Bayesian Inference of Network Formation Models with Payoff Externalities

Cheng Ding* Juan Estrada† Santiago Montoya-Blandón‡

March 1, 2023

Abstract

This paper provides a novel approach to identification and estimation of a network formation model using observed network data, where the model acknowledges possible externalities in agents' utilities of forming connections with one another. The existence of externalities induces an issue of multiple equilibria. We first show that local point identification of the parameters of interest is possible under a mild assumption on the equilibrium selection process. We then propose a Bayesian estimation method to conduct statistical inference of structural payoff coefficients. Implementing the resulting MCMC algorithm requires sampling from the generalized inverse normal distribution, for which we found no efficient sampling algorithm in the literature. A by-product of this paper is to provide such efficient sampling algorithm for the regular and truncated variants of this distribution. Our method also allows us to estimate equilibrium selection probabilities, which requires knowledge on possible equilibrium configurations. We address this issue by proposing a composite likelihood function based on subgraphs of the observed network. We show that the use of a composite likelihood induces misspecification, characterize the Kullback-Leibler divergence that measures this misspecification error and show this measure can be used to tune the composite likelihood weights. We present an empirical application to model the network formation process of individuals creating social connections in villages in Karnataka, India and find strong evidence of homophily effects.

Keywords: Network formation; Multiple Equilibria; Subgraphs; Composite Likelihood; Bayesian Inference.

JEL Codes: C11, C57, D14, D85.

*Department of Economics, Emory University, Rich Building 306, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA. Email: cding25@emory.edu

†Quantitative Theory & Methods Department, Emory University, Rich Building, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA. Email: jjestra@emory.edu

‡University of Glasgow, Adam Smith Business School. University Ave, Glasgow, United Kingdom, G12 8QQ. Email: santiago.montoya-blandon@glasgow.ac.uk

1 Introduction

Social networks are critical to determining how individuals make choices in contexts ranging from labor markets and educational achievement to substance abuse and criminality (Gaviria and Raphael, 2001; Sacerdote, 2001; Bayer et al., 2009; Mas and Moretti, 2009). As network structures arise from individual strategic interaction, it becomes important to understand the underlying network formation process when studying the effects of these structures using observational data (see, e.g., Goldsmith-Pinkham and Imbens, 2013; Johnsson and Moon, 2021). The main contribution of this article is to propose a method to point identify and estimate strategic network formation models using observational data on network links. Our paper contributes to this literature by providing, to the best of our knowledge, the first results on point identification of both preference and selection mechanism parameters.

We then propose a Bayesian method to estimate the structural parameters of preferences and equilibrium selection probabilities that characterize the network formation model. We interpret observed networks as the equilibrium outcome of a complete information game where individuals make connections based on a flexible payoff function that allows for externalities in the utility received from links elsewhere in the network (Pelican and Graham, 2020). In addition, we offer a method to conduct statistical inference on the utility function's parameters that do not rely on asymptotic approximations. This contribution is particularly relevant because our estimation method allows for inference of not only the homophily parameter but also the degree heterogeneity and the payoff externality parameters. The problem of providing asymptotic theory for the degree heterogeneity parameters has proved to be complex because the vector of parameters grows with the sample size (Neyman and Scott, 1948). Moreover, inference for the payoff externality parameters is not standard, given the possibility of multiple equilibria (Pelican and Graham, 2020).

Allowing for strategic externalities is appealing since it provides a rich model that matches documented features of social networks found in the data, such as clustering and homophily (Jackson and Rogers, 2007; Jackson et al., 2012; Sheng, 2020). However, strategic network formation models with utility externalities are plagued with identification and estimation issues. These outstanding problems include multiple equilibria, statistical dependence in large networks, and the curse of dimensionality (Leung, 2015; De Paula et al., 2018). Differing from previous research proposing methods that resolve some problems but not all, this paper provides a technique that simultaneously addresses all these fundamental issues.

We deal with the problem of multiple equilibria by specifying both a payoff function and an equilibrium selection mechanism as the primitives of the network formation model. The selection mechanism determines the probability distribution over possible equilibria for a given realization of exogenous characteristics, individual heterogeneity, and idiosyncratic dyadic shocks. As in Bajari et al. (2010), we take an empirical approach to

characterize the equilibrium selection. Critically, we assume the existence of a predetermined and potentially unknown probability distribution over all possible equilibrium outcomes. We interpret the probability distribution as nature assigning different likelihoods of occurrence to different network structures. We show that this assumption is enough to guarantee local point identification of the structural parameters underlying both utility functions and equilibrium selection probabilities. Our approach differs from previous studies of discrete games of complete information with multiple equilibria such as Bajari et al. (2010) and Bajari et al. (2011) due to the added complexity in the strategy space for the game we consider. This complexity makes it intractable to compute all possible equilibria of the game. Therefore, instead of defining a selection mechanism in the space of individuals' strategies, this paper focuses on selecting network structures that are the game's equilibrium outcomes. One advantage of directly selecting outcomes is that selection of equilibrium networks can result from both pure and mixed strategies (Pelican and Graham, 2020).

To study identification, we introduce assumptions on equilibrium selection based on the idea of *multiplicity regions*. We define a multiplicity region as a subset of the support of idiosyncratic dyadic shocks where the group of all equilibrium networks is the same for any value of the shocks belonging to that region. We show that the total likelihood of observing an equilibrium network can be written as a weighted sum of integrated likelihoods across all possible multiplicity regions. Though critical for showing that the model's parameters are point identified, the formalization in terms of multiplicity regions does not solve the curse of dimensionality problem inherent to the estimation process. This is due to the dimensionality of the idiosyncratic shocks, which result in a likelihood that is computationally intractable, even when using simulation-based methods (McFadden, 1989; Hajivassiliou and McFadden, 1998).

To overcome this issue, we resort to a Bayesian estimation method. Following a data augmentation approach, we introduce the idiosyncratic shocks as latent variables to be sampled along with the necessary preference parameters. This idea forms the basis of many computationally efficient algorithms for estimating models where numerical integration is unfeasible, such as discrete choice models with large choice sets (Albert and Chib, 1993; McCulloch et al., 2000). The backbone of our approach is the fact that, conditional on parameter values, a transformation of the idiosyncratic shocks fully determines admissible equilibrium configurations. We show that once we condition on the transformed values of shocks, only the structure of the externality function remains relevant in determining the distribution of equilibrium networks. Our algorithm alternates between sampling from the utility parameters conditional on transformed shocks and the distribution of shocks given parameters and network data.

We highlight three relevant advantages of using a Bayesian estimation method instead of the standard frequentest approach common in the literature. First, the Bayesian approach allows us to sidestep the issue of high-dimensional numerical integration in order to evaluate the likelihood. Second, we can address the potential

issue of statistical dependence in large networks. The key idea is that, as our estimation method conditions on the current network and does not rely on asymptotic approximations to conduct inference, the distinction between large-games and many-games asymptotics becomes less relevant. Notably, the flexibility in our estimation method can accommodate both large-networks and many-networks data structures, while at the same time generating posterior distributions for the equilibrium selection probabilities, which allows the researcher to simulate the model and perform counterfactual analysis. Finally, the use of random coefficients to perform estimation on the two-way individual heterogeneity components allows to conduct statistical inference while avoiding the issue of asymptotic bias caused by the presence of incidental parameters (Jochmans, 2018; Dzemski, 2019; Yan et al., 2019). The advantage of using the random effects approach can also be extended to longitudinal panel data models as in Fernández-Val and Weidner (2016), where the incidental parameter issue also causes non-standard asymptotic biases.

Implementing our approach requires sampling from the generalized inverse normal ($\mathcal{GIN}$) distribution proposed by Robert (1991). This distribution is conjugate to our likelihood function; assigning a generalized inverse normal prior to (the reciprocal of) the externality coefficient results in a posterior that is also generalized inverse normal. The $\mathcal{GIN}$ distribution can also arise in a simultaneous equation framework with a network structure (Masten, 2014). To the best of our knowledge, there is no efficient algorithm for sampling from the generalized inverse normal distribution in the literature. An additional contribution of our paper is providing such an efficient algorithm. To this end, we build on the tools that constitute the foundation of efficient sampling algorithms for the generalized inverse Gaussian distribution given the similar shape of their kernels.[1] We provide the formal details guaranteeing the algorithm's efficiency for sampling the generalized inverse normal distribution and its truncated variants to the left or to the right of 0 using these tools (Kinderman and Monahan, 1977; Leydold, 2001; Hörmann and Leydold, 2014).

Although our Bayesian estimation approach addresses the computational tractability caused by the dimensionality of the idiosyncratic shocks, the multiplicity of equilibrium outcomes creates two additional issues. The first is in sampling shocks, as the normalizing constant in their posterior distribution remains difficult to calculate. This is because computing the normalizing constant is equivalent to obtaining the set of equilibria for a given configuration of shocks, which is computationally infeasible. The second issue arises when sampling equilibrium selection probabilities, as the space of network configurations is high-dimensional, and finding two isomorphic networks becomes impractical. To tackle both issues, on top of taking a Bayesian estimation approach, we propose a composite likelihood objective function formed by multiplying the marginal distributions of N_K subgraphs forming the complete network $\mathbf{d}$ (Graham, 2020). We show that the correct specification of the

[1] Both distributions arise as different ways to conceptualize the distribution of the reciprocal of a normally distributed random variable.

subgraph marginal distributions in this context is not guaranteed. The misspecification issues cause the classic consistency results of estimators based on the composite likelihood in Cox and Reid (2004) not to apply here. We characterize the Kullback-Leibler divergence that measures the misspecification error and show that it is possible to control it by choosing the weights for each subnetwork such that those with lower external connectivity have a larger weight than those with higher connectivity. Furthermore, using subgraphs allows us to completely characterize the set of equilibria for any given configuration of shocks, making the problem of evaluating the proportionality constant in the posterior distribution of shocks tractable.

To highlight the usability of our approach, we present an empirical application using data from the Social Networks and Microfinance project, which contains the publicly available data of participation in a program of Bharatha Swamukti Samsthe (BSS), a microfinance institution (MFI) in rural southern Karnataka (Banerjee et al., 2013). Crucially, the data include information of thirteen possible relationships among individuals, including visiting each other, praying, borrowing and lending money and goods, obtaining advice, and giving advice. We combine the different social ties into one unique social network encompassing the thirteen dimensions. Based on the constructed social network, we model the network formation process and estimate the payoff parameters of interest. We find strong evidence of homophily for most of the characteristics in our analysis. The most substantial homophily effects happen among the same gender and working status of villagers.

The structure of the paper is as follows. Section 2 introduces the network formation model, our population assumptions, and the main identification result. Section 3 presents the Bayesian algorithm, and introduces the idea of the composite likelihood function. Section 4 presents the data and empirical results. Finally, section 5 concludes.

2 Network Formation Model and Identification

2.1 Network Formation

Following Pelican and Graham (2020) and Sheng (2020), we assume that any observed network structures are the result of a network formation model where a set of N individuals, represented by $\mathcal{I}_N = \{1, 2, \ldots, N\}$, choose connections simultaneously in what is known as a link announcement game. Each individual i is characterized by a K-dimensional vector of observed attributes X_i and unobserved (to the researcher) degree effects A_i. Further, each dyad composed from players i and j (with $i \neq \text{j}$) faces a pair of shocks $(\tilde{U}_{ij}, \tilde{U}_{ji})$ that affect their connection decisions. The collection of all these objects is given by $\mathbf{A} = (A_1, \ldots, A_N)$, $\mathbf{X} = (X_1, \ldots, X_N)$, and $\tilde{\mathbf{U}} = (\tilde{U}_{12}, \ldots, \tilde{U}_{1N}, \ldots, \tilde{U}_{N,N-1})$. Let $\mathbf{d} \in \mathbb{D}_N$ be a realization of a network configuration from the set $\mathbb{D}_N$ of all possible N-player adjacency matrices. We assume payoff functions that allow for players to have

preferences over other individuals' positions in the network. In particular, we assume that individuals have additively separable utility functions divided into three components: degree heterogeneity, assortative matching (also known as homophily) and an externality component that can generate multiple equilibria. In defining our payoff function, for notational convenience, we do not explicitly include A_i and X_i as arguments. Individual's utility from the network configuration is determined by

$$\nu_i\left(\mathbf{d}, \tilde{U}_{ij}\right) = \sum_j d_{ij}\left[A_i + A_j + \lambda_0' W_{ij} + \gamma_0 s_{ij}(\mathbf{d}) - \tilde{U}_{ij}\right], \quad i = 1, \ldots, N \tag{1}$$

where $W_{ij,l} = |X_{i,l} - X_{j,l}|$ for $l = 1, \ldots, p$. The full network $\mathbf{d}$ appears on the right hand side of (1) to reflect the fact that individual i's utility can be potentially affected by other individuals' links in the network. The utility function depends on two individual degree-heterogeneity components A_i and A_j, a dyadic unobserved component $\tilde{U}_{ij}$, a homophily component $\lambda_0' W_{ij}$, and preferences over links other than d_{ij} given by the externality function $s_{ij}(\mathbf{d}) : \mathbb{D}_N \to \mathcal{S}$. This externality function can include relevant cases in the literature such as taste for reciprocated links in directed networks (Pelican and Graham, 2020), and taste for indirect connections and completing triangles in undirected networks (Mele, 2017; Christakis et al., 2020; Sheng, 2020). Similar to Pelican and Graham (2020), a key element of this model is the fact that $s_{ij}(\mathbf{d})$ is discrete-valued, such that $\mathcal{S} = \{s_0, \ldots, s_L\}$ and $|\mathcal{S}| = L + 1$.

As pointed out by Pelican and Graham (2020), choosing an equilibrium selection concept is closely related with the assumption on the directed or undirected nature of individuals' links. Games on directed networks are associated with the Nash equilibrium solution concept, while in the analysis of undirected networks, pairwise stability as equilibrium concept plays a fundamental role (Jackson and Wolinsky, 1996). To accommodate the structure of our empirical application, we assume that individuals form undirected connections. In Appendix B, we show that this assumption is not necessary for our identification and estimation results, as they also apply to a game where individuals can form directed links under a Nash equilibrium solution concept. Thus, following the literature on undirected network formation games, we use pairwise stability under transferable utility as our equilibrium concept.[2] Individuals i and j decide whether or not to form a connection based on their marginal utilities. From the payoff function (1), it follows that the marginal utility individual i receives from linking to j is

$$\Delta\nu_{ij}\left(\mathbf{d}, \tilde{U}_{ij}\right) = A_i + A_j + \lambda_0' W_{ij} + \gamma_0 s_{ij}(\mathbf{d}) - \tilde{U}_{ij}, \quad i = 1, \ldots, N; j = 1, \ldots, N; i \neq j, \tag{2}$$

Finally, define $\Delta\bar{\nu}_{ij}(\mathbf{d}, \tilde{U}_{ij}, \tilde{U}_{ji}) = [\Delta\nu_{ij}\left(\mathbf{d}, \tilde{U}_{ij}\right) + \Delta\nu_{ji}\left(\mathbf{d}, \tilde{U}_{ji}\right)]/2$. Using the definition of marginal utility in equation (2), we can now introduce the definition of pairwise stability. We then discuss the existence of a pairwise

[2] Our initial approach considered non-transferable utility as transferable utility can seem restrictive in this setup. The implications of non-transferable utility can be incorporated in our Bayesian estimation scheme increase at the cost of computational complexity without adding much flexibility nor changing results considerably. We maintain the transferable utility assumption for simplicity.

stable network under the utility function that we introduce in equation (1).

Definition 1 (Pairwise Stability)**.** A network $\mathbf{d}$ is pairwise stable under transferable utility if for any two individuals i and j, we have

(i) $d_{ij} = d_{ji} = 1$ when $\Delta\bar{\nu}_{ij}(\mathbf{d}, \tilde{U}_{ij}, \tilde{U}_{ji}) \geq 0$;

(ii) $d_{ij} = d_{ji} = 0$ when $\Delta\bar{\nu}_{ij}(\mathbf{d}, \tilde{U}_{ij}, \tilde{U}_{ji}) < 0$.

Equilibrium. We assume a game of complete information. Each individual i observes $\{A_i, X_i'\}_{i=1}^N$ and $\{\tilde{U}_{ij}\}_{i\neq j}$, then decides a set of links from the $N-1$ agents. Then, a link is formed if both individuals perceive a positive marginal utility from the connection. Sheng (2020) shows that under transferable utility, assuming that the links are strategic complements, the model can be casted into a supermodular game in which the the existence of an equilibrium follows from the fixed-point theorem for isotone mappings (Topkis, 1979; Milgrom and Roberts, 1990). Therefore, assuming that $\gamma_0 \geq 0$ in our payoff function guarantees a non-empty set of pairwise stable networks from $\mathbb{D}_N$.

Note that under our definitions, link decisions are given as any solution to the simultaneous system of non-linear equations given by

$$d_{ij} = d_{ji} = \mathbb{I}\left[A_i + A_j + \lambda_0' W_{ij} + \gamma_0 \frac{s_{ij}(\mathbf{d}) + s_{ji}(\mathbf{d})}{2} - \frac{\tilde{U}_{ij} + \tilde{U}_{ji}}{2} \geq 0\right], \quad i = 1, \ldots, N; j = 1, \ldots, i-1$$

The externality functions considered in the literature are also symmetric as they satisfy $s_{ij}(\mathbf{d}) = s_{ji}(\mathbf{d})$ for all dyads ij and a given symmetric network $\mathbf{d}$. Defining $U_{ij} \equiv (\tilde{U}_{ij} + \tilde{U}_{ji})/2$, we see that the system reduces to

$$d_{ij} = d_{ji} = \mathbb{I}\left[A_i + A_j + \lambda_0' W_{ij} + \gamma_0 s_{ij}(\mathbf{d}) \geq U_{ij}\right], \quad i = 1, \ldots, N; j = 1, \ldots, i-1 \tag{3}$$

Equation (3) shows that the uncertainty in linking decision comes solely through the average of dyad-level shocks U_{ij}, given the transferibilty assumption (a similar solution appears in Hsieh et al., 2022).

