatic mission in a country likely has strategic aspects. As a contemporary example, consider the decision to maintain diplomatic ties with Syria after the onset of the Syrian Civil War. This decision appears, in part, to be predicated upon the nature of a state's bi-lateral relations with Syria's long term allies Russia and Iran (along the lines of the "friend of my friend is also my friend" principle).[10]

The diplomatic network is also driven by a desire for prestige and practicality. Some nations, like the United States, send and host many ambassadors. Others send and receive very few. The network also has geographic dimensions: countries typically send ambassadors to close neighbors. These aspects of the network naturally generate network transitivity which is *non-strategic*. For example if most countries send an ambassador to, and host one from, the United States, then virtually any additional tie will generate a transitive triad; but this just reflects the superpower status of the United States, not a structural taste for transitivity in relations. Similarly Central American nations may all host ambassadors from, and send them to, one another due to their strong common cultural, economic and security connections. A desire for transitivity need not play a role. For these reasons it is important to allow for degree heterogeneity and (geographic) homophily when assessing whether nations (actively) prefer transitive ties.

[10] Transitivity can also operate negativity, as when mainland China chooses not recognize countries which diplomatically recognize Taiwan, China. Withholding recognition in this case increases transitivity as not doing so would result in an intransitive triad.

Figure 5: Diplomatic Network of the Americas, 2005



<u>Source:</u> Correlates of War Diplomatic Exchange dataset (Bayer, 2006) and authors' calculations.
<u>Notes:</u> Nodes correspond to capital cities. An arc from country i to country j indicates that an ambassador represents i in j. The network is divided into four regions: (i) North America (California Gold), (ii) Central America (Berkeley Blue), (iii) the Caribbean (Lap Lane) and (iv) South America (Golden Gate)

We consider a utility function of the form given in (2) with $s_{ij}(\mathbf{d})$ as defined in (5). We consider tests of the $\gamma_0 = 0$ null based upon the transitivity index (TI) as well as our locally best test statistic. We also consider a variety of null reference distributions. These different reference distributions illustrate the importance of controlling for degree heterogeneity and homophily in practice.

Panel A of Figure 6 plots the distribution of the transitivity index across three null reference sets. First we consider the set of all networks with density equal to that observed empirically (0.399). The degree sequence and cross link matrices are allowed to freely vary. This reference distribution would be appropriate in the absence of any homophily ($\lambda = 0$) and degree heterogeneity ($A_i = B_j = \alpha$ for all i).

In empirical work it is common to compare the observed TI with network density and conclude that a taste for transitivity is present if the TI is substantially greater than density (the observed TI is shown by the vertical bar in the figure). Here this approach results in an absurdly decisive rejection of the null (observed transitivity equals – by a country mile – the 1.0 quantile of the null distribution); this is a "strawman" test.

Second we consider the set of all networks with in- and out-degree sequences held equal to those observed empirically (i.e., networks where the United States host 32 ambassadors from, and sends 26 to, other countries in the Americas and so on). This controls for heterogeneity in prestige and diplomatic activity across countries. This null distribution, however, continues to assumes the absence of any homophily ($\lambda = 0$).

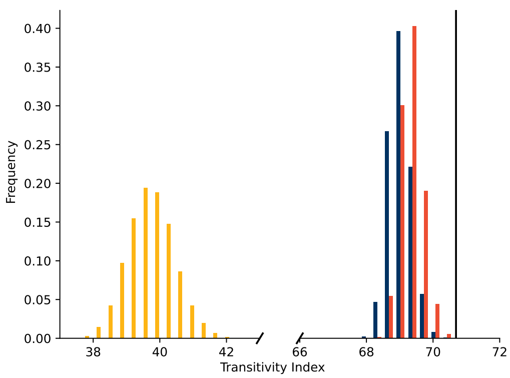
There are several extant methods for simulating uniform draws from this null distribution. For example, Berger and Müller-Hannemann (2009), Kim et al. (2012), Blitzstein and Diaconis (2011), and Del Genio et al. (2010) describe methods for sampling networks with the same degree sequence. Our algorithm easily handles this case as well.

Fixing the degree sequence shifts the null reference distribution substantially to the right. This indicates that much of the observed transitivity in the American diplomatic network can be explained solely by intrinsic variation in baseline diplomatic activity across nations (i.e., degree heterogeneity). While we still easily reject the null in this case as well, the actual TI is much closer to – certainly not a country mile from – the mode of the reference distribution.

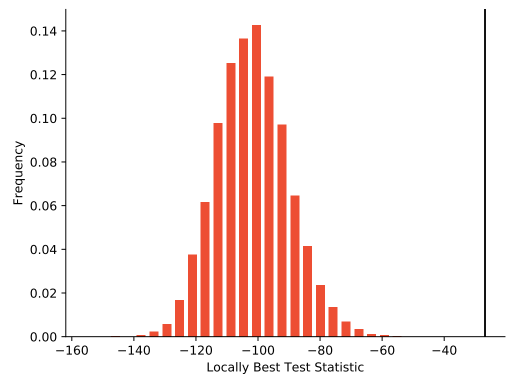
Third, we additionally control for the cross region structure of diplomatic ties (i.e., we now allow $\lambda \neq 0$). We consider the four regions shown in Figure 5. Hence the reference distribution now includes all networks with the same in- and -out degree sequences *and* 4×4 cross link matrix as the observed one. Additionally controlling for geographic homophily further shifts the reference distribution to the right. While we also easily reject the null in this case (p-value $= 2 \cdot 10^{-4}$), the ordering of the three reference distributions demonstrates the potential importance of controlling for degree heterogeneity and homophily in practice.

Figure 6: Testing for strategic transitivity in diplomatic exchanges in the Americas

(a) Transitivity index



(b) Locally best test



Source: Correlates of War Diplomatic Exchange dataset (Bayer, 2006) and authors' calculations.

Notes: Plots of the test statistic null distribution. Panel A: Distribution of the transitivity index (TI) according to three null models. The "California" colored histogram plots the distribution of the TI across the set of networks with density as observed in the empirical network, but no other constraints imposed. The "Berkeley Blue" colored histogram plots the null distribution when the in- and out-degree sequences are additionally held fixed. Finally, the "Golden Gate" colored histogram additionally fixes the structure of cross-region linkage (i.e., the cross link matrix). The actual TI is in, respectively the 1.0, 1.0 and 0.9998 quantiles of the three reference distributions. Panel B: Histogram plot of the null distribution of the locally best test statistic holding the network's degree sequence and cross link matrix fixed. The observed test statistic lies in the 1.0 quantile of the reference distribution. Results based on 10000 simulation draws.

Failing do so could result in mis-interpreted rejections.

Panel B of the figure plots the distribution of our locally best test statistic under the null which fixes both the degree sequence and cross link matrix. The observed locally best test statistic exceeds in value all simulated values from the reference set. While, in this example, it appears to be the case that many "transitivity inspired" statistics would generate rejections, Panel B is suggestive of the power gains associated with using the local best statistics. The locally best statistic is well outside the support of the null distribution. We explore power comparisons further in the Monte Carlo simulations below.

4.2 Buyer-supplier network: the trade-off between direct and indirect customers

Our second example tests for substitutability between direct and indirect customers in a supply chain context. Firms, may, all things equal prefer to have many customers. While there may be a fixed costs to maintaining any customer relationship, having many customers may reduce variation in demand for a firm's products.