2.2 Population Assumptions

We assume that the population is formed by an arbitrarily large set of individuals $\mathcal{I}_N$. As argued by Goldsmith-Pinkham and Imbens (2013), the ability to identify the statistical properties of the network rests on the assumptions on the individuals' dependence in the population given the network structure. Two standard assumptions have been used in the econometrics literature of network formation games. The more straightforward but less realistic case assumes a large number of exchangeable networks. This assumption is consistent with a large population that can be divided into independent networks that are mutually disjoint. A more realistic alternative is to assume a large network defined on the set of nodes $\mathcal{I}_N$. As argued by Leung (2015), the issue with this

approach is that the researcher has to impose conditions on the network dependence between individuals to be able to use asymptotic approximations to conduct inference.

One of our Bayesian estimation method's key advantages is that we do not require an asymptotic approximation to conduct inference. Therefore, our approach can accommodate both of the mentioned population assumptions. However, we need to observe repeated networks to identify the parameters of the equilibrium selection mechanism. The identification argument applies directly to the repeated networks population assumption. In the large networks population assumption case, we have to impose the additional restriction that the analyst can partition the large network into approximately independent sub-networks, see, e.g., Schweinberger and Handcock (2015) and Sheng (2020). It is then possible to estimate selection probabilities when multiple equilibria are present with repeated sub-network sequences.

2.3 Identification

The network externality component $s_{ij}(\mathbf{d})$ in our network formation game induces the potential for multiple equilibria. To specify a likelihood function for this problem, we need to incorporate a way to assign probabilities to different equilibrium outcomes in the set of all possible equilibria. To that end, we define $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) : \mathbb{D}_N \times \mathbb{R}^{N(N-1)/2} \to [0, 1]$ to be a function that assigns probabilities to the set of equilibrium networks, where $\theta = [\lambda', \gamma]'$ is the vector of payoff parameters. We call this function $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ the equilibrium selection function.[3] The set of networks in equilibrium can include both outcomes that result from individuals playing mixed strategies or playing pure strategies. With this definition, the likelihood of observing a network $\mathbf{d}$ is given by

$$P(\mathbf{d}; \theta) = \int_{\mathbf{u} \in \mathbb{R}^{N(N-1)}} \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) \prod_{i=1}^{N} \prod_{j \neq i} f_U(u_{ij}) \, \mathrm{d}\mathbf{u} \ . \tag{4}$$

For identification, we do not need to specify any particular probability distribution for f_U. A common practice in the econometrics of games for incomplete models is to partition the errors' space into different subspaces that determine the potential number of equilibria conditional on given values of the regressors and the parameters of the model (De Paula, 2013). Characterizing these regions in our context is relevant because it allows us to express the likelihood of observing any given outcome as a sum of the probability mass in the areas where the outcome can happen weighted by the relative probability of the outcome in those regions. We call the areas partitioning the errors' space the *multiplicity regions*. In this paper, the definition of multiplicity regions requires conditioning on the values of regressors $\mathbf{X}$, degree heterogeneity $\mathbf{A}$ and parameters θ.

Definition 2 (Multiplicity Region)**.** Conditional on values of $\mathbf{X}$, $\mathbf{A}$, and θ, a multiplicity region $m \in M$ in the

[3]For convenience, we do not make explicit that the equilibrium selection distribution also depends on covariates $\mathbf{X}$ and degree heterogeneity effects $\mathbf{A}$.

space of $\mathbf{U}$ is a region where the set of all possible equilibrium networks is the same for all $\mathbf{u} \in m$. That is, $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) = \mathcal{N}(\mathbf{d}, \mathbf{u}'; \theta)$ for any $d \in \mathbb{D}_N$ and $\mathbf{u}, \mathbf{u}' \in m$. Furthermore, multiplicity regions partition the space of $\mathbf{U}$; i.e., for $m, s \in M$, $m \cap s = \varnothing$ (regions are disjoint) and $\bigcup_{m \in M} m = \mathbb{R}^{N(N-1)}$.

A simple example can clarify the role of multiplicity regions in determining the observed network $\mathbf{d}$.

Example 1 (Multiplicity Regions)**.** *Consider the simple case when $A_i = 0$ for all i and $\Lambda_0 = 0$. Moreover, assume that externalities are given by intransitive triads, $s_{ij}(\mathbf{d}) = d_{ik}d_{kj}$ for all i, for $\gamma \geq 0$. Finally, assume $N = 3$. In this scenario there are only eight possible network configurations given by*

$$d_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, d_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, d_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, d_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, d_5 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$d_6 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, d_7 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \text{ and } d_8 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Determining what network configuration would emerge and whether or not it is unique, depends on the realizations of the errors and the values of the parameters. In this example the space of the vector of shocks $\mathbf{u} = (u_{12}, u_{21}, u_{13}, u_{31}, u_{23}, u_{32})$ can be divided into the following partition that completely determine the admissible equilibrium networks given the externality parameter γ:

$$\underbrace{(-\infty, 0]}_{I_1} \cup \underbrace{(0, \gamma]}_{I_2} \cup \underbrace{(\gamma, \infty)}_{I_3}. \tag{5}$$

Table 1 presents all the possible equilibrium networks given that $u_{13}, u_{31} \in I_1$ and $u_{23}, u_{32} \in I_2$. Each row represents a realization of the errors such that each dyad shock falls into one of the three possible multiplicity regions. Intuitively, if the shocks are too small (both u_{12} and u_{21} fall into I_1), then it is always profitable for individuals to form connections, and the only equilibrium network is d_4. The same intuition holds for large values of the shocks where the only possible equilibrium configuration is the empty network. When both shocks fall into intermediate values (I_2), multiple equilibria arises, both the empty and the complete network are possible outcomes.

The primary identification idea is to separate the likelihood of the problem into all possible multiplicity regions and evaluate whether it is possible to identify the parameters of interest for each of those regions. Identification in this context is in terms of observational equivalence; i.e., the distribution of the data at the

Table 1: Possible Equilibrium Networks in a Simple 2x2 Example

(u_{12}, u_{21})	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
I_1, I_1	0	0	0	0	0	0	0	1
I_1, I_2	0	0	1	0	0	0	0	1
I_1, I_3	0	0	1	0	0	0	0	0
I_2, I_1	0	0	1	0	0	0	0	1
I_2, I_2	0	0	1	0	0	0	0	1
I_2, I_3	0	0	1	0	0	0	0	0
I_3, I_1	0	0	1	0	0	0	0	0
I_3, I_2	0	0	1	0	0	0	0	0
I_3, I_3	0	0	1	0	0	0	0	0

Note: table with all the possible multiplicity regions for an example with intransitive triads externalities. Each row represents a possible combination of realized shocks that follow in one of three possible buckets as described in equation (5). Each column represents the network structures that are possible equilibria. Position i, j in the table equals one if the network configuration j is a possible equilibrium in the multiplicity region i, and zero otherwise.

true parameter is different from that at any other possible parameter value. To motivate our assumption, we introduce the following proposition summarizing a similar discussion in section 2.2 of Pelican and Graham (2020). The proposition relies on the following quantity, which, conditional on parameter values, completely determines whether any given network is an equilibrium or not. Define t_{ij}, for all dyads ij as

$$t_{ij} \equiv \frac{u_{ij} - A_i - A_j - W'_{ij}\lambda}{\gamma} \tag{6}$$

Proposition 1 (Best response characterization)**.** *Recall that* $\mathcal{S} = \{s_0, \ldots, s_L\}$*. Then, for any* $\mathbf{d} \in \mathbb{D}_N$*, for all* $i = 1, \ldots, N$ *and* $j < i$*,*

(i) If $t_{ij} \leq s_0$*, then* $d_{ij} = 1$*,*

(ii) If $t_{ij} > s_L$*, then* $d_{ij} = 0$*,*

(iii) If $s_l < t_{ij} \leq s_{l+1}$*, then* $d_{ij} = 0$ *when* $s_{ij}(\mathbf{d}) \leq s_l$ *and* $d_{ij} = 1$ *when* $s_{ij}(\mathbf{d}) > s_l$*.*

Although the proposition relies on $\gamma > 0$, we can re-define t_{ij} to equal only the numerator of (6) in case $\gamma = 0$, to obtain a similar expression.[4] Let $\mathbf{t}$ collect all t_{ij} across dyads. We impose an assumption on the equilibrium selection in terms of multiplicity regions that is sufficient to point-identify utility function parameters and selection probabilities from observational data on links.

Assumption 1 (Equilibrium Selection Characterization)**.** (i) There exist a probability distribution over the set of all possible networks $\mathbb{D}_N$ denoted by h. (ii) For every $\mathbf{d} \in \mathbb{D}_N$ and $\mathbf{t} \in \mathbb{R}^{N(N-1)/2}$, the distribution only depends on $\mathbf{d}$, which we denote by $h(\mathbf{d})$.

[4] In this case, we would set $s_0 = s_L = 0$. We see that this implies there is only one equilibrium network, as expected. We analyze the case of $\gamma_0 = 0$ in Section 3.

The critical part of Assumption 1 is the fact that the weights are independent of the specific configurations $\mathbf{t}$. It is important to clarify that this does not imply that equilibrium networks are independent of shocks conditional on parameters. Instead, Assumption 1 states that this dependence occurs only through the set of admissible $\mathbb{D}(\mathbf{t})$. With the characterization given in Proposition 1 and following Assumption 1, we can write

$$\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta, h) = \frac{h(\mathbf{d})}{\sum_{\mathbf{w} \in \mathbb{D}_N} h(\mathbf{w}) g(\mathbf{w}, \mathbf{u}; \theta)} g(\mathbf{d}, \mathbf{u}; \theta),$$

where the function $g(\mathbf{d}, \mathbf{u}; \theta)$ takes the form

$$g(\mathbf{d}, \mathbf{u}; \theta) = \prod_{i=1}^{N} \prod_{j<i} \mathbb{I}\left[A_i + A_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}) \geq u_{ij}\right]^{d_{ij}} \cdot \mathbb{I}\left[A_i + A_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}) < u_{ij}\right]^{1-d_{ij}}. \quad (7)$$

As $\mathcal{N}$ is an equilibrium selection function, it should only assign weight to pairwise-stable equilibria. The function g is essentially an indicator that takes the value of 1 when $\mathbf{d}$ is an equilibrium network given the vector of shocks $\mathbf{u}$ preference parameters θ and 0 otherwise. By Proposition 1, it only depends on the transformed $\mathbf{u}$ defining the vector $\mathbf{t}$. We can re-write the denominator in the expression for $\mathcal{N}$ as

$$\sum_{\mathbf{w} \in \mathbb{D}_N} h(\mathbf{w}) g(\mathbf{w}, \mathbf{u}; \theta) = \sum_{\mathbf{w} \in \mathbb{D}(\mathbf{t})} h(\mathbf{w}),$$

where $\mathbb{D}(\mathbf{t})$ is the set of all equilibria that are compatible with $\mathbf{t}$. By definition, $\mathbb{D}(\mathbf{t}) = \mathbb{D}(\mathbf{t}')$ if $\mathbf{t}, \mathbf{t}' \in m$, so that we can focus only on $\mathbb{D}_m$, the set of all possible equilibrium networks within an arbitrary multiplicity region $m \in M$. Then, from the definition of conditional probability, the true equilibrium selection distribution (and thus the likelihood function conditional on $\mathbf{U} = \mathbf{u}$) is such that

$$P(\mathbf{d}|\mathbf{u}; \theta_0, h_0) = \mathcal{N}(\mathbf{d}, \mathbf{u}; \theta_0, h_0) = \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{d' \in \mathbb{D}_m} h_0(\mathbf{d}')} \mathbb{I}\{\mathbf{u} \in m\} g(\mathbf{d}, \mathbf{u}; \theta_0). \quad (8)$$

Therefore, Assumption 1 implies that the selection probability $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta, h)$ only depends on $\mathbf{X}$ and utility components via the multiplicity regions. As we can normalize $\sum_{\mathbf{d} \in \mathbb{D}_N} h(\mathbf{d}) = 1$, we interpret the resulting probability distribution as nature determining what type of equilibrium is more likely a priori. This formulation partially alleviates the potential misspecification issues raised in De Paula (2013). A potential misspecification problem in our context would arise if the probability distribution $h_0(\mathbf{d})$ depends on transformed shocks $\mathbf{t}$ or our model is incorrect for describing the network formation process. However, assuming the model itself is correctly specified, there will not be further dependence of $h_0(\mathbf{d})$ on $\mathbf{X}$ or any utility parameter or potential for misspecifying this relationship.

Combining equation (4) with the definition of the equilibrium selection distribution in Assumption 1, it follows that the likelihood of the problem can be written as

$$\begin{aligned} P(\mathbf{d};\theta,h) &= \int_{\mathbf{u}\in\mathbb{R}^N} \sum_{m\in M} \frac{h(\mathbf{d})}{\sum_{d'\in\mathbb{D}_m} h(\mathbf{d}')} \mathbb{I}\{\mathbf{u}\in m\} g(\mathbf{d},\mathbf{u};\theta) \prod_{i=1}^{N}\prod_{j\neq i} f_U\left(u_{ij}\right) \mathrm{d}\mathbf{u} \\ &= \sum_{m\in M} \frac{h(\mathbf{d})}{\sum_{d'\in\mathbb{D}_m} h(\mathbf{d}')} \int_{\mathbf{u}\in m} g(\mathbf{d},\mathbf{u};\theta) \prod_{i=1}^{N}\prod_{j\neq i} f_U\left(u_{ij}\right) \mathrm{d}\mathbf{u} \end{aligned} \tag{9}$$

For simplicity in notation, let $G(\mathbf{d},m;\theta) = \int_{\mathbf{u}\in m} g(\mathbf{d},\mathbf{u};\theta) \prod_{i=1}^{N}\prod_{j\neq i} f_U\left(u_{ij}\right) \mathrm{d}\mathbf{u}$. It follows that the likelihood can be represented as

$$P(\mathbf{d};\theta,h) = \sum_{m\in M} \frac{h(\mathbf{d})}{\sum_{d'\in\mathbb{D}_m} h(\mathbf{d}')} G(\mathbf{d},m;\theta). \tag{10}$$

Note that the likelihood in (10) varies as a function of θ from the selection probability and the shape of multiplicity regions. When testing for the existing of strategic interactions, Pelican and Graham (2020) define the concept of bucket, which is a mapping from the space of $\mathbf{u}$ to a collection of intervals that determine the connection behavior of individuals i and j. In particular, Pelican and Graham (2020) argue that if the realization of the shock u_{ij} falls into the outer buckets, then ij's connection decision is uniquely determined, while u_{ij} falling into the inner buckets opens the possibility for multiple equilibria. Following Pelican and Graham (2020), we define buckets as follows.

Definition 3 (Buckets)**.** An outer bucket is either the interval $(-\infty, \mu_{ij}+\gamma s_0]$ or $(\mu_{ij}+\gamma s_L, \infty)$, while the inner buckets are $(\mu_{ij}+\gamma s_l, \mu_{ij}+\gamma s_{l+1}]$ for all $l = 0,\ldots,L-1$, where $\mu_{ij} = A_i + A_j + X_i + W'_{ij}\lambda$.

For example, when s_{ij} takes the form of the reciprocity externalities, then $L=1$, $s_0=0$ and $s_L=1$, while if s_{ij} is the intransitive triads externality function, then $L=N-2$ $s_0=0$ and $s_L=N-2$. Example 1 presents a simple case where the definition of $s_{ij}(\mathbf{d})$ allows for a simple characterization of the multiplicity regions. In that example, the multiplicity regions coincide with the buckets as defined by Pelican and Graham (2020). However, for a more general externality measure such as intransitive triads, in which $s_{ij}(\mathbf{d}) = \sum_k d_{ik}d_{kj}$, the multiplicity regions will not coincide with the buckets. Even though the two measures do not coincide, there is a tight relationship between them. All multiplicity regions are composed of a countable number of buckets that also determine their shape. The following Proposition formalizes this discussion.

Proposition 2 (Multiplicity Regions Characterization)**.** *Let $b\in\mathbb{B}$ be an arbitrary bucket, where $\mathbb{B}$ is the set of buckets that partition the errors' space $\mathbb{R}^N$. Then, b must be within one and only one multiplicity region $m\in M$, where $b\subseteq m$.*

The proof of proposition 2 is relegated to Appendix A. Instead, we provide an example with the main intuition of the result. Example 2 shows the geometric intuition from the result in proposition 2 for a simple case of three individuals when the externalities are generated by the taste for completing triangles.

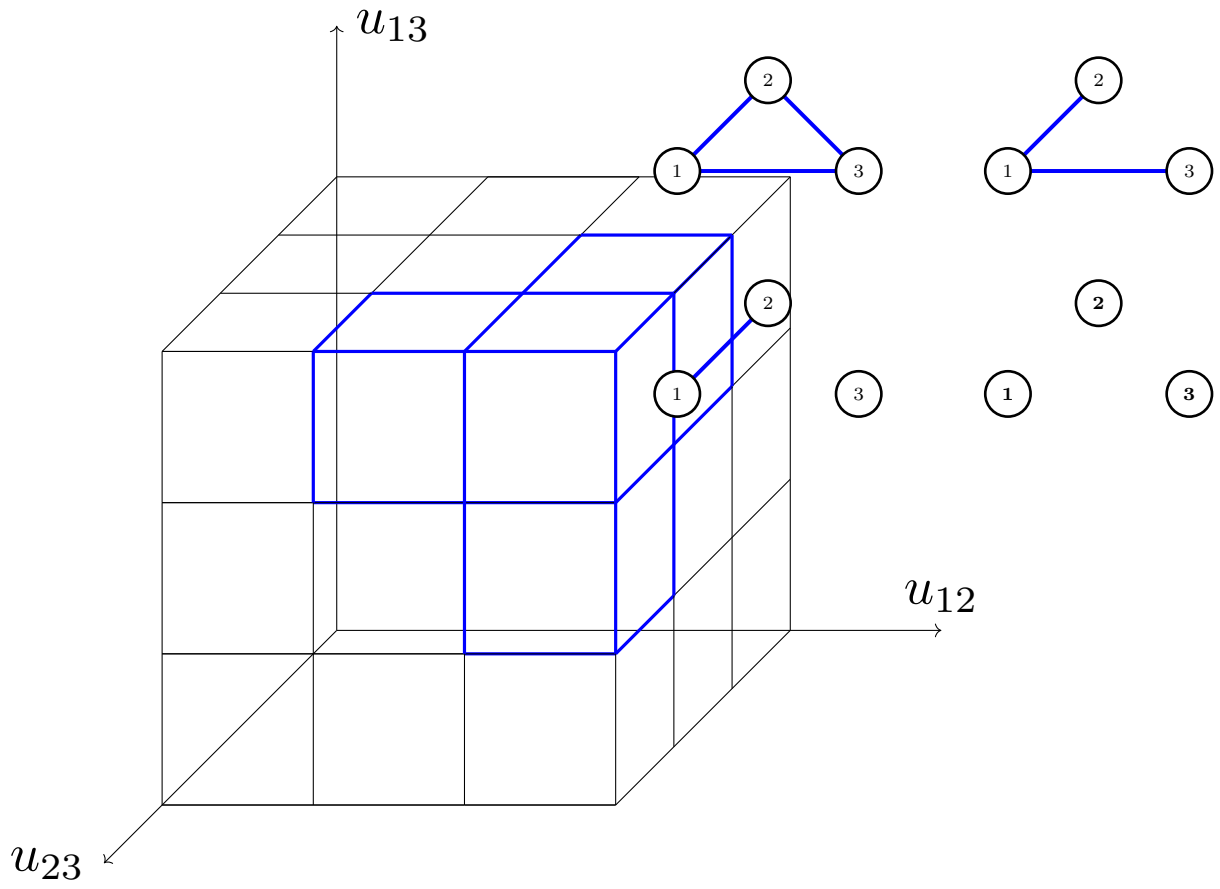
Example 2. *Assume $\mathcal{I}_3 = \{1, 2, 3\}$ with utility functions given by equation (1), where $s_{ij}(\mathbf{d}) = \sum_k d_{ik}d_{kj}$. As shown in equation 7, individuals are only going to form connections if their idiosyncratic shock is low enough. In particular, following definition 3, we can divide the space of each u_{ij} shock into three components given by two outer buckets and one inner bucket. Because there is only one possible intransitive triad for each dyad, then $\bar{s} = 1$. Figure 1 shows the set of all possible buckets for individuals 1, 2 and 3, where for simplicity, we consider a bounded space of $\mathbf{u}$. Panel (a) shows that there are 27 possible buckets represented by the different cubes in the three dimensional graph, which are associated with either unique or multiple equilibria (depending on whether the cubes are in the outer or inner buckets). Panel (b) in the same figure shows all the possible equilibrium networks (up to isomorphisms).*

To exemplify the relationship between buckets and multiplicity regions, Panel (a) highlights in blue the cubes representing outer buckets for all individuals in which only the empty network is an equilibrium outcome. Those blue buckets are all part of the same multiplicity region (where only the empty network is an equilibrium outcome). In addition to being generated by the outer buckets, the empty network can also arise from an inner bucket with the potential to generate multiple equilibria. For instance, the empty network can also occur in the most inner cube (composed of inner buckets), where the empty and complete networks are possible equilibria. This example shows that the same equilibrium outcome can belong to different multiplicity regions composed of different buckets.

The fact that multiplicity regions are made up of buckets guarantees invariability of the shape of the regions to small changes in parameters θ. Intuitively, changes in parameters θ can only affect the shape of the buckets, not the number of partitions of $\mathbb{R}$. Therefore, the number of multiplicity regions does not change when the parameters change. The above discussion matters for identification because the invariability of the number of multiplicity regions to changes in the parameters θ implies that the selection probabilities are independent of changes in the parameters. The following Proposition shows that under the proposed framework, the likelihood in (10) is locally identified.

Theorem 1 (Identification)**.** *Let Assumption 1 hold. Then, the true vector of payoff parameters θ_0 and the selection probabilities $h_1, \ldots, h_J$ with $J = |\mathbb{D}_N|$ are locally identified for all $\mathbf{d}$ from the likelihood $P(\mathbf{d}; \theta, h)$ at (θ_0, h_0) if the following matrix is full rank*

Figure 1: Multiplicity Regions and Possible Equilibrium Networks



Panel (a) Panel (b)

Note: This example shows that multiplicity regions are composed by buckets. Moreover, it exemplifies how multiple equilibria can arise when the utility function (1) includes externalities of the form $s_{ij}(\mathbf{d}) = \sum_k d_{ik}d_{kj}$. Panel (a) displays the 3D space of the errors u_{ij} for individuals $1, 2$, and 3. The blue cubes represent different buckets that form a multiplicity region where only the empty network is an equilibrium. Panel (b) shows all the possible equilibrium networks -up to isomorphisms- for this simplified example with three individuals.

$$\sum_{m\in M}\begin{bmatrix} \frac{1}{H_m}\frac{\partial G_{m1}}{\partial\gamma} & -\frac{G_{m1}}{H_m^2} & \cdots & -\frac{G_{m1}}{H_m^2} & \cdots & -\frac{G_{m1}}{H_m^2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\bar{h}_j}{H_m}\frac{\partial G_{mj}}{\partial\gamma} & -\frac{\bar{h}_j G_{mj}}{H_m^2} & \cdots & \frac{\left(\sum_{i\neq j}\bar{h}_i\right)G_{mj}}{H_m^2} & \cdots & -\frac{\bar{h}_j G_{mj}}{H_m^2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\bar{h}_J}{H_m}\frac{\partial G_{mJ}}{\partial\gamma} & -\frac{\bar{h}_J G_{mJ}}{H_m^2} & \cdots & -\frac{\bar{h}_J G_{mJ}}{H_m^2} & \cdots & \frac{\left(\sum_{i\neq J}\bar{h}_i\right)G_{mJ}}{H_m^2} \end{bmatrix},$$

where $\mathbf{d}_1,\ldots,\mathbf{d}_J$ *is a list of all networks in* $\mathbb{D}_N$, $G_{mj}\equiv G(\mathbf{d}_j,m;\theta)$, $h_j\equiv h_0(\mathbf{d}_j)$, *we define* $\bar{h}_j\equiv h_j/h_1$ *as the relative probability of* $\mathbf{d}_j$ *with respect to* $\mathbf{d}_1$, *and* $H_m\equiv\sum_{\mathbf{d}'\in\mathbb{D}_m}\bar{h}(\mathbf{d}')$ *for each multiplicity region* m. *We can only identify the relative selection probabilities because we have the additional condition that* $\sum_{\mathbf{d}\in\mathbb{D}}h_0(\mathbf{d})=1$. *We chose the normalizing probability to be* h_1 *without loss of generality.*

Theorem 1 shows that, in general, for an arbitrarily large population, the likelihood in (10) contains enough information to point identify both the parameters of interest and selection probabilities. However, in practice, it is unfeasible to construct a probability distribution over a sequence of relatively large networks. The issue is that the space of networks configurations is excessively large even for a moderate number of nodes, such that finding two networks that are isomorphic becomes impractical.[5] To circumvent the issue, we propose to modify the likelihood function in 10 to form a more tractable object. Details are given in Section 3. Another potential concern for identification is that the likelihood of observing certain networks is arbitrarily close to zero. The following remark shows that identification is more robust when the number of network structures associated with zero probability of being observed is larger.

Remark 1. The full rank condition from Theorem 1 is satisfied even if the vector of predetermined probabilities $\mathbf{h}$ has a large number of zero components. Without the loss of generality, we think about the case where there is only one multiplicity region where the networks $\mathbf{d}_1$ and $\mathbf{d}_2$ are possible equilibria. We assume that only the selection probabilities for those two network structures, h_1 and h_2, are the different from zero. For any $\mathbf{d}_j\in\{\mathbf{d}_1,...,\mathbf{d}_J\}$ with $h(\mathbf{d}_j)=0$, all the entries in row j are zero except for the one on the main diagonal. Thus, given that a diagonal matrix is full rank by definition, whether or not the matrix in Theorem 1 is full rank, will only depend on the entrances of the matrix associated with the equilibrium networks with non-zero selection probabilities. For a better illustration, we applied the row transformation to the rank matrix, which results into the following matrix

[5] Obtaining an empirical analogue of a probability distribution over networks would require observing the same network configurations many times to compute relative frequencies, which becomes even more impractical.

$$\begin{bmatrix} \frac{1}{H}\frac{\partial G_1}{\partial \gamma} & -\frac{G_1}{H^2} & \cdots & 0 & \cdots & 0 \\ \frac{\bar{h}_2}{H}\frac{\partial G_2}{\partial \gamma} & \frac{G_2}{H^2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{(1+\bar{h}_2)G_j}{H^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \frac{(1+\bar{h}_2)G_J}{H^2} \end{bmatrix}.$$

One can easily verify the rank condition is satisfied if the 2-by-2 upper left submatrix is full rank, which only depends on the equilibrium with non-zero selection probability.

3 Bayesian Algorithm

In this section we outline our Bayesian specification that allows us to obtain statistical inference on payoff parameters. We set the stage by introducing our assumptions to obtain a tractable Bayesian specification of the estimation problem. These assumptions concern the unobservable $\mathbf{A}$ as well as its relationship to covariates $\mathbf{X}$. Specifically, in this context Pelican and Graham (2020) essentially treat $\mathbf{A}$ as fixed effects in the sense that no assumptions are made about the distribution of $\mathbf{A}$ conditional on $\mathbf{X}$. On the other hand, we take a correlated random effects approach (Mundlak, 1978; Chamberlain, 1982) that maintains the essential features of the problem while making estimation simpler, particularly in a Bayesian framework. Moreover, the random effects approach allows us to provide Bayesian inference for the individual heterogeneity parameters while avoiding non-standard asymptotic theory caused by the incidental parameter issue (Yan et al., 2019). The advantage of using the random effects approach can also be extended to longitudinal panel data models as in Fernández-Val and Weidner (2016), where the incidental parameter issue causes non-standard asymptotic biases.

We introduce the following two assumptions on the distribution of $\mathbf{A}$ conditional on $\mathbf{X}$.

Assumption 2 (Conditionally independent)**.** For all $i = 1, \ldots, N$ and $j = 1, \ldots, N$ with $i \neq j$, it holds that (i) $A_i \perp A_j | X_i, X_j$ and (ii) $f(A_i|X_i, X_j) = f(A_i|X_i)$.

Assumption 3 (Correlated random effects)**.** For all i, we have $A_i|X_i \sim \mathcal{N}(\phi' X_i, \sigma^2)$, where ϕ is a K-dimensional vector of coefficients and σ^2 is a scalar variance.

Assumption 2.(i) simply states that the in- and out-degree heterogeneity effects are independent across individuals once you condition on the covariates of each dyad. Assumption 2.(ii) then states that the joint distribution of these effects will only depend on each individual's covariates. Finally, Assumption 3 is a correlated random effects specification for a network framework, similar to that in Mundlak (1978). This device is widely used in nonlinear panel data methods in order to deal with unobserved heterogeneity (see, e.g., Chapter 11 of

Wooldridge, 2010). In fact, Assumption 2.(*i*) together with linearity in the conditional expectation of A_i, which is part of Assumption 3, is enough to satisfy Assumption 2.(*ii*) without loss of generality. To see this, note that a more general device such as that in Chamberlain (1980) would set $A_i|X_i, X_j \sim \mathcal{N}(\phi_1' X_i + \phi_2' X_j, \sigma^2)$ for each dyad. However, as only the sum of A_i and A_j appears in the marginal utility equation for any dyad, we cannot identify both ϕ_1 and ϕ_2 separately; we can only identify their sum. Thus, we do not gain more flexibility by including both sets of covariates.

Our results so far do not depend on the specific form of the distribution of idiosyncratic errors $f_U(\mathbf{u})$. As mentioned previously, while a standard assumption in the network formation literature is to specify this distribution as logistic, we instead assume that $\mathbf{U}$ is i.i.d. across dyads with a standard normal distribution such that

$$f_U(\mathbf{u}) = \prod_{i=1}^{N} \prod_{j<i} \phi(u_{ij}) \,,$$

where $\phi(\cdot)$ is the standard normal density. This normalization can be achieved by setting the distributions of the original $\tilde{U}_{ij}$ as $\mathcal{N}(0, 2)$ independently across dyads. Setting the variance of this distribution to unity can be seen as an identifying restriction, which is standard in models with binary dependent variables (see, e.g., pp. 476, Cameron and Trivedi, 2005).[6] Indeed, as we only observe whether a specific link was formed or not, all scalings of the idiosyncratic shocks will be observationally equivalent. This assumption results in a normal likelihood for the latent variables and plays well with Assumption . As is standard in Bayesian analysis, other robust alternatives are available and can be incorporated easily into the resulting sampling scheme.

Define $u_{ij}^* \equiv u_{ij} - A_i - A_j - W_{ij}'\lambda$ for $i = 1, \ldots, N$ and $j = 1, \ldots, N$ with $i \neq j$. Using Assumption 3, we can write $A_i = X_i'\phi + a_i$ for all i, where $a_i \sim \mathcal{N}(0, \sigma^2)$ and a_i is independent of $\mathbf{X}$. Replacing into the definition of u_{ij}^*, we have

$$u_{ij}^* = u_{ij} - X_i'\phi - X_j'\phi - W_{ij}'\lambda - a_i - a_j \,. \tag{11}$$

Following the way the elements of an adjacency matrix $d \in \mathbb{D}^N$ are indexed, we can first stack (11) across rows i for a given j to obtain

$$u_{-j}^* = u_{-j} - X_{-j}\phi - \iota_{N-1} X_j'\phi - W_{-j}\lambda - a_{-j} - \iota_{N-1} a_j, \quad j = 1, \ldots, N;$$

where ι_{N-1} is an $(N-1)$-dimensional vector of ones, u_{-j}^* and a_{-j} are $N-1$-dimensional vectors, X_{-j} is a

[6] The use of a standard logistic density in the network formation literature, which sets the variance of u_{ij} equal to $\pi^2/3$, is another such identifying restriction. Of course, other restrictions are possible, such as fixing the value of one coefficient or the sum of the coefficients. However, these restrictions would impact the meaning and interpretation of preference parameters, and so we choose to fix the variance of shocks instead.

$(N-1) \times K$ matrix, W_{-j} is a $(N-1) \times K^2$ matrix and we define

$$u^*_{-j} = \begin{bmatrix} u^*_{1j} \\ \vdots \\ u^*_{j-1,j} \\ u^*_{j+1,j} \\ \vdots \\ u^*_{Nj} \end{bmatrix}, X_{-j} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_{j-1} \\ X'_{j+1} \\ \vdots \\ X'_N \end{bmatrix}, W_{-j} = \begin{bmatrix} W'_{1j} \\ \vdots \\ W'_{j-1,j} \\ W'_{j+1,j} \\ \vdots \\ W'_{Nj} \end{bmatrix}, a_{-j} = \begin{bmatrix} a_1 \\ \vdots \\ a_{j-1} \\ a_{j+1} \\ \vdots \\ a_N \end{bmatrix}, u_{-j} = \begin{bmatrix} u_{1j} \\ \vdots \\ u_{j-1,j} \\ u_{j+1,j} \\ \vdots \\ u_{Nj} \end{bmatrix}.$$

We can then stack across the index j to obtain

$$\mathbf{u}^* = \mathbf{u} - (F+G)\mathbf{X}\phi - \mathbf{W}\lambda - (F+G)\mathbf{a}\,,$$

where $\mathbf{u}^*$ and $\mathbf{u}$ are $N(N-1)$-dimensional vectors, $\mathbf{a}$ is an N-dimensional vector, F and G are $N(N-1) \times N$ matrices, $\mathbf{X}$ is a $N \times K$ matrix, $\mathbf{W}$ is a $N(N-1) \times K^2$ matrix,

$$\mathbf{u}^* = \begin{bmatrix} u^*_{-1} \\ \vdots \\ u^*_{-N} \end{bmatrix}, F = \begin{bmatrix} F_{-1} \\ \vdots \\ F_{-N} \end{bmatrix}, X = \begin{bmatrix} X'_1 \\ \vdots \\ X'_N \end{bmatrix}, \mathbf{W} = \begin{bmatrix} W_{-1} \\ \vdots \\ W_{-N} \end{bmatrix}, a = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_{-1} \\ \vdots \\ u_{-N} \end{bmatrix},$$

$G = I_N \otimes \iota_{N-1}$ and F_{-j} is an identity matrix of order N with column j removed. Using basis vectors $e_i \in \mathbb{R}^N$ that have a 1 at component i and zeros everywhere else, we can write $F_{-j} = [e_1 \cdots e_{j-1}\, e_{j+1} \cdots e_N]'$.