Firms may, all things equal, also prefer to supply firms that themselves have many customers. For example, a customer whose own output is widely sought after may generate a more reliable stream of orders than that of a firm with few customers.

In our baseline model, the supplier-specific effect, A_i, captures heterogeneity in firms' demand for customers. Similarly, the buyer-specific effect, B_j, captures heterogeneity in customer attractiveness. If A_i and B_i positively co-vary, as is allowed by our model, then firms with many customers (i.e., high A_i firms) may be viewed as especially attractive customers themselves (i.e., high B_i firms). These features of our baseline model suggest that it is not suitable for detecting whether firms value many direct (or indirect customers) per se; however we can test for *substitutability* between direct and indirect customers.

To construct such a test we set $s_{ij}(\mathbf{d})$ equal to

$$s_{ij}\left(\mathbf{d}\right) = \left(\sum_{k \neq j} d_{ik}\right)\left(\sum_{k} d_{jk}\right) \overset{def}{\equiv} d_{i+,-j} \times d_{j+} \tag{46}$$

where $d_{i+,-j}$ is notation for the number of customers firm i has excluding any customer relationship with j. With this specification of $s_{ij}(\mathbf{d})$ we can interpret γ_0 in terms of the "cross derivative"

$$\gamma_0 = \frac{\partial^2 MU_{ij}}{\partial d_{i+,-j} \partial d_{j+}}. \tag{47}$$

Here we hypothesize that $\gamma_0 < 0$, so that firms with many *direct* customers value the indirect customers of a firm less than firms with few direct customers.

We implement our test using three industry-specific supply chain networks we constructed using Compustat data: pharmaceuticals (see Figure 1), computers, and motor vehicles (see Panels A and B of Figure 7). We test for whether the marginal benefit of an indirect customer is decreasing in the number of direct customers in each industry separately and using all three networks simultaneously. Pooling is straightforward in our set-up: the broad 4-digit manufacturing sector constitutes an observed firm attribute which is incorporated into the cross link matrix constraint.

The p-value for the null of $\gamma = 0$ equals 0.02 for pharmaceuticals, 0.32 for computers and 0.10 for motor vehicles. Pooling all three networks together yields a p-value of 0.05. These p-values are based upon 1000 simulated networks. The mixing time is chosen, such that each edge is, on average, randomly modified 10 times before the network is considered a random draw from the target set.

4.3 Monte Carlo experiments

For the Monte Carlo experiments we work with the general utility function introduced in Section 1 above. We assume that $A_i \in \mathbb{A} \overset{def}{\equiv} \{\alpha_L, \alpha_H\}$, $B_i \in \mathbb{B} \overset{def}{\equiv} \{\beta_L, \beta_H\}$ and $X_i \in \mathbb{X} \overset{def}{\equiv} \{0, 1\}$. We assume that each support point in $\mathbb{A} \times \mathbb{B} \times \mathbb{X}$ occurs with equal probability (i.e., with probability equal to $\frac{1}{8}$).

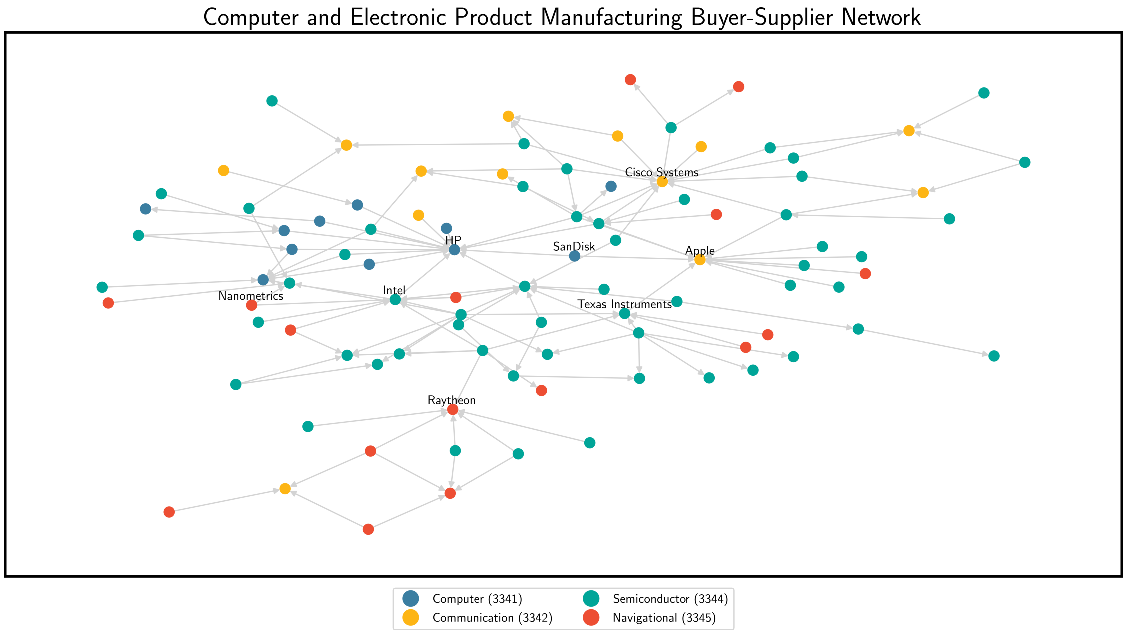
Observe that their are four types of sending agents: $(A_i = \alpha_L, X_i = 0)$, $(A_i = \alpha_H, X_i = 0)$, $(A_i = \alpha_L, X_i = 1)$ and $(A_i = \alpha_H, X_i = 1)$. Similarly there are four types of receiving agents. The null model is therefore fully described by $16 = 4 \times 4$ linking probabilities. We set the parameters $\beta_H = \alpha_H = 1.1$, $\alpha_L = \beta_L = -1.1$, $\lambda_{00} = \lambda_{11} = 0$ and $\lambda_{01} = \lambda_{10} = -2.2$. Our parameter choices generate meaningful degree heterogeneity and homophily under the null. Across $1,000$ Monte Carlo simulations with $N = 48$ average network density was 0.34, average transitivity was 0.53, and the average standard deviation of, respectively, in- and out-degree, was 4.1.

We set the strategic interaction term to $s_{ij}(\mathbf{d}) = \sum_k d_{ik} d_{kj}$; as is appropriate when agents prefer transitive ties. To simulate a network under the alternative we draw $\mathbf{U}$ and then, starting with an empty adjacency matrix, iterate to a fixed point using equation (11). By Tarski's Theorem this finds us the least dense pure strategy Nash Equilibrium.