Finally, we can condense the resulting expression by defining $\widetilde{\mathbf{X}} = [H\mathbf{X}\ W]$, $H = F + G$ and $\beta = [\phi'\ \lambda']'$ to obtain

$$\mathbf{u}^* = \mathbf{u} - \widetilde{\mathbf{X}}\beta - H\mathbf{a}\,. \tag{12}$$

As discussed in Section 2, we ultimately use $t_{ij} \equiv u^*_{ij}/\gamma$ as it contains all sample information relevant for estimating utility parameters. With this definition of latent variables, from (3) we have that an equilibrium network $\mathbf{d} \in \mathbb{D}(\mathbf{t})$ satisfies for all dyads ij,

$$d_{ij} = d_{ji} = \mathbb{I}(s_{ij}(\mathbf{d}) \geq t_{ij})$$

In this way, we can separate the effects of degree-heterogeneity and homophily on marginal utility from those of the externalities. Furthermore, we see that this simplifies our definitions for buckets and multiplicity regions given in Section 2. That is, we can simply re-define buckets and multiplicity regions from the support

of original shocks $\mathbf{u}$ to the transformed $\mathbf{t}$. This change in definition modifies the buckets to be of the form $(-\infty, s_0]$, (s_L, ∞), or $(s_l, \gamma s_{l+1}]$ for $l = 0, \ldots, L-1$, where μ_{ij} and γ are no longer relevant (this change in expectation and scale is included as part of the distribution of $\mathbf{t}$ after the transformation). We can also update our definitions of the equilibrium selection distribution $\mathcal{N}(\cdot)$, as, conditional on $\mathbf{t}$, all of the multiplicity in equilibria come from the dyad-level shocks t_{ij} falling into regions that only depend only on the externality values $s(\mathbf{d}) = (s_{12}(\mathbf{d}), \ldots, s_{N(N-1)/2}(\mathbf{d}))$ and their support. With this change, the equilibrium selection distribution becomes

$$\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta, h) = \mathcal{N}(\mathbf{d}, \mathbf{t}; h) = \sum_{m \in M} \frac{h(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}_m} h(\mathbf{d}')} \mathbb{I}(\mathbf{t} \in m) g(s(\mathbf{d}), \mathbf{t}) \tag{13}$$

where now

$$g(s(\mathbf{d}), \mathbf{t}) = \prod_{i=1}^{N} \prod_{j<i} \mathbb{I}\left[s_{ij}(\mathbf{d}) \geq t_{ij}\right]^{d_{ij}} \cdot \mathbb{I}\left[s_{ij}(\mathbf{d}) < t_{ij}\right]^{1-d_{ij}}$$

We can then express the joint distribution of $\mathbf{d}$ and $\mathbf{t}$ as

$$P(\mathbf{d}, \mathbf{t}; \theta, h) = \mathcal{N}(\mathbf{d}, \mathbf{t}; h) f(\mathbf{t} | \beta, \mathbf{a}, \sigma^2, \gamma) \tag{14}$$

where the conditional distribution $f(\mathbf{t}|\beta, \mathbf{a}, \sigma^2, \gamma)$ is multivariate normal:

$$\mathbf{t} | \beta, \mathbf{a}, \sigma^2, \gamma \sim \mathcal{N}\left(-\frac{\tilde{X}\beta + H\mathbf{a}}{\gamma}, \frac{1}{\gamma^2}\right) \tag{15}$$

To complete a Bayesian specification of the problem, we let the joint prior distribution be

$$\pi(\mathbf{t}, \beta, \mathbf{a}, \sigma^2, \gamma) = f(\mathbf{t}|\beta, \mathbf{a}, \sigma^2, \gamma) \pi(\beta) \pi(\mathbf{a}|\sigma^2) \pi(\sigma^2) \pi(\gamma).$$

As we assume in our identification results that $\gamma \geq 0$, we can incorporate this restriction into our prior $\pi(\gamma)$ by choosing a distribution with support on the non-negative reals or by truncating to this part of the parameter space. We will assume the standard conditionally conjugate priors given as

$$\begin{aligned} \beta &\sim \mathcal{N}(\underline{\beta}, \underline{B}), \\ \sigma^2 &\sim \mathcal{IG}(\underline{v}/2, \underline{\sigma}^2/2), \end{aligned}$$

where $\mathcal{IG}$ is the inverse Gamma distribution. Symbols with an underline are prior hyperparameters and the updated (posterior) quantities will be denoted using an overline. We obtain tractable estimation of γ by using

the generalized inverse normal prior proposed in Robert (1991).

$$\frac{1}{\gamma} \sim \mathcal{GIN}^{+}(\underline{\nu}, \underline{\gamma}, \underline{\tau}^2)$$

As noted by Robert (1991), this prior is conjugate for mean mixture models where the mean is a multiple of the standard deviation. The generalized inverse normal distribution can place substantially large density on $(0, \infty)$ for given parameter values. With large sample size N, the density placed in this part of the support increases. However, as a truncated version of the distribution would maintain conjugacy, we directly use the truncated generalized inverse normal to the positive reals as our prior. As noted below, this corresponds to the notation $\mathcal{GIN}^{+}$.

An important consequence of including the latent $\mathbf{t}$ into our sampling scheme is that now the likelihood (14) separates into two terms. The preference parameters enter the likelihood through the second term. This immediately implies that the conditional posterior distributions of $\beta, \mathbf{a}$, σ^2 and γ will only depend on the networks through values of $\mathbf{t}$. To see this, note that the joint conditional posterior is

$$\begin{aligned}\pi(\beta, \mathbf{a}, \sigma^2, \gamma | \mathbf{d}, \mathbf{t}) &\propto \mathcal{N}(\mathbf{d}, \mathbf{t}) f(\mathbf{t}, \beta, \mathbf{a}, \sigma^2, \gamma) \\ &\propto f(\mathbf{t}|\beta, \mathbf{a}, \sigma^2, \gamma)\pi(\beta)\pi(\mathbf{a}|\sigma^2)\pi(\sigma^2)\pi(\gamma)\end{aligned}$$

Thus, we can obtain a Gibbs sampling algorithm for $\beta, \mathbf{a}$, σ^2, and γ by focusing on the posterior with likelihood given by the conditional distribution of $\mathbf{t}$. Combining the likelihood in (14) with our prior specification allows us to find the posterior distribution of all quantities of interest. For the coefficients β and random effects $\mathbf{a}$, we can obtain a joint posterior as $\pi(\beta, \mathbf{a}|\mathbf{t}, \sigma^2, \gamma, \mathbf{d}) = \pi(\beta|\mathbf{t}, \sigma^2, \gamma, \mathbf{d})\pi(\mathbf{a}|\mathbf{t}, \beta, \sigma^2, \gamma, \mathbf{d})$. Standard updates (see, e.g., Chib, 2008) result in posteriors

$$\begin{aligned}\beta|\mathbf{t}, \sigma^2, \gamma, \mathbf{d} &\sim \mathcal{N}(-\bar{\beta}, \bar{B}), \\ \mathbf{a}|\mathbf{t}, \beta, \sigma^2, \gamma, \mathbf{d} &\sim \mathcal{N}(-\gamma\bar{\mathbf{a}}, \bar{V}), \\ \sigma^2|\mathbf{t}, \beta, \mathbf{a}, \mathbf{d} &\sim \mathcal{IG}(\bar{v}, \bar{\sigma}^2)\end{aligned} \tag{16}$$

As the chosen generalized inverse normal prior is conjugate for this problem, the posterior of the externality parameter is

$$\frac{1}{\gamma}|\mathbf{t}, \beta, \sigma^2, \mathbf{d} \sim \mathcal{GIN}^{+}(\bar{\nu}, \bar{\gamma}, \bar{\tau}^2), \tag{17}$$

Define $\Omega \equiv I_{N(N-1)} + \sigma^2 HH'$. The updated posterior hyperparameters are given by

$$\begin{aligned}
\bar{B} &= (\widetilde{\mathbf{X}}'\Omega^{-1}\widetilde{\mathbf{X}} + \underline{B}^{-1})^{-1}\,, & \bar{v} &= \underline{v} + N\,, \\
\bar{\beta} &= \bar{B}(\gamma\widetilde{\mathbf{X}}'\Omega^{-1}\mathbf{t} + \underline{B}^{-1}\underline{\beta})\,, & \bar{\nu} &= \underline{\nu} + N\,, \\
\bar{V} &= (H'H + \sigma^{-2}I_N)^{-1}\,, & \bar{\gamma} &= \frac{\underline{\tau}^2\mathbf{t}'(\tilde{X}\beta + H\mathbf{a}) + \underline{\gamma}}{\underline{\tau}^2\mathbf{t}'\mathbf{t} + 1}\,, \\
\bar{\mathbf{a}} &= \bar{V}H'(\mathbf{t} - \widetilde{\mathbf{X}}\beta)\,, & \bar{\tau}^2 &= \frac{\underline{\tau}^2}{\underline{\tau}^2\mathbf{t}'\mathbf{t} + 1}\,, \\
\bar{\sigma}^2 &= \mathbf{a}'\mathbf{a} + \underline{\sigma}^2\,.
\end{aligned} \tag{18}$$

A Gibbs sampler that cycles through the conditional posteriors in (16) and (17) requires obtaining samples from the (truncated) generalized inverse normal distribution. To the best of our knowledge, no such algorithm exists in the literature. Therefore, an additional contribution of our paper is to provide an efficient sampling algorithm for this distribution and its truncated variants. We now provide a short digression on the theoretical results supporting this sampling algorithm, but leave the full extension to Appendix D.

To establish the main result supporting our sampling algorithm for the generalized inverse normal distribution, we require the definition of T_c-concavity (see Hörmann and Leydold, 2014). We then provide a proposition establishing that the kernel of this distribution is T_c concave with a value of $c = -1/\nu$.

Definition 4 (Hörmann and Leydold, 2014)**.** A function $f(x)$ is called T_c-concave (for $c \neq 0$) if $\text{sgn}(c)f(x)^c$ is concave. It is called T_0-concave if it is log-concave.

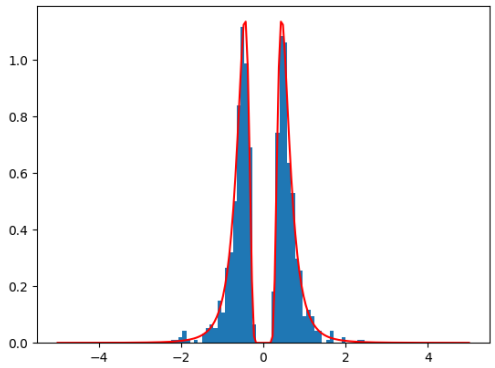
Proposition 3. *The kernel of the standard generalized inverse normal distribution $\mathcal{GIN}(\nu, \gamma, 1)$, given by $g(\cdot; \nu, \gamma)$ is $T_{-1/\nu}$-concave.*

The sampling algorithm we derive in this paper is based on the ratio-of-uniforms method with mode shift proposed by Kinderman and Monahan (1977). Leydold (2001) and Hörmann and Leydold (2014) show that this algorithm is efficient for $T_{-1/2}$-concave distributions and is especially useful in the varying parameters setting that characterizes Bayesian sampling schemes. The following theorem shows that the generalized inverse normal distribution is in fact $T_{-1/2}$-concave in a region of the parameter space, which follows easily from the previous proposition.

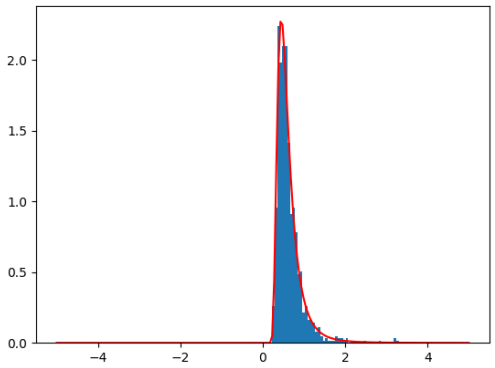
Theorem 2. *The kernel $g(\cdot; \nu, \gamma)$ is $T_{-1/2}$-concave when $\nu > 2$.*

Using these two results, Appendix D presents pseudo-code and the full details of Algorithms 1–3 for sampling from the $\mathcal{GIN}$ distribution. The first two implement the sampling for the truncated variants to either the negative or positive reals, while the third one does so for the full distribution supported on the real line. We

Figure 2: Sampling from the Generalized Inverse Normal distribution (Robert, 1991)



(a) $\mathcal{GIN}(5, 0, 1)$ distribution



(b) $\mathcal{GIN}(5, 0, 1)$ distribution truncated to $\mathbb{R}_+$.

Notes: The solid line is the theoretical density and the histogram represents draws from the distribution using Algorithms 1–3 in Appendix D.

provide Python and `R` routines that implement all generating algorithms described in the paper. A showcase of the draws obtained from the procedures are given in Figure 2.

Remark 2. It is important to acknowledge that our algorithm will not work as is for values of $\nu \in (1, 2]$. One could try to accommodate this part of the parameter space by deriving an envelope function over the log-convex parts of the distribution (e.g., Hörmann and Leydold, 2014, for the generalized inverse Gaussian distribution). However, we can show the restriction of our sampling algorithm to $\nu > 2$ is not of concern if the goal is to carry out Bayesian estimation. As in Eq. (18), when using a $\mathcal{GIN}(\underline{\nu}, \underline{\gamma}, \underline{\tau})$ distribution as the prior for a scale, the posterior distribution is also generalized inverse normal with updated parameter $\bar{\nu} = \underline{\nu} + N$, where N is the sample size. Thus, even an uninformative prior that sets $\underline{\nu} = 1 + \varepsilon$ for some small $\varepsilon > 0$, would have $\bar{\nu} = 1 + N + \varepsilon > 2$, as long as $N \geq 1$; i.e., when there is at least one observation. This means that for posterior sampling, we are effectively always drawing from distributions where $\nu > 2$. Thus, in this paper we do not pursue such an optimization for the algorithm and defer it for future research.

To complete our sampling scheme, we need to obtain the conditional posterior for $\mathbf{t}$, which is given by

$$\begin{aligned}\pi(\mathbf{t}|\mathbf{d}, \beta, \mathbf{a}, \sigma^2, \gamma) &\propto \mathcal{N}(\mathbf{d}, \mathbf{t}) f(\mathbf{t}|\beta, \mathbf{a}, \sigma^2) \\ &= \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}_m} h_0(\mathbf{d}')} \mathbb{I}(\mathbf{t} \in m) g(\mathbf{d}, \mathbf{t}) f(\mathbf{t}|\beta, \mathbf{a}, \sigma^2, \gamma) \end{aligned} \tag{19}$$

To make sense of this expression, first note that $g(\mathbf{d}, \mathbf{t})$ imposes constraints on each t_{ij}, as they need to be such that they satisfy our equilibrium concept for a given equilibrium network $\mathbf{d}$. Furthermore, this latent component

of marginal utility should be such that equilibrium solutions arising from $\mathbf{t}$ belong to the same multiplicity region as $\mathbf{d}$. Finally, as $\mathbf{d}$ can be an equilibrium network for several multiplicity regions, the posterior of $\mathbf{t}$ weights each of these regions according to the underlying network distribution from Assumption 1. Putting everything together implies that the posterior of $\mathbf{t}$ is a mixture of truncated multivariate normals across multiplicity regions, with mixture weights given by the conditional probability of observing network $\mathbf{d}$ in each multiplicity region.[7]

The main issue in obtaining draws for $\mathbf{t}$ is that we must now invert the mapping of shocks to equilibrium networks. That is, we need to ensure sampled $\mathbf{t}$ satisfy the equilibrium restrictions for a given network. To accomplish this, we need a way to evaluate the equilibrium selection probabilities; i.e., we need to compute $\mathcal{N}(\mathbf{d}, \mathbf{t})$. As this is unfeasible given the current setup, we introduce one further simplifying assumption in the form a composite likelihood that replaces the full likelihood (14) for sampling the two remaining components.

3.1 Composite Likelihood Function

Let $\mathbf{d}_k^{(K)}$ be the subgraph of size K induced by the network $\mathbf{d}$. Construct the sequence of N choose K subnetworks $\{\mathbf{d}_1^{(K)}, \ldots, \mathbf{d}_{N_K}^{(K)}\}$, where N_K is the total number of subnetworks that emerge when forming all possible combinations of nodes of size K. Therefore, the marginal probability of observing network $\mathbf{d}$ can be constructed by the joint probability of observing the sequence of subnetworks $\{\mathbf{d}_1^{(K)}, \ldots, \mathbf{d}_{N_K}^{(K)}\}$. We impose the following set of assumptions on the behavior of individuals when deciding how to form connections in each of the N_K subnetworks.

Assumption 4 (Subnetworks Individuals' Behavior)**.** (i) each subnetwork $\mathbf{d}_k^{(K)}$ forms following the Pairwise Stability criteria in Definition 1, and (ii) the game structure and payoffs are the same for any subgraph.

Assumption 4 guarantees that all the form of the likelihood of observing the complete network is inherited by the subnetwork structures, see Schweinberger and Handcock (2015) for a similar set of assumptions on large networks formed following an Exponential Random Graph model. Under these assumptions, the marginal likelihood of each subgraph will be given by

$$P(\mathbf{d}_k^{(K)}; \theta, \tilde{h}_0) = \sum_{m \in M_K} \frac{\tilde{h}_0(\mathbf{d}_k^{(K)})}{\sum_{\mathbf{d}' \in \mathbb{D}_m} \tilde{h}_0(\mathbf{d}')} \int \mathbb{I}(\mathbf{u}_k^{(K)*} \in m) g(\mathbf{d}_k^{(K)}, \mathbf{u}_k^{(K)*}; \gamma) f(\mathbf{u}_k^{(K)*} | \theta) \mathrm{d}\mathbf{u}_k^{(K)*},$$

for all the possible multiplicity regions M_K for subnetworks of size K, and for all k in the sequence of N_K subnetworks. The distribution $\tilde{h}_0$ is a predetermined distribution over subnetworks and can be expressed in terms of the original predetermined distribution h_0 (details can be found in Appendix A).

[7] If a network $\mathbf{d}$ cannot arise as an equilibrium network for shocks in multiplicity region m (i.e., $\mathbf{d} \notin \mathbb{D}_m$), then the associated mixture weight of multiplicity region m is 0 because $\mathbf{t} \notin m$.