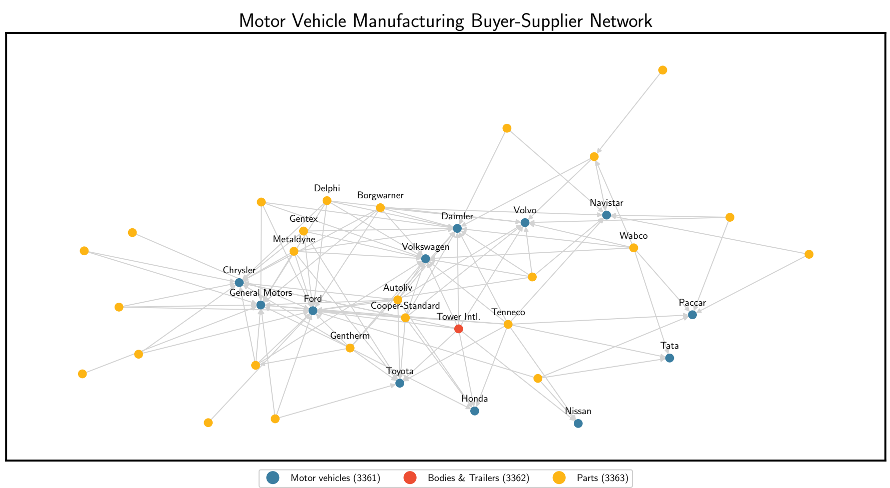
We compare the performance of three tests. The infeasible locally best test that is based upon the true value of δ_0. The feasible version of this test which replaces δ_0 with its maximum likelihood estimate computed under the null (see Graham (2017), Dzemski (2018) and Yan et al. (2018) for a discussion of this particular MLE problem). Finally we construct an *ad hoc* test based upon the transitivity index (TI). This last test is the one most often used in

Figure 7: Buyer-supplier networks used to test for substitutability between direct and indirect buyers

(a) Computers Buyer-Supplier Network, 2015



(b) Motor Vehicles Buyer-Supplier Network, 2015



Source: Compustat and authors' calculations.
Notes: Plots of the computers and motor vehicles buyer-supplier networks in 2015 based upon Compustat data. The head of each arc denotes the buying firm. Nodes colored differently according to their sub-industry as listed in the legends. Largest weakly-connected component is shown.

Table 1: Monte Carlo Design Parameterization

Calibrated Link	Probability
$F(\alpha_H + \beta_H + \lambda_{00})$	0.90
$F(\alpha_H + \beta_L + \lambda_{00})$	0.50
$F(\alpha_L + \beta_L + \lambda_{00})$	0.10
$F(\alpha_H + \beta_H + \lambda_{01})$	0.50
$F(\alpha_H + \beta_L + \lambda_{01})$	0.10
$F(\alpha_L + \beta_L + \lambda_{01})$	0.012

Source: Authors' calculations.
Notes: Utility parameters for the Monte Carlo experiments were chosen to calibrate the null link probabilities listed in Column 1 to equal those values listed in Column 2.

practice.

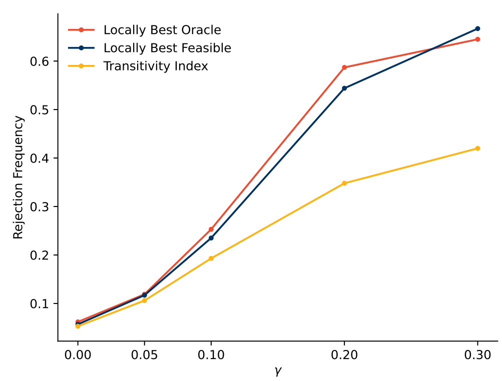
Figure 8 summarizes our main findings. The two panels of the figure correspond to the two network sizes we considered: $N = 24$ and $N = 48$. The horizontal axes of the figures in each panel correspond to different values of the strategic interaction parameter, γ_0; the vertical axes to rejection frequencies. With 1000 Monte Carlo replications the standard error of our simulation estimate of size is $\sqrt{(0.05\,(1-0.05)\,/1000)} \approx 0.007$.

As expected, the actual size of our test is indistinguishable (i.e., equal up to simulation error) from its nominal size. For the designs considered here the power gains associated with using the locally best test statistic derived in Section 2 are considerable. Furthermore the feasible locally best test, which replaces δ_0 with its MLE (computed under the null), performs about as well as the infeasible locally best test based on the actual value of δ_0.

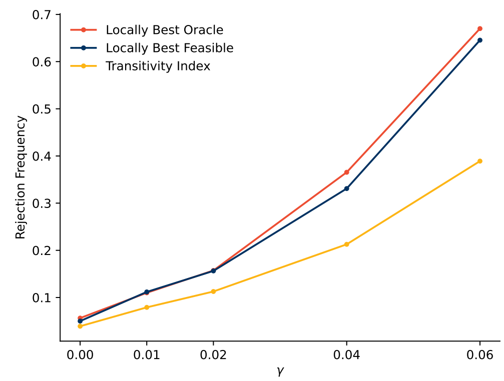
The Monte Carlo experiments highlight that the locally best test, which upweights episodes of "unexpected" transitivity, is more powerful than the ad hoc test based on comparing the transitivity index (TI) with its null distribution. Note both tests are "valid" and correctly-sized.

Figure 8: Power Analysis

(a) $N = 24$ agents/nodes



(b) $N = 48$ agents/nodes



Source: Authors' calculations.
Notes: The figures plot the frequency with which $H_0 : \gamma_0 = 0$ is rejected across 1,000 Monte Carlo replications for, respectively, networks with $N = 24$ and 48 agents. The y-axis reports the estimated rejection frequency, the x-axis gives the value of the strategic interaction parameter, γ_0. The sparsest pure strategy NE is use to simulate each network. For each simulation a total of 400 MCMC draws from $\mathbb{D}_{\mathbf{s},\mathbf{m}}$ were used to compute critical values. The mixing time was chosen such that every edge is randomly modified before the network is considered a uniform draw. The marginal utility function equals $A_i + B_j + X_i' \Lambda_0 X_i + \gamma_0 s_{ij}(\mathbf{d}) - U_{ij}$ with $s_{ij}(\mathbf{d}) = \sum_k d_{ik} d_{kj}$. The distribution of (A_i, B_i, X_i) is as described in the main text; other model parameters are given in Table 1.

5 Extensions

Graham and Pelican (2020) introduced an econometric model for network formation with transferable utility (cf., Bloch and Jackson, 2007); appropriate for the analysis of undirected networks. They use the importance sampling algorithm of Blitzstein and Diaconis (2011) to simulate the null distribution of their test statistic. Their set-up does not allow for homophily under the null; nor do they consider optimal test statistics. These extension could be accomplished use the simulation algorithms for undirected networks developed in Pelican (2019) and the ideas of this paper.

Our approach to testing could also be applied to settings where the econometrician observes many independent games, each with a small number of players (see de Paula (2013) for a review); this arises when a small set of competing firms make entry decisions across a large number of independent markets. Testing here is relatively better studied (e.g., Chen et al. (2018) and the references therein). An attractive feature of our framework for empirical researchers is that is can to easily handle, for example, "market level" unobservables (albeit at the cost of assuming logistic utility shocks).

A primary advantage of our approach – complete agnosticism about equilibrium selection under the null – is also a limitation. It is not obvious how to adapt our method to, for example, construct an identified set for γ_0 (or a confidence interval for this set). Research in this direction would be useful.

While much additional research remains to be done, we have provided a feasible and powerful test for a key scientific hypothesis on the nature of network formation, whilst maintaining a rich and realistic null model structure. Our MCMC simulation algorithm is also likely to be of independent interest.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981 – 2014.

Anderson, J. E. (2011). The gravity model. *Annual Review in Economics*, 3(1):133 – 160.

Andrews, I., Stock, J. H., and Sun, L. (2019). Weak instruments in instrumental variables regression: theory and practice. *Annual Review of Economics*, 11:727 – 753.

Apicella, C. L., Marlowe, F. W., Fowler, J. H., and Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature*, 481(7382):497 – 501.

Atalay, E., Hortaçsu, A., Roberts, J., and Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199 – 5202.

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, 68(5):1181 – 1229.

Barabási, A.-L. (2016). *Network Science*. Cambridge University Press, Cambridge.

Bayer, R. (2006). Diplomatic exchange data set, v2006.1. Technical report, The Correlates of War Project.

Berger, A. and Müller-Hannemann, M. (2009). Uniform sampling of undirected and directed graphs with a fixed degree sequence. *CoRR*, abs/0912.0685.

Bickel, P., Choi, D., Chang, X., and Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Annals of Statistics*, 41(4):1922 – 1943.

Blitzstein, J. and Diaconis, P. (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4):489 – 522.

Bloch, F. and Jackson, M. O. (2007). The formation of networks with transfers among players. *Journal of Economic Theory*, 113(1):83 – 110.

Boss, M., Elsinger, H., Summer, M., and Thurner, S. (2004). Network topology of the interbank market. *Quantitative Finance*, 4(6):677 – 684.

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

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

Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability*, 21(4):1400 – 1435.

Chen, X., Christensen, T. M., and Tamer, E. (2018). Monte carlo confidence sets for identified sets. *Econometrica*, 86(6):1965 – 2018.

Cox, D. R. (2006). *Principles of Statistical Inference*. Cambridge University Press, Cambridge.

Czabarka, E., Szekely, L. A., Toroczkai, Z., and Walker, S. (2017). An algebraic monte-carlo algorithm for the bipartite partition adjacency matrix realization problem. *arXiv preprint arXiv:1708.08242*.

de Paula, A. (2013). Econometric analysis of games with multiple equilibria. *Annual Review of Economics*, 5:107–131.

de Paula, Á., Richards-Shubik, S., and Tamer, E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, 86(1):263 – 288.

Del Genio, C. I., Kim, H., Toroczkai, Z., and Bassler, K. E. (2010). Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *Plos One*, 5(4):1–7.

Dutta, B. and Jackson, M. O. (2000). The stability and efficiency of directed communication networks. *Review of Economic Design*, 5(3):251 – 272.

Dzemski, A. (2018). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics*. University of Mannheim.

Erdős, P., G. Hartke, S., van Iersel, L., and Miklós, I. (2017). Graph realizations constrained by skeleton graphs. *Electronic Journal of Combinatorics*, 24(2):1 – 18.

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.

Gao, C., Lu, Y., and Zhou, H. H. (2015). Rate-optimal graphon estimation. *Annals of Statistics*, 43(6):2624 – 2652.

Goyal, S. (2009). *Connections: An Introduction to the Economics of Networks*. Princeton University Press, Princeton, NJ.

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

Graham, B. S. (2020). *Handbook of Econometrics*, volume 7A, chapter Network data. North-Holland, Amsterdam.

Graham, B. S. and Pelican, A. (2020). *The Econometric Analysis of Network Data*, chapter Testing for externalities in network formation using simulation, pages 65 – 82. Academic Press, London.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1360 – 1380.

Jackson, M. O., Rogers, B. W., and Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, 55(1):49 – 95.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44 – 74.

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business and Economic Statistics*, 36(4):705 – 713.

Kaido, H. and Zhang, Y. (2019). Robust likelihood ratio tests for incomplete economic models. arXiv Working Paper arXiv:1910.04610v2 [econ.EM], Boston Univerity and Jinan University.

Kim, H., Genio, C. I. D., Bassler, K. E., and Toroczkai, Z. (2012). Constructing and sampling directed graphs with given degree sequences. *New Journal of Physics*, 14(2):023012.

Kinne, B. J. (2014). Dependent diplomacy: signaling, strategy, and prestige in the diplomatic network. *International Studies Quarterly*, 58(2):247 – 259.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data*. Springer, New York.

König, M. D., Liu, X., and Zenou, Y. (2019). R&d networks: theory, empirics and policy implications. *Review of Economics and Statistics*, 101(3):476 – 491.

McDonald, J. W., Smith, P. W. F., and Forster, J. J. (2007). Markov chain monte carlo exact inference for social networks. *Social Networks*, 29(1):127 – 136.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27(1):415 – 444.

Menzel, K. (2016). Strategic network formation with many agents. New York University.

Mitzenmacher, M. and Upfal, E. (2005). *Probability and Computing*. Cambridge University Press, Cambridge.

Miyauchi, Y. (2016). Structural estimation of a pairwise stable network with nonnegative externality. *Journal of Econometrics*, 195(2):224 – 235.

Moreira, M. J. (2009). Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics, 152(2), 131-140*, 152(2):131 – 140.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48 – 49.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press, Oxford.

Pelican, A. (2019). Uniform sampling of graphs with fixed degree sequence under partition constraints.
https://www.fernuni-hagen.de/MATHEMATIK/DMO/pubs/Master_Andrin_Pelican.pdf. Accessed: 2020-01-01.

Rao, A. R., Jana, R., and Bandyopadhyay, S. (1996). A markov chain monte carlo method for generating random (0,1)-matrices with given marginals. *Sankhya*, 58(2):225 – 242.

Sheng, S. (2014). A structural econometric analysis of network formation games. Mimeo, University of California - Los Angeles.

Sinclair, A. (1993). *Algorithms for Random Generation and Counting: A Markov Chain Approach*. Progress in Theoretical Computer Science. Birkhäuser, Boston.

Tao, T. (2016). An improved mcmc algorithm for generating random graphs from constrained distributions. *Network Science*, 4(1):117 – 139.

Tarski, A. (1955). A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285 – 309.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. Twentieth Century Fund, New York.

Werner, K. (2012). *Maß- und Integrationstheorie*. FernUniversität in Hagen.

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. (2018). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association*.

A Supplemental Web Appendix: Proofs

The appendix includes proofs of the theorems stated in the main text as well as statements and proofs of supplemental lemmata. All notation is as established in the main text unless stated otherwise. Equation number continues in sequence with that established in the main text.

A.1 Proof of Theorem 1.1

For the equation (13) to be well defined we must show that $\mathcal{N}(\mathbf{d}, \cdot; \theta)$ is measurable. For a network $\mathbf{d}$ we can define a function $\mathcal{N}(\mathbf{d}, \cdot; \theta)$, which assigns to a realization $\mathbf{U} = \mathbf{u}$ the Nash equilibrium weight of the pure strategy which corresponds to $\mathbf{d}$. We now show that there is a measurable function $\mathcal{N}(\mathbf{d}, \cdot; \theta)$ satisfying these conditions.

First we consider the case where the utility shocks are bounded. Let $M > 1$ be given, we show that $\mathcal{N}(\mathbf{d}, \cdot; \theta)$ is measurable on $\mathbb{U}^c = [-M, M]^n$. Observe that every realization of the taste shock $\mathbf{u}$ corresponds to a game in normal form γ. Here we use γ to denote a table containing, for each pure strategy combination (or equivalently network) $\mathbf{d}$, the utility of each agent according to equation (2). We use γ to denote this table to be consistent with the game theory literature (recognizing this may be confusing since γ is used in the main text, and elsewhere in this appendix, to denote the strategic interaction parameter of parameter of interest.

Utility is defined for every player. The mapping $g : \mathbb{R}^n \to \mathbb{R}^{N2^n}$ assigns to each taste shock realization $\mathbf{u}$ the corresponding game $g(\mathbf{u})$. In a game each player can choose among 2^{N-1} strategies (corresponding to which of the $N-1$ other agents she chooses to link to); thus there are $2^{(N-1)N}$ pure strategy combinations.