Characterizing the joint distribution over the sequence $\{\mathbf{d}_1^{(K)}, \ldots, \mathbf{d}_{N_K}^{(K)}\}$ is still unfeasible. We follow the idea in Graham (2020) and propose a composite likelihood that is formed by multiplying the marginal distributions of N_K subgraphs forming $\mathbf{d}$,

$$\tilde{P}(\mathbf{d}; \theta, \tilde{h}_0) = \left[\prod_{k=1}^{N_K} P(\mathbf{d}_k^{(K)}; \theta, \tilde{h}_0) \right]^{\omega_k}, \tag{20}$$

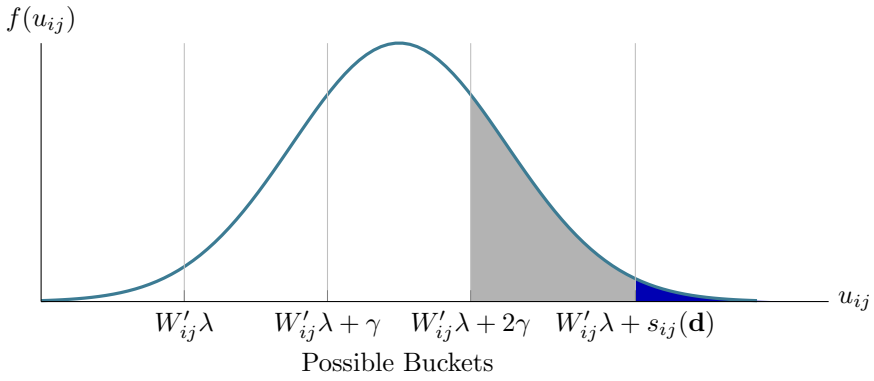
where ω_k weights the importance of the marginal of tetrad k in the composite likelihood. The weights in equation (20) have to be such that $\sum_k^K \omega_k = 1$. The advantage of a composite likelihood is that, even though it fails to correctly represent the dependence structures across different subgraphs, if the marginal distributions are correctly specified, estimators for θ based on the composite likelihood are consistent for the true population parameter (Cox and Reid, 2004; Varin et al., 2011). In addition, by choosing values of K that make the space of potential network configurations manageable, it is possible to provide a tractable probability distribution over the set of possible equilibrium networks. As expected, when $K \to N$, the composite likelihood in (20) collapses to (9), and we are back to the situation where the identification of selection probabilities is an intractable problem.

In our case, we cannot guarantee the correct specification of the subgraph marginal distributions. The reason is that when we split the complete network $\mathbf{d}$ into subnetworks $\mathbf{d}_k^{(K)}$, the externality values in the subnetworks $s_{ij}(\mathbf{d}_k^{(K)})$ will underestimate the true externality values $s_{ij}(\mathbf{d})$ for several interesting externality functions. For example, if s_{ij} represents taste for completing triangles, the maximum number of indirect connections that a node can have is $K-2$, while in the complete network that number is given by $N-2$. The following example illustrates the approximation error that arises when using the misspecified composite likelihood.

Example 3. *Consider the case where $A_i = 0$ for all i, externalities are given by intransitive triads $s_{ij}(\mathbf{d}) = d_{ik}d_{kj}$, $\gamma \geq 0$ and assume $K = 4$. In this case, there are only two possible indirect connections for each i in the subnetworks $\mathbf{d}_k^{(4)}$, while in the complete network that number is $N-2$. Figure 3 illustrates the misspecification error induced by the composite likelihood under the assumption that u_{ij} are iid from a normal distribution. The error will only arise when we underestimate the total number of indirect links between i and j. In the Figure, we represent the true number of indirect connections by $s_{ij}(\mathbf{d})$. Because the true externality value is located to the right of 2 –the maximum number of indirect links in a tetrad subnetworks– there will be a misspecification error in this case. The error is given by the log of the ratio between the blue and the gray areas. As discussed by White (1982), we can interpret this ratio as the Kullback and Leibler (1951) (KL) divergence criteria. As we would expect, when we let the size of the subnetworks K approach the total number of nodes N, the number of intransitive triads will approach the true value. Therefore, the gray and blue areas ratio will approach one, meaning zero approximation error.*

From Figure 3 it is easy to see that both payoff parameters have a direct influence on the specification error.

Figure 3: Example of the Composite Likelihood Misspecification Error



Note: This example illustrates the misspecification error that arises when using the incorrectly specified composite likelihood. We separate the space of the shocks for arbitrary individuals i and j into buckets following Definition 3. When using the subnetworks to separate the shocks' space, the further right bucket starts in 2γ. Let $s_{ij}(\mathbf{d})$ represent the true number of indirect connections between i and j. Then, under the assumption that u_{ij} are iid from a normal distribution, the misspecification error arises for underestimating the total amount of indirect links, and it is given by the ratio between the blue and the gray area.

The utility parameters change the limits of the outer bucket delimited by $\lambda+\gamma$,$W'_{ij}\lambda+2\gamma$ and its distance to the true externality value. Because the distribution determining the gray and the blue areas is non-linear, changes in the limits of the outer bucket will change the specification error even for proportional changes such as those that happen after varying the value of λ.

Example 3 discusses the approximation error that arises because of the misspecification induced by the fact that we are underestimating the true externalities value $s_{ij}(\mathbf{d})$. However, from equation (9), it follows that the likelihood is also a function of the equilibrium selection probabilities h_0. The dependence of the complete likelihood on those probabilities raises the question of whether the composite likelihood approximation also induces approximation errors when trying to identify h_0. The following proposition shows that, because of the multiplicative form, the misspecification error only affects the utility function parameters λ and γ. Proposition 4 also characterizes the approximation error. As expected, the error for each tetrad k depends on the distance between the externality values $s_{ij}(\mathbf{d})$ and $s_{ij}(\mathbf{d}_k^{(K)})$ weighted by ω_k for each $k \in K$.

Proposition 4 (Misspecification Error)**.** *Let (θ_0, h_0) be the true parameter vector, and denote the misspecification error as $\mathcal{KL}(\mathbf{d};\theta_0,h_0) = \log P(\mathbf{d};\theta_0,h_0) - \log \tilde{P}(\mathbf{d};\theta_0,h_0)$, where we use $\mathcal{KL}$ to represent the Kullback-Leibler divergence measure. Let Assumption 4 hold. Then,*

$$\mathcal{KL}(\mathbf{d};\theta_0,h_0) = -\left[\sum_{k}^{N_K}\omega_k \log q_k(\mathbf{d}_k^{(K)}) + \sum_{k=1}^{N_K}\omega_k \sum_{i=1}^{N}\sum_{j<i}\log\left(\frac{\int_{u_{ij}}\mathbb{I}\left(s_{ij}(\mathbf{d})\geq -t_{ij}\right)f(t_{ij})dt_{ij}}{\int_{t_{ij}}\mathbb{I}\left(s_{ij}(\mathbf{d}_k^{(K)})\geq t_{ij}\right)f(t_{ij})dt_{ij}}\right)^{d_{ij}}\right.$$
$$\left.\times\left(\frac{\int_{t_{ij}}\mathbb{I}\left(s_{ij}(\mathbf{d})< t_{ij}\right)f(t_{ij})dt_{ij}}{\int_{t_{ij}}\mathbb{I}\left(s_{ij}(\mathbf{d}_k^{(K)})< t_{ij}\right)f(t_{ij})dt_{ij}}\right)^{1-d_{ij}}\right],$$

We provide the proof of Proposition 4 in Appendix A. In the proof, we also characterize the marginal effect of changing λ and γ on the KL divergence. Importantly, we find that the marginal effect of the utility function parameters can be decreased by choosing the weights ω_k for each subnetwork k. Intuitively, we want to have higher weights on subnetworks for which the distance between $s_{ij}(\mathbf{d})$ and $s_{ij}(\mathbf{d}_k^{(K)})$ is smaller. The subnetworks with low values of $s_{ij}(\mathbf{d}) - s_{ij}(\mathbf{d}_k^{(K)})$ can be found by inspecting the connectivity of the subnetwork k with the rest of the network. Therefore, the subnetworks with lower external connectivity should have a larger weight than those with higher connectivity.

In our Bayesian specification, the use of the composite likelihood introduces an additional consideration for sampling. Given that the preference parameters carry a structural interpretation in Equation (1), we decide against allowing for specifications where the preference parameters vary across tetrads. Thus, all of the information present in the sequence of $\mathbf{t}^{(K)}$ can be used to sample from the distribution of β, $\mathbf{a}$, σ^2 and γ. This is equivalent to pooling the information present across tetrads (with associated weights ω) but maintaining the specification given in (12). This process can be implemented by combining the information from $\mathbf{X}$ and $\mathbf{a}$ (information on each *node*) across tetrads. This is the process that we undertake in the empirical results for our Bayesian estimation.

4 Empirical Application

This section presents an empirical application that highlights the relevance and practicality of our proposed Bayesian estimator. We use data from Banerjee et al. (2013), whose primary objective is to study learning diffusion through social networks in the context of microfinance participation for 77 villages in Karnataka, India. Banerjee et al. (2013) examines microfinance take-up diffusion in a network, arguing that the information seed is exogenous because it is decided by a third-party institution that is not explicitly maximizing any individual or aggregate objective function. Interestingly, though, the authors acknowledge that the social network itself is likely to be endogenous. They mention homophily reasons arguing that connected individuals tend to exhibit strong similarities. As we have argued, there are additional strategic consideration that can affect individuals

decisions to form connections such as degree heterogeneity and payoff externalities. Given the availability of granular and rich network information, our data is particularly well-suited to investigate which channels are indeed relevant for individuals to form social ties.

4.1 Network Data

The data comes from the Social Networks and Microfinance project which contains the publicly available data of participation in a program of Bharatha Swamukti Samsthe (BSS), a microfinance institution (MFI) in rural southern Karnataka. The data were collected in 2006 for 77 villages and include information of thirteen possible dimensions in which individuals can be connected, including visiting each other, going to pray, borrowing and lending money and goods, obtaining advise, and giving advise (Banerjee et al., 2013). Additionally, the data contain a village questionnaire and a full census including some information on all households in the villages. Individuals' characteristics include gender, age, religion, caste, sub-caste, mother tongue, whether the individual is a village native, education, work frequency and occupations. There are two levels of aggregation in the data: individuals and households. Given that our objective is to estimate the utility function parameters of individuals making strategic choices, we will use the individual level data to fit our model.

Regarding the construction of the social network, we follow Banerjee et al. (2013) when defining our network of interest. We consider the connections to be undirected, so we use the symmetric version of the adajacency matrices capturing the connections between individuals. Therefore, we consider two individuals to be neighbors in the network if at least one of them mentions the other as a contact in response to some network question. Finally, instead of considering each of the thirteen possibles ways in which people interact in these villages a different network, we consider two people linked if they mentioned each other in at least one type of relationship. Table 2 presents some summary statistics for the social networks of the 20 most populated villages in the sample. The networks consistently show relatively high average degrees between approximately 8 and 10 connections. The levels of transitivity are across the board lower than 0.5, meaning that there are more incomplete triangles than complete ones. These transitivity values present initial evidence that individuals may not have a substantial payoff from completing triangles (Jackson and Rogers, 2007). These networks also feature a relatively high average shortest path length (distance) and a large number of disconnected components. All these attributes are suggestive of highly clustered networks that are sparse across clusters.

4.2 Data on Individual Characteristics

The Social Networks and Microfinance project also includes a battery of variables characterizing individuals in the villages. We have information on gender, individuals' role in the household (head of the household, spouse

Table 2: Summary Network Statistics

Village	Nodes	Average Degree	Transitivity	Average Distance	Components
60	1775	8.67	0.45	4.63	16
28	1612	9.40	0.52	4.36	18
59	1599	8.37	0.44	4.68	20
52	1525	10.28	0.36	4.13	12
71	1387	8.26	0.33	4.44	13
3	1380	7.78	0.35	4.36	21
39	1339	8.96	0.45	4.29	14
29	1337	7.61	0.39	4.53	21
65	1331	9.22	0.30	4.09	11
25	1313	9.33	0.46	4.21	10
64	1286	8.67	0.46	4.46	10
46	1257	7.71	0.39	4.55	12
23	1252	8.40	0.39	4.27	19
36	1214	8.61	0.31	4.07	21
32	1181	9.42	0.42	4.10	15
55	1180	7.84	0.48	4.84	12
76	1154	8.15	0.45	4.48	13
18	1146	9.10	0.38	4.13	5
19	1134	9.19	0.40	4.18	5
40	1097	7.89	0.40	4.58	11

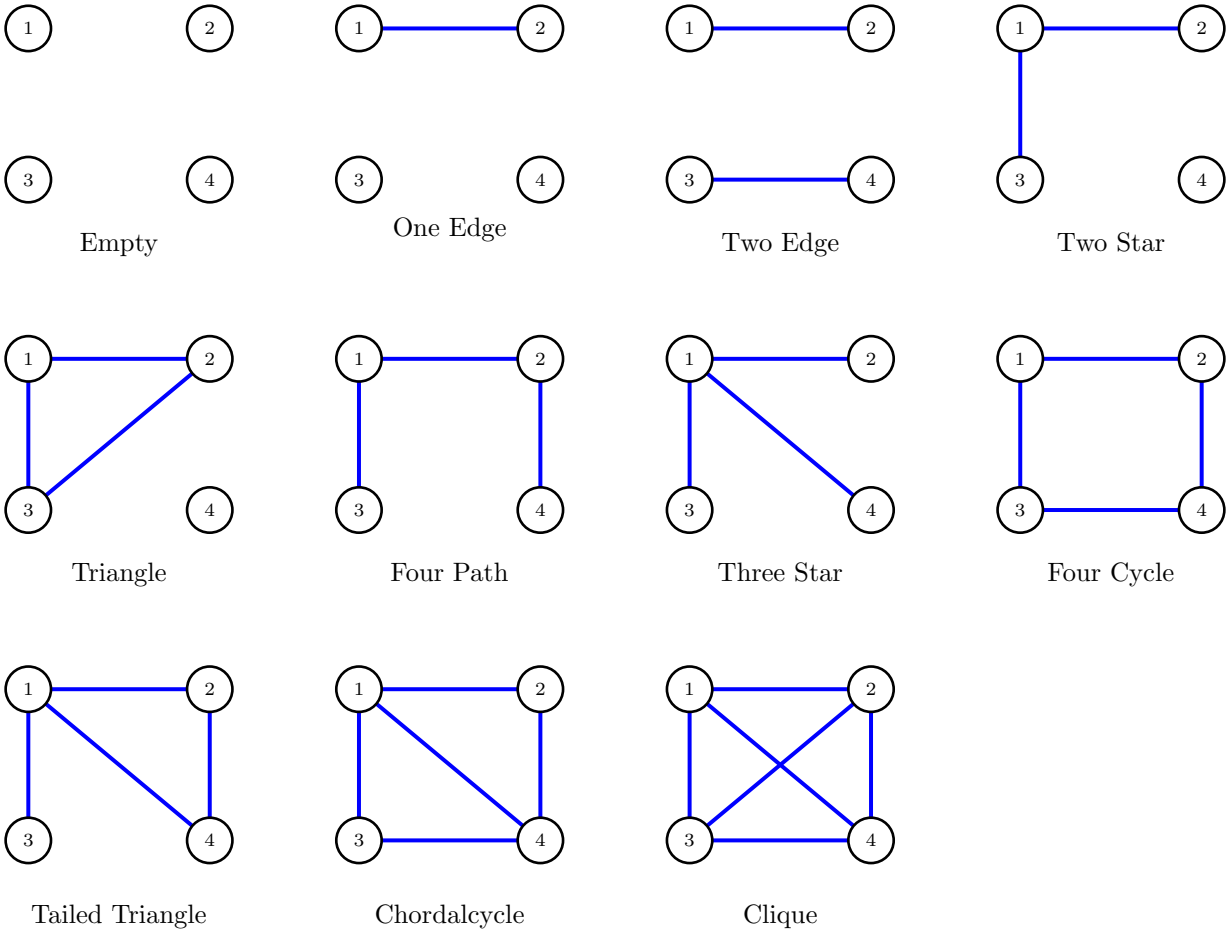
Note: table with the network statistics for the 20 most populated villages out of the 77 villages in the data. The average distance is calculated as the maximum distance in all possible connected components.

of the head, or other), age, religion, cast, sub-cast, languages that the individuals speak, working status, saving behavior, and participation in the financial market. We choose a subset of those characteristics to construct our homophily measure in equation $W_{i,j}$ from Equation (7). In particular, we use the working status, gender, individuals' role in the household, cast, and whether or not the individual is native from the village, and construct a set of dummy variables. Using this information, the measure of homophily is whether two individuals match in the value of each of the dummy variables. If they match, the homophily variable takes the value of zero, and it takes the value of one if they mismatch.

4.3 Subgraphs and Selection Probabilities

As we argue in Section 2.3, it is untractable to estimate probability distribution over a sequence of large networks in practice. The data we are using contains large networks for each village, with a sample size between 780 and 1775 nodes. Section 3.1 presents a solution for this practical issue based on the idea of subgraphs and composite likelihoods. One relevant parameter that we need to choose is the number of subgraphs K. As argued by Graham (2017), working with subgraphs of size four (tetrads) provides enough variation to identify homophily parameters while at the same time having the the advantage of yielding a criterion function that is easy to evaluate and maximize. Tetrad subgraphs also contain enough nodes to provide variation in the number of complete and

Figure 4: Tetrads Isomorphisms



Note: each of the graphs represent one possible tetrad configuration. A tetrad can be wired in up to 64 different ways, but the 11 configurations in this figure are the unique isomorphism classes. Any tetrad can be represented by one of the graphs in this figure up to rotating indexes (Graham, 2017).

incomplete triads. Based on these arguments, we choose $K = 4$ to perform our Bayesian estimation algorithm. An additional convenient feature of working with tetrads, is that it is possible to completely characterize all the possible configuration of the subnetworks. Figure 4 presents the 11 unique isomorphism classes in which tetrads can be wired.