Looking at equation (2) it is apparent that g is continuous and therefore measurable. Because of continuity we have $\sup\{||g(\mathbf{u})|| : \mathbf{u} \in [\mathbf{-M}, \mathbf{M}]^{\mathbf{N(N-1)}}\} := \mathbf{L} < \infty$. We set $\Gamma := [-L, L]^{N2^n}$; the set of all games with bounded payoffs.

Lemma A.1. *Let Σ be the set of all mixed strategies combinations of the players. The set*

$$E := \{(\gamma, \sigma) : \sigma \textit{ is a NE of } \gamma \in \Gamma\} \subset \Gamma \times \Sigma \tag{48}$$

is compact.

Proof. E is bounded. Thus it is sufficient to show that E is closed. The NE is defined over a set of inequalities, which have to be fulfilled (each player cannot strictly increase her payoff by replacing their strategy with any other pure strategy, holding the other players strategies constant). The utilities are continuous functions on $\Gamma\times\Sigma$. Now assume $x\in\Gamma\times\Sigma$ is not in E, then there exists a inequality which is not satisfied. The inequality is violated by $\mathbf{u}$. Because the function on both sides of the inequality are continuous we can choose a δ environment of x such that it is violated for all the elements in the environment. Therefore the complement of E is open, which proves the statement. □

Lemma A.2. *Let $\Gamma\subset\mathbb{R}^n$ and $\Sigma\subset\mathbb{R}^m$, $E\subset\Gamma\times\Sigma$ be compact sets. Further*

$$\forall x\in\Gamma\exists y\in\Sigma:(x,y)\in E. \tag{49}$$

There is a measurable function $f:\Gamma\to\Sigma$ with

$$\forall x\in\Gamma:(x,f(x))\in E. \tag{50}$$

Proof. Because of compactness there are $a,b\in\mathbb{R}$ such that $\Gamma\times\Sigma\subset[a,b)^{n+m}$. For each k we partition $[a,b)$ in 2^k intervals: $[a,a+\frac{b-a}{2^k}),..,[b-\frac{b-a}{2^k},b)$ and correspondingly partition $[a,b)^{n+m}$ into hyper-cubes with side length $\frac{b-a}{2^k}$. To a cube $[a_1,b_1)\times[a_2,b_2)\times..\times[a_h,b_h)$ we associate the *characteristic* vector $(a_1,a_2,..,a_h)$. We order vectors such that $a^1>a^2$ if the first coordinate i, for which entry of the two vectors is not equal is $a^1_i>a^2_i$. The ordering of the vectors implies an ordering of the boxes by their characteristic vectors. Let $\mathcal{C}^k$ be the set of hyper-cubes constructed as described above with side length $\frac{b-a}{2^k}$ covering $[a,b)^{n+m}$. Let $p:\mathbb{R}^{n+m}\to\mathbb{R}^n$ be the projection.

Now we define $f_k:\Gamma\to\Sigma$ as follows. Select $x\in\Gamma$ find all cubes in $C\in\mathcal{C}^k$ with $x\in p(C)$ and $C\cap E\neq\emptyset$. We denote these cubes as $\mathcal{C}_x$. We call $x_1,x_2\in\Gamma$ equivalent if and only if $\mathcal{C}_{x_1}=\mathcal{C}_{x_2}$. In this way Γ is partitioned into finitely many equivalence classes. By (49) $\mathcal{C}_x$ is not empty. If more than one cube is in $\mathcal{C}_x$ chose C, the highest one in the sense of the ordering defined above. We select an arbitrary vector $v\in C\cap E$ and set $f_k(x'):=(v_{n+1},v_{n+2},..,v_{n+m})$ for all x' equivalent to x.

$f_k(\Gamma)$ is finite. And each element of $f_k(\Gamma)$ has as a preimage the intersection of Γ with a finite number of n dimensional cubes. Because Γ is measurable (because of compactness) and the n dimensional cube are measurable, f_k is a measurable function. Note f_k does not have property (50).

We now want to show pointwise convergence for $(f_k)_{k\in\mathbb{N}}$. We fix $x\in\Gamma$ and have a sequence $(x^k,y^k)=(x,f_k(x))$ in $\Gamma\times\Sigma$. Because $\Gamma\times\Sigma$ is compact, (x^k,y^k) has a convergent subsequence (x^{k_z},y^{k_z}) with limit (x^*,y^*). Let $||v||$ be the maximum norm of a vector $v\in\mathbb{R}^{n+m}$. Assume $(x^*,y^*)\notin E$, then $\inf\{||(x^*,y^*)-e||:e\in E\}:=\epsilon>0$ because E is compact. Now we can choose z (and thereby k) such that the side length of the cube is $\frac{\epsilon}{2}$. Then the cube containing (x^*,y^*) does not contain any $e\in E$. A contradiction of the way f_k is constructed.

Now assume that (x^k, y^k) has subsequence converging to point $(\tilde{x}^*, \tilde{y}^*) \neq (x^*, y^*)$. Without loss of generality assume $(\tilde{x}^*, \tilde{y}^*) < (x^*, y^*)$. Let i be the first index of y in which the two vectors don't equal. We have $\tilde{y}_i^* < y_i^*$. Therefore $|\tilde{y}_i^* - y_i^*| =: \epsilon > 0$. Now we can choose a threshold K such that for any $k > K$ the side length of the cube is less then $\frac{\epsilon}{2}$. Then for all $k > K$ the cube containing $(\tilde{x}^*, \tilde{y}^*)$ and the cube containing (x^*, y^*), do not intersect. The cube containing (x^*, y^*) is higher in the ordering then the cube containing $(\tilde{x}^*, \tilde{y}^*)$. Therefore for all $k > K$ (x^k, y^k) is not in a $\frac{\epsilon}{2}$ environment of $(\tilde{x}^*, \tilde{y}^*)$. A contradiction to another limit of the sequence. This shows the pointwise convergence for $(f_k)_{k\in\mathbb{N}}$. The limit function we denote with f^*.

For a sequence of measurable functions, which converges pointwise, the limit function is also measurable Werner (2012). f^* is measurable and satisfies condition (50). □

Proof of Theorem 1.1

Now we are ready to prove the result stated in the main text.

For a $z \in \mathbb{Z}^n$ we define a function $r^z : [z_1, z_1 + 1] \times ... \times [z_n, z_n + 1] \to \Sigma$, which assigns to each taste shock in the hyper cube $[z, z + 1]$ a NE. According to Lemma A.1 and Lemma A.2 and the fact that g from above is measurable we can define r^z such that it is measurable. We now define $\bar{r}^z : \mathbb{R}^n \to \mathbb{R}^{N2^n}$ with $\bar{r}^z(x) = r^z(x)$ if $x \in [z, z + 1)$ and $\bar{r}^z(x) = 0$ otherwise. Let $r : \mathbb{R}^n \to \Sigma$ with

$$r(x) = \sum_{z\in\mathbb{Z}^n} \bar{r}^z(x) \tag{51}$$

Note in equation (51) we sum over a countable set and and the sum converges absolutely for each x. Therefore r is measurable; specifically it is a measurable function which assigns to each taste shock to a NE. Let $h_{\mathbf{d}} : \Sigma \to [0, 1]$ be the function which assigns to every mixed strategy the probability of $\mathbf{d}$ by multiplying the mixed strategies weights corresponding to $\mathbf{d}$. Since multiplication is a measurable operation h_G is measurable. $\mathcal{N}(\mathbf{d}, \cdot; \theta) := h_{\mathbf{d}} \circ r$ satisfies the desired properties.

A.2 Optimal test statistic proofs

Preliminary results

Lemma A.3. *Any differentiable function $f \in \mathcal{O}(\gamma^2)$ with $f(0) = 0$ has a derivative of zero at point zero.*

Proof. For $f \in \mathcal{O}(\gamma^2)$ we have, for some $C > 0$ and $\epsilon > 0$, that

$$|f(\gamma)| < C\gamma^2 \tag{52}$$

for all $\gamma \in [-\epsilon, \epsilon]$. The derivative of f at $\gamma = 0$ equals

$$f'(0) = \lim_{\gamma \to 0} \frac{f(\gamma) - f(0)}{\gamma} = \lim_{\gamma \to 0} \frac{f(\gamma)}{\gamma}, \tag{53}$$

with the second equality because $f(0) = 0$. As $\gamma \to 0$, we will have $\gamma < \epsilon$ so that

$$f'(0) = \lim_{\gamma \to 0} \frac{f(\gamma)}{\gamma} \leq \lim_{\gamma \to 0} \frac{C\gamma^2}{\gamma} = \lim_{\gamma \to 0} C\gamma \tag{54}$$

which goes to zero as $\gamma \to 0$ as claimed. $\square$

Proof of Theorem 2.1 (i.e., derivation of the form of the locally best test statistic)

We begin with the likelihood decomposition (35) given in the main text. The number of summands in (35) depends on the partition that $s_{ij}(\mathbf{d})$ induces on $\mathbb{R}$. For a positive γ, the number neither depends on the exact value of γ, nor on the other covariates and parameters. Intuitively, as long as γ is positive, there is a positive probability that $\mathbf{U}$ falls in any combination of buckets. The number of summands in (35) is typically large. The buckets $\mathbf{b}$ of $\mathbb{B}^n$ and the function $\mathcal{N}$ depend on γ.

We have that

$$\begin{aligned}
\frac{\partial P(\mathbf{d}; \theta, \mathcal{N})}{\partial \gamma} &= \frac{\partial}{\partial \gamma} \left\{ \sum_{\mathbf{b} \in \mathbb{B}^n} \int_{\mathbf{u} \in \mathbf{b}} \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) f_{\mathbf{U}}(\mathbf{u}) \, \mathrm{d}\mathbf{u} \right\} \\
&= \sum_{\mathbf{b} \in \mathbb{B}^n} \frac{\partial}{\partial \gamma} \int_{\mathbf{u} \in \mathbf{b}} \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) f_{\mathbf{U}}(\mathbf{u}) \, \mathrm{d}\mathbf{u},
\end{aligned}$$

The switching of summation and derivative operator is possible because the number of summands does not depend on γ. We could try to take the derivative of each summands integral boundaries and of $\mathcal{N}(\mathbf{d}, .; \theta)$. But there is no need to boil the ocean, because regardless of $\mathcal{N}(\mathbf{d}, .; \theta)$ most of the summands are 0. To show this we consider three cases of summands.

Case 1: more than two buckets in B are inner buckets

Recall that the boldface subscripts $\mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots$ index the $n = N(N-1)$ directed dyads in arbitrary order. Consider a set of buckets $\mathbf{b}$ where two or more of them are inner buckets. Without loss of generality assume that the $L \geq 2$ inner buckets correspond to $b_{\mathbf{1}}, \ldots, b_{\mathbf{L}}$ of $\mathbf{b} = (b_{\mathbf{1}}, \ldots, b_{\mathbf{n}})$. The shape of the $\mathbf{l}^{th}$ bucket is $(\gamma \underline{s}_{\mathbf{l}}, \gamma \bar{s}_{\mathbf{l}}]$ with $\underline{s}_{\mathbf{l}} < \bar{s}_{\mathbf{l}}$ coinciding with the bucket borders induced by the precise form of strategic interaction specified under the alternative. We normalize the dyad-specific systematic utility component $\mu_{ij} = 0$ without loss of generality.

Recall that $\tilde{\mathbb{B}}^n$ is the set of bucket configurations with two or more inner buckets. For any $\mathbf{b} \in \tilde{\mathbb{B}}^n$ we can derive the upper bound:

$$\begin{aligned}
\int_{\mathbf{u}\in\mathbf{b}} \mathcal{N}(\mathbf{d},\mathbf{u};\theta) f_{\mathbf{U}}(\mathbf{u})\,\mathrm{d}\mathbf{u} &= \int_{\mathbf{u}\in\mathbf{b}} \mathcal{N}(\mathbf{d},\mathbf{u};\theta)\left[\prod_{\mathbf{i}} f_U(u_{\mathbf{i}})\right]\mathrm{d}\mathbf{u} \\
&\leq \int_{\gamma\underline{s}_{\mathbf{1}}}^{\gamma\overline{s}_{\mathbf{1}}} f_U(u_{\mathbf{1}}) \times \cdots \times \int_{\gamma\underline{s}_{\mathbf{L}}}^{\gamma\overline{s}_{\mathbf{L}}} f_U(u_{\mathbf{L}}) \int_{\mathbf{u}_{-L}\in\mathbf{b}_{-L}} f_{\mathbf{U}_{-L}}(\mathbf{u}_{-L})\,\mathrm{d}\mathbf{u} \\
&< \int_{\gamma\underline{s}_{\mathbf{1}}}^{\gamma\overline{s}_{\mathbf{1}}} f_U(u_{\mathbf{1}}) \times \cdots \times \int_{\gamma\underline{s}_{\mathbf{L}}}^{\gamma\overline{s}_{\mathbf{L}}} f_U(u_{\mathbf{L}})\,\mathrm{d}u_{\mathbf{1}}\cdots\mathrm{d}u_{\mathbf{L}} \\
&< \int_{\gamma\underline{s}_{\mathbf{1}}}^{\gamma\overline{s}_{\mathbf{1}}} 1 \times \cdots \times \int_{\gamma\underline{s}_{\mathbf{L}}}^{\gamma\overline{s}_{\mathbf{L}}} 1\mathrm{d}u_{\mathbf{1}}\cdots\mathrm{d}u_{\mathbf{L}} \\
&= \gamma^L\left(\overline{s}_{\mathbf{1}} - \underline{s}_{\mathbf{1}}\right) \times \cdots \times \left(\overline{s}_{\mathbf{L}} - \underline{s}_{\mathbf{L}}\right)
\end{aligned}$$

where $\mathbf{u}_{-L}$ denotes the vector $\mathbf{u}$ after removal of its first L components and similarly for $\mathbf{b}_{-L}$. The first equality follows from independence of the components of $\mathbf{u}$, the second (weak) inequality from the fact that $\mathcal{N}(\mathbf{d},\mathbf{u};\theta) \leq 1$ for all $\mathbf{u} \in \mathbb{U}$. The third (strict) inequality follows because $f_{\mathbf{U}_{-L}}(\mathbf{u}_{-L})$ is a density and the integration is not over all of $\mathbb{R}^{n-L}$. The fourth (strict) inequality arises because when $f_U(u)$ is the logistic density we have that $f_U(u) = F_U(u)[1 - F_U(u)] < 1$ for all u on a compact interval of the real line. We conclude that any summand where $\mathbf{b}$ has two or more inner buckets is $\mathcal{O}(\gamma^2)$ for $\gamma \to 0$.