4.3.1 Selection Probability Estimator

Based on the 11 isomorphism classes in Figure 4 and under Assumption 1, it is possible to non-parametrically estimate the selection probabilities for each isomorphism given the observation of one large network. In obtaining these results, as well for all those that follow, we assume for the sake of simplicity that $\Delta\nu_{ij}(\mathbf{d}, U_{ij}) = \Delta\nu_{ji}(\mathbf{d}, U_{ji})$.

Under this simplifying assumption, the pairwise stability criteria in Definition 1 reduces to the requirements that (i) if $d_{ij} = 1$, then $\Delta\nu_{ij}(\mathbf{d}, U_{ij}) \geq 0$, and (ii) if $d_{ij} = 0$, then $\Delta\nu_{ji}(\mathbf{d}, U_{ij}) < 0$. This allows to focus only on one shock per dyad, which greatly reduces the computational complexity of the algorithms presented in Section 3.

The estimator for the selection probabilities is then given as follows: from the observations of a sampled network $\mathbf{d}$ of size n, construct the sequence of n choose 4 possible tetrads in the network and claculate the frequency of each isomorphism. The frequency values are our estimator for the predetermined probabilities $h_0(\mathbf{d}_k^{(4)})$. After estimating the predetermined probabilities, conditional on the form of the externality component $s_{ij}(\mathbf{d})$, it is possible to form the buckets as defined in 3, construct the multiplicity region following proposition 2, and form the quotient $h_0(\mathbf{d}_k^{(4)})/\sum_{d' \in \mathbb{D}_m}$ for each multiplicity region m. For instance, if $s_{ij}(\mathbf{d})$ is the test for completing triangles externality, for each tetrad, there are a total of three values for the possible number of intransitive triads for individual i. Following the same idea as in equation 5, it is possible to construct four buckets. Given that the links are undirected, there are six possible ways in which the upper triangular elements of the tetrads adjacency matrices can be arrange that can be mapped to the isomorphism. Therefore, there are a total of 4,096 possible combinations of matrix configurations and buckets, and we can know exactly to what bucket each isomorphism belongs.

Table 3 presents the estimates of the predetermined probability of each tetrad isomorphism for the three villages with the larger sample sizes. We see that the most likely configuration to emerge is the One Edge followed by the Empty isomorphisms. These two network configuration are the most sparse among all possible isomorphisms, which suggest that the social networks in our sample are relatively sparse. This result is cosistent with the relatively low transitivity index and large average distance from the network statistics presented in Table 2. Based on the predetermined probabilities presented in Table 3, it is then possible to calculate the relative frequency of each isomorphisms with respect to the other network configurations that belong to its multiplicity region. Again, that calculation completely depend on the specification of the payoff externality form in $s_{ij}(\mathbf{d})$.

4.3.2 Misspecification Error Analysis

As we discussed in section 3.1, the composite likelihood approach we propose in this paper is sensitive to potential misspecification errors caused by differences in the values of externalities between the complete network and the subnetworks. However, we also showed the misspecification error decreases when the subnetwork externality values for a dyad ij are not larger than those in the complete network. In particular, for our empirical application we focus on the intransitive triads externalities, $s_{ij}(\mathbf{d}) = d_{ik}d_{kj}$, and subnetworks of size $K = 4$. In this case, in principle, the maximum number of intransitive triads for a dyad ij can be $N - 4$ while the maximum number of subnetwork intransitive triads is 2. The three most populated Villages, 60, 28 and 59, have a total of 1775, 1612 and 1599 nodes, which creates a situation where the misspecification error can be potentially large.

Table 3: Estimators of Tetrads Probabilities

Tetrad	Village 60	Village 28	Village 59
Empty	0.324	0.352	0.358
One Edge	0.449	0.471	0.476
Two Edge	0.049	0.049	0.048
Two Star	0.061	0.047	0.047
Triangle	0.068	0.064	0.056
Four Path	0.009	0.006	0.006
Three Star	0.002	0.001	0.001
Four Cycle	0.000	0.000	0.000
Tailed Triangle	0.007	0.005	0.005
Chordalcycle	0.001	0.001	0.001
Clique	0.029	0.004	0.002

Note: table with the non-parametric estimates of the tetrads probabilities based on the three villages with larger sample sizes. We compute the n choose 4 sequence of tetrads where n represents the number of nodes in each village, and then use that sequence to calculate the proportions of each tetrad isomorphisms.

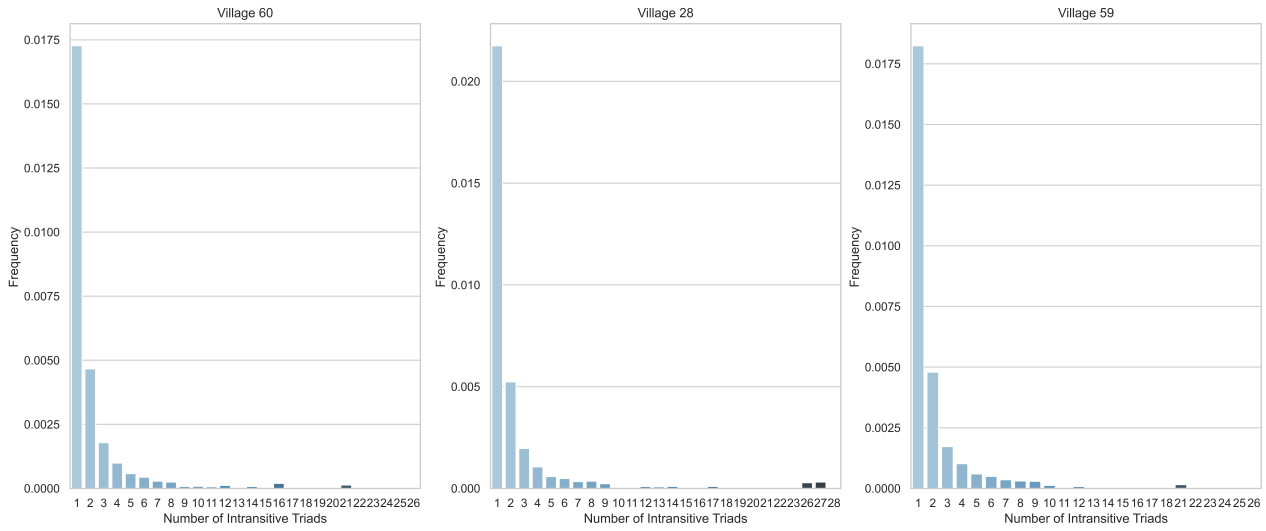
We can empirically check the potential prevalence of misspeficication errors by calculating the distribution of the total number of intransitive triads for all possible dyads. Figure 5 present the empirical frequency of the total number of intransitive triads for Villages 60, 28 and 59. We excluded the values for zero intransitive triads because they represent the largest proportion of dyads. In particular, the proportion of zero intransitive triads is 97.2%, 96.6%, and 97.1% for Villages 60, 28 and 59, respectively. From the figure, we can conclude that less than 1% of dyads have a value of the externalities that is larger than 2. Given that the proportion of dyads for which the value of complete network intransitive triads is larger than the subnetwork maximum intransitive triads, we argue that the misspecification error for the parameters of interest is low, increasing the validity of our empirical results.

4.4 Empirical Results: Simplified Algorithm

This section presents the empirical results of estimating the network formation rule described in section 2.1 without the presence of payoff externalities. This estimators can be seen as the Bayesian-Correlated Effects version of the tetrads fixed effects maximum likelihood estimator proposed by Graham (2017). We focus on the results for the homophily parameters where we fit the network formation model using the observed characteristics described in section 4.2. Table 4 presents the means and standard deviations of the homophily parameters for work status, gender, head of the household, spouse of the head, casts and native variables. For this analysis we focus only on village 60 which contains the largest sample size among all villages.

We find strong evidence of homophily for most of the characteristics in our analysis. The strongest homophily effects happen among same the gender and working status of villagers. This results have repercussions in the

Figure 5: Empirical Frequency of The Number of Triads



Note: figure with the empirical frequency of the total number of intransitive triads for the three villages with larger sample sizes. The values for zero intransitive triads are excluded as the represent a large proportion of dyads. In particular, the proportion of zero intransitive triads is 97.2%, 96.6%, and 97.1% for Villages 60, 28 and 59, respectively.

Table 4: Mean and Standard Deviation for the Posterior Distribution of λ

	Mean	Standard Deviation
Different Work Status	-0.012	0.006
Different Gender	-0.043	0.023
Head of the Household/Others	0.007	0.076
Spouse of the Head/Others	-0.011	0.078
Different Caste	-0.011	0.006
Native and Non-Native	-0.012	0.006

Note: the table presents the mean and standard deviation for the posterior distribution of λ using the Bayesian algorithm described in section 3. The posterior distribution takes into account the fact that there may exists degree heterogeneity in the network formation rule.

context of learning dynamics, because the shape of the network critically depends on the characteristics of the individuals. Therefore, we can expect to see higher levels of information transition across similar people, which can affect policies determining interventions such as for example injection points (Banerjee et al., 2013).

4.5 Empirical Results: Full Algorithm

This section extends the empirical results from the previous section to add the Bayesian posterior distribution of the externality parameter γ. Table 5 presents the mean and standard deviation for the posterior distribution of the parameters of interest θ. In terms of qualitative results, most of the parameters have the same signs and have similar magnitudes. In addition to the homophily parameters, Table 5 presents the results for the externality parameters. We use the taste for completing triangles as the source of externalities. For the village 60 and using

Table 5: Mean and Standard Deviation for the Posterior Distribution of λ and γ

	Mean	Standard Deviation
Different Work Status	0.0003	0.0121
Different Gender	−0.0044	0.0158
Head of the Household/Others	0.002	0.0108
Spouse of the Head/Others	0.0031	0.0138
Different Caste	−0.0111	0.0796
Native and Non-Native	−0.0024	0.0080
Externality	0.0063	0.0075

Note: the table presents the mean and standard deviation for the posterior distribution of λ and γ using the Bayesian algorithm described in section 3. The posterior distribution takes into account the fact that there may exists degree heterogeneity in the network formation rule.

a subsample of 200 individuals, we do not find evidence of positive externalities.

5 Conclusion

This paper provides a novel approach to identify and perform inference on the utility function parameters of a network formation model with payoff externalities. Our identification results rely on the assumption that there is a predetermined probability distribution over the set of all possible equilibrium networks. Under this assumption, we can characterize the selection probabilities relying on the fact that the space of idiosyncratic payoff shocks and the type of payoff externalities determine entirely the regions where multiple equilibria happen–we call them *multiplicity regions*. With the characterization of the selection probability, we show that the payoff parameters of interest are identified from the likelihood function that aggregates the likelihoods of observing a network across different multiplicity regions.

We propose a Bayesian algorithm to estimate the parameters of interest. The use of Bayesian methods allows us to sidestep the issue of high-dimensional numerical integration. It also enables us to address the potential problem of statistical dependence in large networks because it does not rely on asymptotic theory to conduct inference. Because we recover the posterior distribution of the parameters of interest, we can perform statistical inference based on that distribution. One potential issue when translating the identification results to the estimation method is that it is unfeasible to construct a probability distribution over a sequence of relatively large networks in practice. We address this issue by building a composite likelihood based on the marginal distribution of all the possible subgraphs forming the network of interest.

Using the proposed estimation methods, we present an empirical application to model the network formation process of individuals creating social connections in villages in Karnataka, India (Banerjee et al., 2013). We characterize the probability distribution of tetrads subgraphs for the social networks and use our Bayesian algorithm to estimate the homophily effects. We find strong evidence of homophily for most of the characteristics

in our analysis. The most substantial homophily effects happen among the same gender and working status of villagers.

References

Abramowitz, M. and Stegun, I. A. (1965). Handbook of mathematical functions with formulas, graphs, and mathematical tables. *Applied Mathematics Series*.

Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, 88(422):669–679.

Bajari, P., Hahn, J., Hong, H., and Ridder, G. (2011). A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models. *International Economic Review*, 52(3):807–824.

Bajari, P., Hong, H., and Ryan, S. P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568.

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

Bayer, P., Hjalmarsson, R., and Pozen, D. (2009). Building criminal capital behind bars: Peer effects in juvenile corrections. *The Quarterly Journal of Economics*, 124(1):105–147.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies*, 47(1):225–238.

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics*, 18(1):5–46.

Chib, S. (2008). Panel Data Modeling and Inference: A Bayesian Primer. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, volume 46 of *Advanced Studies in Theoretical and Applied Econometrics*, pages 479–515. Springer.

Christakis, N., Fowler, J., Imbens, G. W., and Kalyanaraman, K. (2020). An empirical model for strategic network formation. In *The Econometric Analysis of Network Data*, pages 123–148. Elsevier.

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.

De Paula, A. (2013). Econometric analysis of games with multiple equilibria. *Annu. Rev. Econ.*, 5(1):107–131.

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

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics*, 101(5):763–776.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312.

Gaviria, A. and Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83(2):257–268.

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

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

Graham, B. S. (2020). Chapter 2 - network data. In Durlauf, S. N., Hansen, L. P., Heckman, J. J., and Matzkin, R. L., editors, *Handbook of Econometrics, Volume 7A*, volume 7 of *Handbook of Econometrics*, pages 111–218. Elsevier.

Hajivassiliou, V. A. and McFadden, D. L. (1998). The method of simulated scores for the estimation of ldv models. *Econometrica*, pages 863–896.

Hörmann, W. and Leydold, J. (2014). Generating generalized inverse gaussian random variates. *Statistics and Computing*, 24(4):547–557.

Hsieh, C.-S., König, M. D., and Liu, X. (2022). A structural model for the coevolution of networks and behavior. *Review of Economics and Statistics*, 104(2):355–367.

Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1857–97.

Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915.

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 & Economic Statistics*, 36(4):705–713.

Johnsson, I. and Moon, H. R. (2021). Estimation of peer effects in endogenous social networks: Control function approach. *Review of Economics and Statistics*, 103(2):328–345.

Kinderman, A. J. and Monahan, J. F. (1977). Computer generation of random variables using the ratio of uniform deviates. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):257–260.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

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

Leydold, J. (2001). A simple universal generator for continuous and discrete univariate t-concave distributions. *ACM Transactions on Mathematical Software (TOMS)*, 27(1):66–82.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–45.

Masten, M. A. (2014). Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies*, 85:1193–1250.

McCulloch, R. E., Polson, N. G., and Rossi, P. E. (2000). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99(1):173–193.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, pages 995–1026.

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

Milgrom, P. and Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society*, pages 1255–1277.

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

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Pelican, A. and Graham, B. S. (2020). An optimal test for strategic interaction in social and economic network formation between heterogeneous agents. Technical report, National Bureau of Economic Research.

Robert, C. (1991). Generalized inverse normal distributions. *Statistics & Probability Letters*, 11(1):37–41.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics*, 116(2):681–704.

Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):647–676.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, 88(5):1829–1858.

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

Topkis, D. M. (1979). Equilibrium points in nonzero-sum n-person submodular games. *Siam Journal on control and optimization*, 17(6):773–787.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society*, pages 1–25.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, second edition edition.

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association*, 114(526):857–868.

Appendix A Proofs of Main Results

Proof of Proposition 1. Recall that we assume $\gamma > 0$ for the transformation in (6) to be defined. To show (i), simply note that $s_ij(\mathbf{d}) \geq s_0$ for all ij and $\mathbf{d}$, so that if $t_{ij} \leq s_0$, then $t_{ij} \leq s_ij(\mathbf{d})$ or $\nu_{ij}(\mathbf{d}, u_{ij}) \geq 0$. This

guarantees $d_{ij} = 1$. For (ii), we use the upper bound $s_i j(\mathbf{d}) \leq s_L$ for all ij and $\mathbf{d}$, such that $t_{ij} > s_L \geq s_i j(\mathbf{d})$ or $\nu_{ij}(\mathbf{d}, u_{ij}) < 0$, from which $d_{ij} = 0$. Finally, for (iii), consider both cases when $s_l < t_{ij} \leq s_{l+1}$. If $s_{ij}(\mathbf{d}) \leq s_l$, then $s_{ij}(\mathbf{d}) < t_{ij}$ or $\nu_{ij}(\mathbf{d}, u_{ij}) < 0$. If $s_{ij}(\mathbf{d}) > s_l$, then $s_{ij}(\mathbf{d}) \geq s_{l+1}$ by discreteness of the support of the externality function. Thus, $t_{ij} \leq s_{l+1} \leq s_{ij}(\mathbf{d})$ or $\nu_{ij}(\mathbf{d}, u_{ij}) \geq 0$. □

Proof of Proposition 2. We prove this proposition by contradiction. Assume there exists a bucket b and two multiplicity regions m_1 and m_2, such that $b \cap m_1 = b_1$ and $b \cap m_2 = b_2$. By the definition of multiplicity region, there must exist some network equilibrium d such that $d \in m_1$ and $d \notin m_2$. Thus, we have $d \in b_1$ and $d \notin b_2$, which is a contradiction. □

Proof of Theorem 1. As a reminder, we note that we define $\bar{h}_j = h(\mathbf{d}_j)/h(\mathbf{d}_1)$ be the relative probability of $\mathbf{d}_j$ with respect to $\mathbf{d}_1$, where $h(\mathbf{d}_1)$ was chosen as the normalizing probability without loss of generality. Similarly, $G_{mj} = G(\mathbf{d}_j, m; \theta)$. We also defined $H_m = H_m$ Defining $J = |\mathbb{D}_N|$, the likelihood function of the problem in equation (10) can be written as a system of equation as follows

$$
\begin{aligned}
P(\mathbf{d}_1; h, \theta) &= \sum_{m \in M} \frac{1}{H_m} G_{m1}, \\
&\vdots \\
P(\mathbf{d}_j; h, \theta) &= \sum_{m \in M} \frac{\bar{h}(\mathbf{d}_j)}{H_m} G_{mj}, \\
&\vdots \\
P(\mathbf{d}_J; h, \theta) &= \sum_{m \in M} \frac{\bar{h}(\mathbf{d}_J)}{H_m} G_{mJ}.
\end{aligned}
$$

Without loss of generality, we assume there is one multiplicity region for the rest of the proof. The final result will simply be a sum over all the matrices for each multiplicity region. We take the total differentiation of both sides of the system of equations defined before to get

$$\Delta P(\mathbf{d}_1; h, \theta) = \frac{1}{H_m}\left[\sum_{ij}\frac{\partial G(\mathbf{d}_1)}{\partial W_{ij}\lambda}W_{ij}\nabla\lambda + \frac{\partial G(\mathbf{d}_1)}{\partial\gamma}\Delta\gamma\right]$$
$$-\sum_{l=1}^{J}\frac{G(\mathbf{d}_1;\theta)}{H_m^2}\Delta\bar{h}(\mathbf{d}_{N,l})$$
$$\vdots$$
$$\Delta P(\mathbf{d}_j; h, \theta) = \frac{\bar{h}(\mathbf{d}_j)}{\sum_{\ell\in m}\bar{h}(\mathbf{d}_j)}\left[\sum_{ij}\frac{\partial G(\mathbf{d}_j)}{\partial W_{ij}\lambda}W_{ij}\nabla\lambda + \frac{\partial G(\mathbf{d}_j)}{\partial\gamma}\Delta\gamma\right] + \frac{G(\mathbf{d}_j;\theta)}{H_m}\Delta\bar{h}(\mathbf{d}_j)$$
$$-\sum_{l=1}^{J}\frac{\bar{h}(\mathbf{d}_j)G(\mathbf{d}_j;\theta)}{H_m^2}\Delta\bar{h}(\mathbf{d}_{N,l})$$
$$\vdots$$
$$\Delta P(\mathbf{d}_J; h, \theta) = \frac{\bar{h}(\mathbf{d}_J)}{\sum_{\ell\in m}\bar{h}(\mathbf{d}_j)}\left[\sum_{ij}\frac{\partial G(\mathbf{d}_J)}{\partial W_{ij}\lambda}W_{ij}\nabla\lambda + \frac{\partial G(\mathbf{d}_J)}{\partial\gamma}\Delta\gamma\right] + \frac{G(\mathbf{d}_J;\theta)}{\sum_{\ell\in m}\bar{h}(\mathbf{d}_{i,N})}\Delta\bar{h}(\mathbf{d}_J)$$
$$-\sum_{l=1}^{J}\frac{\bar{h}(\mathbf{d}_J)G(\mathbf{d}_J;\theta)}{(\sum_{\ell\in m}\bar{f}_D(\mathbf{d}_j))^2}\Delta\bar{h}(\mathbf{d}_{N,l})$$

The identification idea is then to show that when the total derivative of the likelihood function equals zero, the vector of parameters θ and the selection probabilities $h_D(\mathbf{d})$ do not vary. Let $\Delta P(\mathbf{d}_j; f_D, \theta) = 0$ for all $j = 1, ..., J$. Defining the following matrix,

$$A = \begin{bmatrix} \frac{\partial G(\mathbf{d}_1)}{\partial\gamma} & -\frac{G(\mathbf{d}_1;\theta)}{H_m} & \cdots & -\frac{G(\mathbf{d}_1;\theta)}{H_m} & \cdots & -\frac{G(\mathbf{d}_1;\theta)}{H_m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \bar{h}(\mathbf{d}_j)\frac{\partial G(\mathbf{d}_j)}{\partial\gamma} & -\frac{\bar{h}(\mathbf{d}_j)G(\mathbf{d}_j;\theta)}{H_m} & \cdots & \frac{(\sum_{i\neq j}\bar{h}(\mathbf{d}_j))G(\mathbf{d}_j;\theta)}{H_m} & \cdots & -\frac{\bar{h}(\mathbf{d}_j)G(\mathbf{d}_j;\theta)}{H_m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \bar{h}(\mathbf{d}_J)\frac{\partial G(\mathbf{d}_J)}{\partial\gamma} & -\frac{\bar{h}(\mathbf{d}_J)G(\mathbf{d}_J;\theta)}{\sum_{\ell\in m}\bar{h}(\mathbf{d}_j)} & \cdots & -\frac{\bar{h}(\mathbf{d}_J)G(\mathbf{d}_J;\theta)}{H_m} & \cdots & \frac{(\sum_{\ell\neq J}\bar{h}(\mathbf{d}_j))G(\mathbf{d}_J;\theta)}{H_m} \end{bmatrix},$$

the system of equations defined before can be represented in matrix form as

$$\mathbf{A}\begin{bmatrix}\Delta\gamma \\ \Delta\bar{h}(\mathbf{d}_2) \\ \vdots \\ \Delta\bar{h}(\mathbf{d}_J)\end{bmatrix} = \begin{bmatrix}-\sum_{ij}\frac{\partial G(\mathbf{d}_1)}{\partial W_{ij}\lambda}W_{ij}\nabla\lambda \\ -\bar{h}(\mathbf{d}_2)\sum_{ij}\frac{\partial G(\mathbf{d}_2)}{\partial W_{ij}\lambda}W_{ij}\nabla\lambda \\ \vdots \\ -\bar{h}(\mathbf{d}_J)\sum_{ij}\frac{\partial G(\mathbf{d}_J)}{\partial W_{ij}\lambda}W_{ij}\nabla\lambda\end{bmatrix}, \tag{A-21}$$

Under the full rank assumption the matrix $\mathbf{A}$ is invertible. Given that W_{ij} are random variables, we cannot guarantee that the equality in A-21 will hold unless $\nabla\lambda = \mathbf{0}$. Therefore, we have that $\Delta P(\mathbf{d}_j; f_D, \theta) = 0$ for all

$j = 1, ..., J$ if and only if $\nabla\lambda = \mathbf{0}$, and

$$\begin{bmatrix} \Delta\gamma \\ \Delta\bar{h}(\mathbf{d}_2) \\ \vdots \\ \Delta\bar{h}(\mathbf{d}_J) \end{bmatrix} = \mathbf{0},$$

where $\mathbf{0}$ is a vector of zeros. Therefore, the vector of parameters θ and $h_D(\mathbf{d}_N)$ for all $\mathbf{d} \in \mathbb{D}_N$ are locally identified from the likelihood $P(\mathbf{d}_N; h, \theta)$ at (h_{D0}, θ_0). □

Proof of Proposition 3. We are required to show that $-g(z;\nu,\gamma)^{-1/\nu}$ is a concave function, so we proceed by showing that $-\frac{\partial^2 g(z;\nu,\gamma)^{-1/\nu}}{\partial z^2} < 0$ for all $z \in \mathbb{R}$, $\gamma \in \mathbb{R}$, and $\nu > 1$. To see this, we first need to obtain the derivatives of the kernel. As it is much easier to work with the log-kernel, we get

$$\begin{aligned} \log g(z;\nu,\gamma) &= -\nu \log(|z|) - \frac{1}{2}\left(\frac{1}{z} - \gamma\right)^2, \\ \frac{\partial}{\partial z}\log g(z;\nu,\gamma) &= -\frac{1}{z^3}(\nu z^2 + \gamma z - 1), \\ \frac{\partial^2}{\partial z^2}\log g(z;\nu,\gamma) &= \frac{1}{z^4}(\nu z^2 + 2\gamma z - 3). \end{aligned}$$

Using the property $\frac{d\log f(x)}{dz} = \frac{1}{f(x)}\frac{df(x)}{dz}$, we can see that

$$\begin{aligned} \operatorname{sgn}(c)\frac{\partial^2 g(z;\nu,\gamma)^c}{\partial z^2} &= \operatorname{sgn}(c)c\left[(c-1)g(z;\nu,\gamma)^{c-2}\left(\frac{\partial g(z;\nu,\gamma)}{\partial z}\right)^2 + g(z;\nu,\gamma)^{c-1}\frac{\partial^2 g(z;\nu,\gamma)}{\partial z^2}\right], \\ &= |c| g(z;\nu,\gamma)^c\left[c\left(\frac{\partial \log g(z;\nu,\gamma)}{\partial z}\right)^2 + \frac{\partial^2 \log g(z;\nu,\gamma)}{\partial z^2}\right]. \end{aligned}$$

Replacing the values for the log-derivatives, we finally obtain

$$\operatorname{sgn}(c)\frac{\partial^2 g(z;\nu,\gamma)^c}{\partial z^2} = |c||z|^{-(\nu+6)}\exp\left\{-\frac{1}{2}\left(\frac{1}{z} - \gamma\right)^2\right\} \cdot p_c(z;\nu,\gamma), \tag{A-22}$$

where $p_c(z;\nu,\gamma)$ is a quartic polynomial in z, given by

$$p_c(z;\nu,\gamma) = \nu(c\nu+1)z^4 + 2(c\nu+1)\gamma z^3 + [c(\gamma^2 - 2\nu) - 3]z^2 - 2c\gamma z + c.$$

Note that all terms in (A-22) aside from the polynomial one are non-negative for all values of the variable and parameters. Thus, the sign of this polynomial term controls the concavity or convexity of the function.

At the value $c = -1/\nu$, both the quartic and cubic terms drop out of $p_c(z;\nu,\gamma)$, such that $p_{-1/\nu}(z;\nu,\gamma) = (-1/\nu)[(\gamma^2+\nu)z^2 - 2\gamma z + 1]$. As this is simply a parabola, we know that it is concave since the coefficient

of the squared term is negative. It also achieves its maximum at $z^* = \gamma/(\gamma^2 + \nu)$ and at this maximum, $p_{-1/\nu}(z^*; \nu, \gamma) = -1/(\gamma^2 + \nu) < 0$ for all $\gamma \in \mathbb{R}$ and $\nu > 1$. Thus, $p_{-1/\nu}(z; \nu, \gamma) < 0$ for all $z \in \mathbb{R}$, $\gamma \in \mathbb{R}$, and $\nu > 1$. This implies our desired result. □

Proof of Theorem 2. Lemma 3 from Hörmann and Leydold (2014) states that if $f(x)$ is T_c-concave, it is also $T_{c'}$-concave for any $c' \leq c$. When $\nu > 2$, $-1/2 < -1/\nu$, such that $g(z; \nu, \gamma)$ is also $T_{-1/2}$-concave. □

Proof of Proposition 4. Let (θ_0, h_0) be the true vector of parameters generating the data. We want to characterize the Kullback and Leibler (1951) distance between the true likelihood $P(\mathbf{d}; \theta_0, h_0)$ and the composite likelihood $\tilde{P}(\mathbf{d}; \theta, \tilde{h}_0)$ caused by the misspecification embedded in the later. Let $\tilde{h}_0$ be the selection probability of a subgraph induced by the selection probability of the complete network d, which we call h_0. Define the function $Q : \mathbb{D}_N \to [0, 1]$, mapping networks $d \in \mathbb{D}_N$ from the space of possible network configurations $\mathbb{D}_N$ to the zero-one interval. It effectively computes the proportion of subnetworks of size K for each network $\mathbf{d} \in \mathbb{D}_N$. Let N_K be the total number of network configurations of size K. The vector of proportions is such that $\sum_k q_k(d_k^{(K)}) = 1$. With function Q, we can express $\tilde{h}_0(d_k^{(K)}) = q_k(d_k^{(K)})h_0(d)$. With this characterization of the equilibrium, we can characterize the approximation error due to the partition of d into subgraph, $d_k^{(K)}$. Assume there is one multiplicity region without loss of generality. The true likelihood can be writen as

$$
\begin{aligned}
P(\mathbf{d}; \theta, h_0) &= \int_{\mathbf{u}^*} \frac{h_0(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}} h_0(\mathbf{d}')} g(\mathbf{d}, \mathbf{u}^*; \gamma) f(\mathbf{u}^*) \mathrm{d}\mathbf{u}^* \\
&= \frac{h_0(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}} h_0(\mathbf{d}')} \int_{\mathbf{u}^*} g(\mathbf{d}, \mathbf{u}^*; \gamma) f(\mathbf{u}^*) \mathrm{d}\mathbf{u}^*
\end{aligned}
$$

Similarly, we compute the composite likelihood as

$$
\begin{aligned}
\tilde{P}(\mathbf{d}; \theta, h_0) &= \prod_{k=1}^{N_K} \tilde{P}(\mathbf{d}_k^{(K)}; \theta, h_0)^{\omega_k} \\
&= \prod_{k=1}^{N_K} \left[\int_{\mathbf{u}_k^{*(K)}} \frac{\tilde{h}_0(\mathbf{d}_k^{(K)})}{\sum_{\mathbf{d}' \in \mathbb{D}} \tilde{h}_0(\mathbf{d}')} g(\mathbf{d}_k^{(K)}, \mathbf{u}_k^{*(K)}; \gamma) f(\mathbf{u}_k^{*(K)}) d\mathbf{u}_k^{*(K)} \right]^{\omega_k} \\
&= \left(\frac{h_0(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}} h_0(\mathbf{d}')} \right) \prod_{k=1}^{N_K} \left[q_k(\mathbf{d}_k^{(K)}) \int_{\mathbf{u}_k^{*(K)}} g(\mathbf{d}_k^{(K)}, \mathbf{u}_k^{*(K)}; \gamma) f(\mathbf{u}_k^{*(K)}) d\mathbf{u}_k^{*(K)} \right]^{\omega_k} \\
&= \left(\frac{h_0(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}} h_0(\mathbf{d}')} \right) \left(\prod_{k=1}^{N_K} q_k(\mathbf{d}_k^{(K)}) \right)^{\omega_k} \\
&\quad \left(\int_{\mathbf{u}_k^{*(K)}} \prod_{ij} \left[\mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) \geq -u_{ij}^* \right) \right]^{d_{ij}} \times \left[1 - \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) \geq -u_{ij}^* \right) \right]^{1 - d_{ij}} f(\mathbf{u}_k^{*(K)}) d\mathbf{u}_k^{*(K)} \right)^{\omega_k}
\end{aligned}
$$

Following Kullback and Leibler (1951), we define the KB distance as the log difference between the true and the composite (misspecified) likelihoods. The distance is given by $\mathcal{KL}(\mathbf{d}, \mathbf{u}^*; \theta_0, h_0) = \log P(\mathbf{d}, \mathbf{u}^*; \theta_0, h_0) - \log \tilde{P}(\mathbf{d}, \mathbf{u}^*; \theta_0, h_0)$. From the previous set of equations, it follows that $h_0(\mathbf{d})$ the selection probabilities are not affected by the misspecified composite likelihood.

$$\mathcal{KL}(\mathbf{d}, \mathbf{u}^*; \theta_0, h_0) = -\left[\sum_k^{N_K} \omega_k \log q_k(\mathbf{d}_k^{(K)}) + \sum_{k=1}^{N_K} \omega_k \sum_{ij} \log\left(\frac{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) \geq -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) \geq -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}\right)^{d_{ij}}\right.$$
$$\left. \times \left(\frac{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) < -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) < -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}\right)^{1-d_{ij}}\right].$$

As discussed by White (1982), to maximize the composite likelihood is equivalent to minimize of the log difference defined above. The first component of the difference is independent of the value of (θ_0, h_0), which shows that the selection probabilities are not affected by the misspecification error. Intuitively, the utility function parameters are affected by the misspecification error because We underestimate $s_{ij}(d)$ by truncating it with $s_{ij}(d_k^{(K)})$. To compensate the reduction of variation of s_ij across links, the variation of estimated λ and γ relative to their true values have to increase. The direction of bias would depend on whether how connected the network is or the average value of d_{ij}. Finally, we characterize the the marginal impact of changing λ and γ on the KL distance.

$$\frac{\partial \mathcal{KL}}{\partial \lambda} = -\sum_{k=1}^{N_K} \omega_k \sum_{ij} \left(\frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}))}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) \geq -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}} - \frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}_k^{(K)})}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) \geq -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}\right)^{d_{ij}} W_{ij} +$$
$$\left(\frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}))}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) < -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}} - \frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}_k^{(K)})}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) < -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}\right)^{1-d_{ij}} W_{ij}$$

$$\frac{\partial \mathcal{KL}}{\partial \gamma} = -\sum_{k=1}^{N_K} \omega_k \sum_{ij} \left(\frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}))s_{ij}(\mathbf{d})}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) \geq -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}} - \frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}_k^{(K)})s_{ij}(\mathbf{d}_{k,k})}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) \geq -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}\right)^{d_{ij}} +$$
$$\left(\frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}))s_{ij}(\mathbf{d})}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) < -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}} - \frac{f(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}_k^{(K)})s_{ij}(\mathbf{d}_k^{(K)})}{\int_{u_{ij}^*} \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}_k^{(K)}) < -u_{ij}^*\right) f(u_{ij}) \mathrm{d}u_{ij}}\right)^{1-d_{ij}}$$

□

Appendix B Equilibrium in Directed Networks

In this section we show that our identification results are consistent with a network formation model where individuals can choose directed links. We follow the model by Pelican and Graham (2020) where individuals form directed links based on a payoff function that includes preferences over other individuals' positions in the network. As in the main text, individual's utility from the network configuration is determined by

$$\nu_i\left(\mathbf{d}_i, \mathbf{d}_{-i}; \mathbf{U}\right) = \sum_j d_{ij}\left[A_i + B_j + X_i'\Lambda_0 X_i + \gamma_0 s_{ij}(\mathbf{d}) - U_{ij}\right], \tag{B-23}$$

where the components of the utility are described in the main text (equation (1)). We assume a game of complete information where individual i observes $\{A_i, B_i, X_i'\}_{i=1}^N$ and $\{U_{ij}\}_{i\neq j}$, then decides what $N-1$ agents she will send links to. Following the literature of directed network formation, we use the Nash Equilibrium (NE) solution concept. A pure strategy NE corresponds to a strategy $\mathbf{d}^*$ such that $\nu_i\left(\mathbf{d}_i^*, \mathbf{d}_{-i}^*, \mathbf{u}\right) \geq \nu_i\left(\mathbf{d}_i, \mathbf{d}_{-i}^*, \mathbf{u}\right)$.