We have, directly from this argument, that $Q(\mathbf{d};\theta,\mathcal{N}) \in \mathcal{O}(\gamma^2)$ and furthermore that $Q\left(\mathbf{d};(0,\delta')',\mathcal{N}\right) = 0$ (since inner buckets have zero probability when $\gamma = 0$). Hence, by Lemma A.3, we have that

$$\left.\frac{\partial Q(\mathbf{d};\theta,\mathcal{N})}{\partial\gamma}\right|_{\gamma=0} = 0.$$

This is enough to show equation (40) of the main text. This simplification is essential to the overall result, as it allows us to proceed without knowing any details about the form of the equilibrium selection rule $\mathcal{N}$ when $\mathbf{U}$ takes values which admit multiple equilibrium networks.

Case 2: No bucket in b is an inner bucket (i.e., all buckets are outer buckets)

If all components of $\mathbf{u}$ fall in either their first or last buckets, then the network is uniquely defined. This occurs because agent-level preferences for forming (or not forming) a link are so strong that they do not depend on the presence or absence of other links in the network.

Each agent i either prefers to send a link to j, regardless of the actions taken by others, or does not wish to send a link. Put differently, each agent has a pure link formation strategy which is strictly dominating in such games; therefore $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ is either zero or one.

For a particular network $\mathbf{d}$, $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) = 1$ if, for all (directed) dyads ij such that $d_{ij} = 1$, we have that u_{ij} falls in the first bucket and for all dyads ij such that $d_{ij} = 0$ we have that u_{ij} falls in the last bucket. These considerations give the equality

$$\int_{\mathbf{u}\in\mathbf{b}} \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) f_{\mathbf{U}}(\mathbf{u}) \,\mathrm{d}\mathbf{u} = \prod_{i\neq j} \left[\int_{-\infty}^{\mu_{ij}+\gamma\underline{s}} f_U(u_{ij}) \,\mathrm{d}u_{ij}\right]^{d_{ij}} \left[\int_{\mu_{ij}+\gamma\bar{s}}^{\infty} f_U(u_{ij}) \,\mathrm{d}u_{ij}\right]^{1-d_{ij}} \tag{55}$$

$$= \prod_{i\neq j} \left[F_U(\mu_{ij} + \gamma\underline{s})\right]^{d_{ij}} \left[1 - F_U(\mu_{ij} + \gamma\bar{s})\right]^{1-d_{ij}} \tag{56}$$

Taking logarithms of the expression above, differentiating with respect to γ, evaluating at $\gamma = 0$, and multiplying by $P_0(\mathbf{d}; \delta)$ yields a derivative for summands where all buckets in $\mathbf{b}$ are outer buckets of

$$P_0(\mathbf{d}; \delta) \sum_{i\neq j} \left[d_{ij}\underline{s}\frac{f_U(\mu_{ij})}{F_U(\mu_{ij})} - (1 - d_{ij})\,\bar{s}\frac{f_U(\mu_{ij})}{1 - F_U(\mu_{ij})}\right]. \tag{57}$$

Case 3: Exactly one bucket in b is an inner bucket

If all but one component of $\mathbf{u}$ falls into its first or last bucket, then the resulting network is uniquely defined except for the presence or absence of one arc, say, ij. For any such draw of $\mathbf{u}$, since all other links are formed according to a strictly dominating strategy, player i will either benefit from forming the ij arc or not. Hence $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ is also either zero or one in this case.

For a particular network $\mathbf{d}$, $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ will equal one if two conditions hold. First, for all directed dyads $kl \neq ij$ such that $d_{kl} = 1$ we have that u_{kl} falls in the first bucket and for all dyads $kl \neq ij$ such that $d_{kl} = 0$ we have that u_{kl} falls in the last bucket. Second, for the dyad ij with u_{ij} falling in an inner bucket, we require that if $u_{ij} \in [\mu_{ij} + \gamma\underline{s}, \mu_{kl} + \gamma s_{ij}(\mathbf{d}))$ that $d_{ij} = 1$, while if $u_{ij} = [\mu_{kl} + \gamma s_{ij}(\mathbf{d}), \mu_{ij} + \gamma\bar{s})$ we require that $d_{ij} = 0$. The overall likelihood contribution for this case therefore equals:

$$\begin{aligned}\int_{\mathbf{u}\in\mathbf{b}} \mathcal{N}\left(\mathbf{d},\mathbf{u};\theta\right) f_{\mathbf{U}}\left(\mathbf{u}\right)\mathrm{d}\mathbf{u} = & \prod_{kl\neq ij}\left[\int_{-\infty}^{\mu_{kl}+\gamma\underline{s}} f_U\left(u_{kl}\right)\mathrm{d}u_{kl}\right]^{d_{kl}}\left[\int_{\mu_{kl}+\gamma\bar{s}}^{\infty} f_U\left(u_{kl}\right)\mathrm{d}u_{kl}\right]^{1-d_{kl}} \\ & \times\left[\int_{\mu_{ij}+\gamma\underline{s}}^{\mu_{ij}+\gamma s_{ij}(\mathbf{d})} f_U\left(u_{ij}\right)\mathrm{d}u_{ij}\right]^{d_{ij}}\left[\int_{\mu_{ij}+\gamma s_{ij}(\mathbf{d})}^{\mu_{ij}+\gamma\bar{s}} f_U\left(u_{ij}\right)\mathrm{d}u_{ij}\right]^{1-d_{ij}} \\ = & \prod_{kl\neq ij}\left[F_U\left(\mu_{kl}+\gamma\underline{s}\right)\right]^{d_{kl}}\left[1-F_U\left(\mu_{kl}+\gamma\bar{s}\right)\right]^{1-d_{kl}} \\ & \times\left[F_U\left(\mu_{ij}+\gamma s_{ij}\left(\mathbf{d}\right)\right)-F_U\left(\mu_{ij}+\gamma\underline{s}\right)\right]^{d_{ij}} \\ & \times\left[F_U\left(\mu_{ij}+\gamma\bar{s}\right)-F_U\left(\mu_{ij}+\gamma s_{ij}\left(\mathbf{d}\right)\right)\right]^{1-d_{ij}}.\end{aligned}$$

Recall the restriction $s_{ij}\left(\mathbf{d}\right) = s_{ij}\left(\mathbf{d}-ij\right) = s_{ij}\left(\mathbf{d}+ij\right)$. Because the last two terms in $[\cdot]$ in the expression above are zero at $\gamma = 0$ we only need to consider their derivative (by the product rule the other term equals zero at $\gamma = 0$). Differentiating the last two terms with respect to γ (and multiplying by the balance of preceding terms) yields