Based on the marginal utility defined in equation (2) in the main text, an observed Nash Equilibrium network is one that satisfies the following $N(N-1)$ set of non-linear equations:

$$d_{ij} = \mathbb{I}\left(A_i + B_j + W_{ij}'\lambda + \gamma_0 s_{ij}(\mathbf{D}) \geq u_{ij}\right),$$

for $i = 1, \ldots, N$ and $i \neq j$. Following Miyauchi (2016) and Pelican and Graham (2020), we define the mapping $\varphi(\mathbf{d}) : \mathbb{D}_N \to \mathbb{I}_{N(N-1)}$ as

$$\varphi(\mathbf{d}) \equiv \begin{bmatrix} \mathbb{I}\left(\Delta\nu_{12}\left(\mathbf{d}_i, \mathbf{d}_{-i}; u_{ij}\right) \geq 0\right) \\ \mathbb{I}\left(\Delta\nu_{13}\left(\mathbf{d}_i, \mathbf{d}_{-i}; u_{ij}\right) \geq 0\right) \\ \vdots \\ \mathbb{I}\left(\Delta\nu_{NN-1}\left(\mathbf{d}_i, \mathbf{d}_{-i}; u_{ij}\right) \geq 0\right) \end{bmatrix},$$

and use Tarski's (1955) fixed point theorem to argue that for $\gamma_0 \geq 0$, an equilibrium exists and the set of all equilibria constitutes a non-empty complete lattice (see Proposition 1 in Miyauchi, 2016). An important insight is that the change in equilibrium concept does not change the idea of multiplicity regions in Definition 2. The reason is that multiplicity regions definition takes any arbitrary equilibrium concept as its input, and serves as a way to partition the space of shocks $\mathbf{u}$. Therefore, the likelihood of observing a network $\mathbf{d}_N$ under the Nash Equilibrium concept, can also be written as in equation (9) in the main text

$$P(\boldsymbol{d}_N; \theta) = \sum_{m \in M} \frac{h_D(\boldsymbol{d}_N)}{\sum_{i \in m} h_D(\boldsymbol{d}_{N,\ell})} \int_{\boldsymbol{u}_N \in \mathbb{R}^N} \mathbb{I}\{\boldsymbol{u}_N \in m\} g(\boldsymbol{d}_N, \boldsymbol{u}_N; \theta) \prod_{i \neq j} f_U\left(u_{N,ij}\right) \mathrm{d}\boldsymbol{u}_N,$$

where in this case the function $g(\boldsymbol{d}_N, \boldsymbol{u}_N; \theta)$ takes the form

$$g(\boldsymbol{d}_N, \boldsymbol{u}_N; \theta) = \prod_i \prod_j \mathbb{I}\left(A_i + B_j + W'_{ij}\lambda \geq u_{ij}\right)^{d_{ij}} \times \mathbb{I}\left(A_i + B_j + W'_{ij}\lambda < u_{ij}\right)^{1-d_{ij}}$$

to be consistent with the Nash Equilibrium solution concept. Given that the likelihood function can be still separated into a weighted sum of multiplicity regions probability masses, the identification result in 1 also holds for the directed network formation game.

Appendix C Computational Process

Our approach to calculate the parameters of interest relies on the idea of subgraphs. In particular, we partition our observed network $\mathbf{d}$ into all possible tetrads (subgraphs of size 4). After the partition, the tetrads $\mathbf{d}_4$ become our unit of analysis to evaluate the likelihood. Therefore, we focus on the relative probabilities $\sum_{m \in M} h_D(\boldsymbol{d}_4) / \sum_{i \in m} h_D(\boldsymbol{d}_{4,\ell})$, for the multiplicity regions $m \in M$. From Proposition 2, all multiplicity regions are composed by buckets. Thus, summing across all multiplicity regions is equivalent to summing across all buckets. Based on the above, we need to determine all possible buckets for each type of tetrad network configuration $\boldsymbol{d}_{4,k}$.

Given that we are focusing only on undirected networks, the total number of tetrad configurations can be represented by the total number of configurations of the upper (or lower) triangular part of the adjacency matrix representing $\boldsymbol{d}_{4,k}$. In the tetrads case, there are $2^6 = 64$ possible configurations of triangular matrices. To construct the buckets, we need to consider the type of externality. If we focus on intransitive triads externalities, each dyad could have either 0, 1 or 2 intransitive triads. Based on those numbers, we can partition the space of errors into 4 different spaces:

$$\underbrace{(-\infty, 0]}_{I_1} \cup \underbrace{(0, \gamma]}_{I_2} \cup \underbrace{(\gamma, 2\gamma]}_{I_4} \cup \underbrace{(2\gamma, \infty)}_{I_4}.$$

Given that there are 6 possible dyads in a directed tetrad, we need to consider all the possible combinations of regions where the shocks can fall when constructing the possible buckets. Therefore, there are $4^6 = 4096$ possible buckets when focusing on tetrads and intransitive triads.

Appendix D Generalized Inverse Normal generation

The generalized inverse normal distribution was introduced in Robert (1991). This appendix presents details on deriving an efficient algorithm for sampling random variates from this distribution. Let $Z \sim \mathcal{GIN}(\nu, \gamma, \tau)$. Then

the density of Z, denoted as $f_Z(z)$, is given by

$$f_Z(z) = \frac{1}{C(\nu,\gamma,\tau)}|z|^{-\nu}\exp\left\{-\frac{1}{2\tau^2}\left(\frac{1}{z}-\gamma\right)^2\right\}, \tag{D-24}$$

where the proportionality constant can be expressed as

$$\begin{aligned} C(\nu,\gamma,\tau) &= \int_{-\infty}^{\infty}|z|^{-\nu}\exp\left\{-\frac{1}{2\tau^2}\left(\frac{1}{z}-\gamma\right)^2\right\}dz \\ &= \tau^{\nu-1}\exp\left\{-\frac{\gamma^2}{2\tau^2}\right\}2^{\frac{\nu-1}{2}}\Gamma\left(\frac{\nu-1}{2}\right){}_1F_1\left(\frac{\nu-1}{2};\frac{1}{2};\frac{\gamma^2}{\tau^2}\right), \end{aligned}$$

and ${}_1F_1$ represents the confluent hypergeometric function (Abramowitz and Stegun, 1965). As shown in Robert (1991), when $Z \sim \mathcal{GIN}(\nu,\gamma/\tau,1)$ and $Y = Z/\tau$, then $f_Y(y) = f_Z(\tau y)\tau$ so that $Y \sim \mathcal{GIN}(\nu,\gamma,\tau)$. Thus, we only need to devise an algorithm for sampling from the $\mathcal{GIN}(\nu,\gamma,1)$ distribution. We denote the kernel for the standardized version of this distribution as

$$g(z;\nu,\gamma) = |z|^{-\nu}\exp\left\{-\frac{1}{2}\left(\frac{1}{z}-\gamma\right)^2\right\} \tag{D-25}$$

The generalized inverse normal distribution $\mathcal{GIN}(\nu,\gamma,1)$ is bimodal with modes at

$$m^- = \frac{-\gamma-\sqrt{\gamma^2+4\nu}}{2\nu} \quad \text{and} \quad m^+ = \frac{-\gamma+\sqrt{\gamma^2+4\nu}}{2\nu}, \tag{D-26}$$

with $m^- < 0 < m^+$. This means it is easier to sample from the positive and negative regions separately using the ratio-of-uniforms with mode shift. To this end, we study the generalized inverse normal distribution truncated to the negative or positive reals, respectively denoted as $\mathcal{GIN}^-(\nu,\gamma,\tau)$ and $\mathcal{GIN}^+(\nu,\gamma,\tau)$. That is, the densities for the $\mathcal{GIN}^-(\nu,\gamma,\tau)$ and $\mathcal{GIN}^+(\nu,\gamma,\tau)$ distributions, f_Z^- and f_Z^+ respectively, are given by

$$f_Z^-(z) = \frac{f_Z(z)\mathbb{I}(z<0)}{F_Z(0)}, \tag{D-27}$$

$$f_Z^+(z) = \frac{f_Z(z)\mathbb{I}(z>0)}{1-F_Z(0)}, \tag{D-28}$$

where $F_Z(0) = \int_{-\infty}^0 f_Z(z)dz$ is the cumulative distribution function (CDF) of the generalized inverse normal evaluated at 0. By the same argument to before, if we let $Z \sim \mathcal{GIN}(\nu,\gamma/\tau,1)$ and $Y = Z/\tau$, then

$$F_Y(0) = \int_{-\infty}^0 f_Y(y)dy = \int_{-\infty}^0 f_Z(z)dz = F_Z(0),$$

so that we can focus on the case $\tau = 1$. Define $C^-(\nu,\gamma) \equiv \int_{-\infty}^{0} g(z;\nu,\gamma)dz$ and $C^+(\nu,\gamma) \equiv \int_{0}^{\infty} g(z;\nu,\gamma)dz$. Then, $C^-(\nu,\gamma) + C^+(\nu,\gamma) = C(\nu,\gamma,1)$, and we can simplify

$$f_Z^-(z) = \frac{g(z;\nu,\gamma)\mathbb{I}(z<0)}{C^-(\nu,\gamma)}\,,$$
$$f_Z^+(z) = \frac{g(z;\nu,\gamma)\mathbb{I}(z>0)}{C^+(\nu,\gamma)}\,.$$

Using the intermediate results in Robert (1991), we can find closed-form expressions for these proportionality constants as

$$C^-(\nu,\gamma) = e^{-\frac{\gamma^2}{4}}\Gamma(\nu-1)D_{-(\nu-1)}(-\gamma)\,,$$
$$C^+(\nu,\gamma) = e^{-\frac{\gamma^2}{4}}\Gamma(\nu-1)D_{-(\nu-1)}(\gamma)\,,$$

and D is the parabolic cylinder function (Abramowitz and Stegun, 1965). As expected, the kernels of these truncated distributions are the same as the original kernel $g(z;\nu,\gamma)$ with the additional indicators for the respective truncated region. In the remainder of the section we describe the procedure of drawing from the positive region (i.e., drawing from f_Z^+), which is the one used in the paper. Full details of both procedures are provided in Algorithms 1 and 2.

The ratio-of-uniforms method is usually implemented by drawing a point uniformly from the minimal bounding rectangle $\mathcal{R} = \{(u,v) : u_{\min} \le u \le u_{\max}, 0 \le v \le v_{\max}\}$, where

$$v_{\max} = \sup_{z>0}\sqrt{g(z;\nu,\gamma)} = \sqrt{g(m^+;\nu,\gamma)}\,,$$
$$u_{\min} = \inf_{0<z\le m^+}(z-m^+)\sqrt{g(z;\nu,\gamma)}\,,$$
$$u_{\max} = \sup_{z>m^+}(z-m^+)\sqrt{g(z;\nu,\gamma)}\,.$$

To compute $u_{\min}$ and $u_{\max}$ we consider two approaches available in the literature. The first, due to Leydold (2001), uses information on the proportionality constant and sets

$$u_{\max} = \frac{C^+(\nu,\gamma)}{v_{\max}}\,,$$
$$u_{\min} = -u_{\max}\,. \tag{D-29}$$

The original formulation in Leydold (2001) can also exploit information on the CDF at the mode m^+, but we were not able to obtain a closed-form expression for this quantity. Thus, we turn to our second and preferred method, which uses Cardano's formula for solving cubic equations, as done in Hörmann and Leydold (2014).

Optimizing the function $(z - m^+)\sqrt{g(z;\nu,\gamma)}$ using standard calculus yields the first-order condition

$$\left\{\sqrt{g(z;\nu,\gamma)}\left[1 - \frac{z-m^+}{2z^3}(\nu z^2 + \gamma z - 1)\right]\right\}_{z=z^*} = 0\,.$$

Simplifying this expression, we see that any solution z^* must be a root of the cubic polynomial equation

$$p(z;\nu,\gamma,m^+) \equiv (2-\nu)z^3 - (\gamma - \nu m^+)z^2 + (1 + \gamma m^+)z - m^+ = 0\,. \tag{D-30}$$

We let z_1 denote the root such that $0 < z_1 \leq m^+$ and z_2 be such that $m^+ < z_2 < \infty$.[8] We then set

$$\begin{aligned} u_{\min} &= (z_1 - m^+)\sqrt{g(z_1;\nu,\gamma)}\,, \\ u_{\max} &= (z_2 - m^+)\sqrt{g(z_2;\nu,\gamma)}\,. \end{aligned} \tag{D-31}$$

Everything remains the same for the distribution truncated to the negative region, except that we use $C^-(\nu,\gamma)$ for the Leydold (2001) method. For the method similar to Hörmann and Leydold (2014), we use m^- for the shift and choose the roots z_1 and z_2 such that $-\infty < z_1 \leq m^-$ and $m^- < z_2 < 0$.

Algorithm 1 Generate from $\mathcal{GIN}^-(\nu,\gamma,\tau)$ distribution

Input: $\nu > 2, \gamma \in \mathbb{R}, \tau > 0$
$\mu \leftarrow \gamma/\tau$
$m^- \leftarrow -(\mu + \sqrt{\mu^2 + 4\nu})/2\nu$
$v_{\max} \leftarrow \sqrt{g(m^-;\nu,\mu)}$
if using the method of Hörmann and Leydold (2014) **then**
    Obtain roots z_1 and z_2 of $p(z;\nu,\mu,m^-)$ using Cardano's formula
    $u_{\min} \leftarrow (z_1 - m^-)\sqrt{g(z_1;\nu,\mu)}$
    $u_{\max} \leftarrow (z_2 - m^-)\sqrt{g(z_2;\nu,\mu)}$
else if using the method of Leydold (2001) **then**
    $u_{\max} \leftarrow C^-(\nu,\gamma)/v_{\max}$
    $u_{\min} \leftarrow -u_{\max}$
end if
repeat
    Draw U uniformly from $(u_{\min}, u_{\max})$
    Draw V uniformly from $(0, v_{\max})$
    $Z \leftarrow U/V + m^-$
until $U^2 \leq g(Z;\nu,\mu)$
$Y \leftarrow Z/\tau$
return Y

From these two algorithms, we can devise an efficient algorithm to sample from the full generalized inverse normal distribution. Let S be distributed Bernoulli with success probability p. Define the random variable Y

[8] The third root will always be negative and yield the unconstrained minimum of $(z - m^+)\sqrt{g(z;\nu,\gamma)}$.

Algorithm 2 Generate from $\mathcal{GIN}^{+}(\nu, \gamma, \tau)$ distribution

Input: $\nu > 2, \gamma \in \mathbb{R}, \tau > 0$

$\mu \leftarrow \gamma/\tau$
$m^{+} \leftarrow (-\mu + \sqrt{\mu^2 + 4\nu})/2\nu$
$v_{\max} \leftarrow \sqrt{g(m^{+}; \nu, \mu)}$
if using the method of Hörmann and Leydold (2014) **then**
    Obtain roots z_1 and z_2 of $p(z; \nu, \mu, m^{+})$ using Cardano's formula
    $u_{\min} \leftarrow (z_1 - m^{+})\sqrt{g(z_1; \nu, \mu)}$
    $u_{\max} \leftarrow (z_2 - m^{+})\sqrt{g(z_2; \nu, \mu)}$
else if using the method of Leydold (2001) **then**
    $u_{\max} \leftarrow C^{+}(\nu, \gamma)/v_{\max}$
    $u_{\min} \leftarrow -u_{\max}$
end if
repeat
    Draw U uniformly on $(u_{\min}, u_{\max})$
    Draw V uniformly on $(0, v_{\max})$
    $Z \leftarrow U/V + m^{+}$
until $U^2 \leq g(Z; \nu, \mu)$
$Y \leftarrow Z/\tau$
return Y

such that

$$Y|S = s \sim \begin{cases} \mathcal{GIN}^{-}(\nu, \gamma, \tau) & \text{if } s = 0 \\ \mathcal{GIN}^{+}(\nu, \gamma, \tau) & \text{if } s = 1 \end{cases}$$

Without loss of generality we set $\tau = 1$, and let $p = C^{+}(\nu, \gamma)/C(\nu, \gamma, 1)$. Then, the marginal distribution of Y is $\mathcal{GIN}(\nu, \gamma, 1)$, since

$$\begin{aligned} f_Y(y) &= f_Z^{-}(y) \cdot \Pr(S = 0) + f_Z^{+}(y) \cdot \Pr(S = 1)\,, \\ &= \frac{g(y; \nu, \gamma)\mathbb{I}(y < 0)}{C^{-}(\nu, \gamma)} \cdot \left[1 - \frac{C^{+}(\nu, \gamma)}{C(\nu, \gamma, 1)}\right] + \frac{g(y; \nu, \gamma)\mathbb{I}(y > 0)}{C^{+}(\nu, \gamma)} \cdot \frac{C^{+}(\nu, \gamma)}{C(\nu, \gamma, 1)}\,, \\ &= \frac{g(y; \nu, \gamma)}{C(\nu, \gamma, 1)}\,. \end{aligned}$$

We summarize the generation procedure in Algorithm 3. We provide Python and `R` routines that implement all generating algorithms described in the paper.

Algorithm 3 Generate from $\mathcal{GIN}(\nu, \gamma, \tau)$ distribution

Input: $\nu > 2, \gamma \in \mathbb{R}, \tau > 0$

$\mu \leftarrow \gamma/\tau$
$p \leftarrow C^+(\nu, \mu)/C(\nu, \mu, 1)$
Draw U uniformly on (0, 1)
if $U > p$ **then**
    Draw Z from $\mathcal{GIN}^-(\nu, \mu, 1)$ using Algorithm 1
else
    Draw Z from $\mathcal{GIN}^+(\nu, \mu, 1)$ using Algorithm 2
end if
$Y \leftarrow Z/\tau$
return Y