$$\begin{aligned} & \prod_{kl\neq ij}\left[F_U\left(\mu_{kl}+\gamma\underline{s}\right)\right]^{d_{kl}}\left[1-F_U\left(\mu_{kl}+\gamma\bar{s}\right)\right]^{1-d_{kl}} \\ & \times\left[s_{ij}\left(\mathbf{d}\right)f_U\left(\mu_{ij}+\gamma s_{ij}\left(\mathbf{d}\right)\right)-\underline{s}f_U\left(\mu_{ij}+\gamma\underline{s}\right)\right]^{d_{ij}} \\ & \times\left[\bar{s}f_U\left(\mu_{ij}+\gamma\bar{s}\right)-s_{ij}\left(\mathbf{d}\right)f_U\left(\mu_{ij}+\gamma s_{ij}\left(\mathbf{d}\right)\right)\right]^{1-d_{ij}} \\ = & \prod_{i\neq j}\left[F_U\left(\mu_{ij}+\gamma\underline{s}\right)\right]^{d_{ij}}\left[1-F_U\left(\mu_{ij}+\gamma\bar{s}\right)\right]^{1-d_{ij}} \\ & \times\left[s_{ij}\left(\mathbf{d}\right)\frac{f_U\left(\mu_{ij}+\gamma s_{ij}\left(\mathbf{d}\right)\right)}{F_U\left(\mu_{ij}+\gamma\underline{s}\right)}-\underline{s}\frac{f_U\left(\mu_{ij}+\gamma\underline{s}\right)}{F_U\left(\mu_{ij}+\gamma\underline{s}\right)}\right]^{d_{ij}} \\ & \times\left[\bar{s}\frac{f_U\left(\mu_{ij}+\gamma\bar{s}\right)}{F_U\left(\mu_{ij}+\gamma\bar{s}\right)}-s_{ij}\left(\mathbf{d}\right)\frac{f_U\left(\mu_{ij}+\gamma s_{ij}\left(\mathbf{d}\right)\right)}{F_U\left(\mu_{ij}+\gamma\bar{s}\right)}\right]^{1-d_{ij}}.\end{aligned}$$

Summing this expression over all potential arcs (and evaluating at $\gamma = 0$) gives a total contribution of "one inner bucket in $\mathbf{b}$" summands to the derivative of:

$$\begin{aligned} P_0\left(\mathbf{d};\delta\right)\sum_{i\neq j} d_{ij}&\left[s_{ij}\left(\mathbf{d}\right)\frac{f_U\left(\mu_{ij}\right)}{F_U\left(\mu_{ij}\right)}-\underline{s}\frac{f_U\left(\mu_{ij}\right)}{F_U\left(\mu_{ij}\right)}\right] \\ & +\left(1-d_{ij}\right)\left[\bar{s}\frac{f_U\left(\mu_{ij}\right)}{F_U\left(\mu_{ij}\right)}-s_{ij}\left(\mathbf{d}\right)\frac{f_U\left(\mu_{ij}\right)}{F_U\left(\mu_{ij}\right)}\right]. \end{aligned} \tag{58}$$

Summing (55) and (58) then gives the expression in the statement of Theorem 2.1. Using similar methods we can show that $P\left(\mathbf{d};\theta\right)$ can be differentiated with respect to γ twice as claimed.

A.3 MCMC Proofs

Proof of Lemma 3.1

Let $A_{G,G'}$ be the set of arcs form the node v_G to the node $v_{G'}$. We construct a bijection $\varphi : A_{G,G'} \to A_{G',G}$. Then we show that the probability of an arc $p(a)$ is equal to $p(\varphi(a))$. If that is proven, the probability of a transition form v_G to $v_{G'}$ is

$$\begin{aligned}\sum_{a \in A_{G',G}} p(a) &= \sum_{a \in A_{G',G}} p(\varphi(a)) \\ &= \sum_{\varphi^{-1}(a') \in A_{G',G}} p(a') \\ &= \sum_{a' \in \varphi(A_{G',G})} p(a') \\ &= \sum_{a' \in A_{G,G'}} p(a')\end{aligned}$$

which is the probability for a transition from $v_{G'}$ to v_G.
For the construction of the bijection consider that every arc $A_{G,G'}$ corresponds uniquely to a schlaufen-sequence $\mathcal{R} = (R_1, .., R_h)$. Let $R_k = (i_1, .., i_m, .., i_l)$ with i_m the start of the cycle (if there is no cycle in R, we set $R = \bar{R}$). We define $\bar{R}_k = (i_1, .., i_m, i_{l-1}, ..i_{m+1}, i_l)$ and $\bar{\mathcal{R}} = (\bar{R}_1, .., \bar{R}_h)$.

Note that the $R_1, .., R_h$ are link disjoint and as soon as the cycle of R_k is switched $\bar{R}_k$ is a schlaufe. The violation matrix of $\bar{R}_k$ is the negative violation matrix of R_k. This implies that if $\mathcal{R}$ is a feasible schlaufen sequence for G which defined a arc in $A_{G,G'}$ then $\bar{\mathcal{R}}$ is a feasible schlaufen-sequence for G' and defines an arc $A_{G',G}$.

We define now φ as the function which maps the arc in $A_{G,G'}$ with schlaufen sequence $\mathcal{R}$ to the arc in $A_{G',G}$ with schlaufen sequence $\bar{\mathcal{R}}$. By construction φ is injective, which implies $|A_{G,G'}| \leq |A_{G',G}|$. By symmetry we conclude $|A_{G',G}| \geq |A_{G,G'}|$, which implies $|A_{G',G}| = |A_{G',G}|$ and that φ is bijective.

It remains to show that the probability of an arc $p(a)$ is equal to $p(\varphi(a))$. For any node there are equally many feasible active / passive outlinks in G as in G'. If for a node one outlink is marked due to an link in R_k then for the same node one outlink is marked in $\bar{R}_k$. Therefore is $r_{G'_k}(i)$ equal to $r_{G_k}(i)$ for an active as well as a passive step. Looking at equation (44) the $p_{G_k}(R_k)$ is only different to $p_{G'_k}(\bar{R}_k)$ in the numbering of the factors. But in a cycle of a schlaufe the start node i_m and the end node i_l are such that $m - l \mod 2 = 0$. The reordering leaves even indexes even and odd indexes odd. Therefore $p_G(R_k) = p'_G(\bar{R}_k)$. From equation (45) it follows directly that $p_G(\mathcal{R}) = p_{G'}(\bar{\mathcal{R}})$ which completes the proof.

Proof of Lemma 3.2

The symmetric difference of two realizations of $\mathbb{G}_{s,m}$, which we denote by G and G' is a set of alternating cycles. Cycles are in particular schlaufen. We order them arbitrary $(R_1, .., R_h)$. The sum of the violation matrices is 0. Therefore $(R_1, .., R_h)$ is either a feasible schlaufen-sequence or a concatenation of feasible schlaufen-sequences. In the first case there is an arc from v_G to $v_{G'}$. In the second case, all the feasible schlaufen-sequence define an arc to a new node, resulting in a directed path starting at v_G and ending in $v_{G'}$. Thus between any two vertexes in Φ there is a directed path.

Proof of Theorem 3.3

Every time the Algorithm 1 arrives at step 2 a new arc is crossed. At step 2 the algorithm follows a loop arc of type 1 with probability q (here "refers" to the numbered points used to describe the state graph in the main text). Others it proceeds to step 3. In step 3 a schlaufen-sequence $\mathcal{R}$ is constructed. If the violation matrices of this schlaufen-sequence sum up to 0, the in it cycles are switched and an arc of type 2 is followed with probability $p_G(\mathcal{R})$. If the violation matrices do not sum up to 0, then an arc of type 3 is followed. All the cases, in which the violation matrices don't sum up to 0, correspond to the residual probability. Therefore Algorithm 1 is a a random walk on the state graph Φ.

According to lemma 3.1 Φ is (weighted) symmetric and according to lemma 3.2 it is strongly connected. Due to the self-loops, Φ is not bipartite. Therefore the limit distribution is uniform.