Identification and Estimation of Network Statistics with Missing Link Data*

Matthew Thirkettle†

This version: November 5, 2019

Please download the latest version at matthewthirkettle.com

Abstract

I obtain informative bounds on network statistics in a partially observed network whose formation I explicitly model. Partially observed networks are commonplace due to, for example, partial sampling or incomplete responses in surveys. Network statistics (e.g., centrality measures) are not point identified when the network is partially observed. Worst-case bounds on network statistics can be obtained by letting all missing links take values zero and one. I dramatically improve on the worst-case bounds by specifying a structural model for network formation. An important feature of the model is that I allow for positive externalities in the network-formation process. The network-formation model and network statistics are set identified due to multiplicity of equilibria. I provide a computationally tractable outer approximation of the joint identified region for preferences determining network-formation processes and network statistics. In a simulation study on Katz-Bonacich centrality, I find that worst-case bounds that do not use the network formation model are 44 times wider than the bounds I obtain from my procedure.

Keywords: Social networks, strategic network formation, partially observed network, multiple equilibria, subnetworks, partial identification, moment inequalities, Katz Bonacich centrality

*I am grateful to Francesca Molinari, Levon Barseghyan, David Easley, and Jörg Stoye for their support and feedback. I thank Isha Agarwal, Abhishek Ananth, Panle Jia Barwick, Giulia Brancaccio, Anne Burton, Angela Cools, Maura Coughlin, Tommaso Denti, Thomas Eisenberg, James Elwell, Amanda Eng, Andrew J. Fieldhouse, Jorgen Harris, Katelyn Heath, Jean-Francois Houde, Malin Hu, Eleonora Patacchini, Penny Sanders, Seth Sanders, Caroline Walker, and seminar participates at Cornell University for valuable comments. Research support was provided in part by the Tapan Mitra and the L.R. "Red" Wilson endowments.

†Department of Economics, Cornell University, mkt68@cornell.edu.

1 Introduction

A wide array of economic outcomes of interest are generated through processes that involve the social interaction of individuals. In a classroom setting, for example, students' schooling effort and subsequent test scores are determined in part by their friends' effort provision through the process of knowledge spillovers. As another example, information about vaccines in developing countries and consequently vaccine uptake is spread through word of mouth. The position of an individual determines, at least in part, her economic outcomes as well as her influence on the economic outcomes of others. Centrality measures are network statistics that allow researchers to parsimoniously capture different features of an individual's network position. In practice, however, parts of the network are often unobserved due to subsampling or low response rates to surveys inquiring about social interactions (e.g., Banerjee, Chandrasekhar, Duflo, and Jackson (2013)). Consequently, centrality measures and other statistics of the network are not point identified.

I obtain informative bounds on network statistics and their impact on economic outcomes of interest in a partially observed network whose formation I explicitly model. My method to recover bounds on a network statistic applies to *social networks*, e.g., friendships between students in a classroom. I propose a model in which the network is endogenous and individuals choose their friends based on their preferences. A distinctive feature of my network-formation model is that preferences depend on the topology of the network, e.g., the popularity of others. As a result, individuals strategically choose their friends and some individuals may only form friendships with popular individuals. I next use this network-formation model to infer the missing portion of the network. The network-formation model admits multiple equilibria, resulting in partial identification of the model. I recover bounds on the missing portion of the network and hence bounds on network statistics of interest. My main theoretical result is in obtaining a joint outer region for both the preferences determining network formation processes and for network statistics. This result applies broadly to a range of network statistics including intercentrality (Ballester, Calvó-Armengol, & Zenou, 2006), diffusion centrality (Banerjee et al., 2013), and Katz-Bonacich centrality (KBC) (Katz, 1953; Bonacich, 1987).[1]

The problem of partially observed network data is common in applied research. Due to resource constraints, researchers may only elicit social interaction information from a subsample of the population of interest – whether households in remote villages (Banerjee et al., 2013) or students in a school (Add Health, Harris (2009)). Low response rates also result

[1]Many more examples are given in Bloch, Jackson, and Tebaldi (2019).

in partially observed networks. For example, the response rate in the Add Health dataset ranges from 77.5% to 88.6%. My model allows for this type of partial network information as well as any setting where network links are missing at random, but the identity of the individuals are known. I assume that the researcher partially observes a large number of undirected social networks. Each network is a cross-sectional snapshot and can be thought of as a market with a fixed population of individuals. These partially observed networks may, for example, come from a random survey.[2] I assume the characteristics that influence social connections, e.g, gender, race, ect. are observed for all individuals in the network. While I do allow for characteristics that are unobservable to the researcher, these characteristics must satisfy an independence assumption commonly maintained in the literature.

The network is endogenously formed according to a structural model, which governs how people choose their friends. For example, students in a classroom form friendships with other students based on their characteristics and the popularity of other students. This structural framework allows me to both partially reconstruct the unobserved component of the network and execute counterfactual analysis where, for instance, a set of individuals or links are removed from the network. A distinctive feature of my network-formation model relative to previous work on identification with sampled networks (Chandrasekhar & Lewis, 2016) is that it allows for positive externalities. Social actors derive positive utility from connections with popular individuals and from individuals with whom they have many mutual friends. Allowing for positive externalities is important for obtaining good model fit. Standard models without spillovers predict that triads – three people that are connected with one another – form at much lower frequencies than what is found empirically, and allowing for spillovers is critical in correcting this issue (Jackson et al., 2008; Graham, 2016). Positive externalities also imply strategic behavior; there is ample evidence to suggest that individuals act strategically when forming friendships.[3]

Identification and estimation in my framework is challenging. The network-formation model admits multiple equilibria. Consequently, the model is incomplete (Tamer, 2003) without further restrictions on the function that selects which particular equilibrium is realized in

[2] In the survey example, the researcher obtains information on all the social interactions between individuals in the subsample. In addition, the researcher can observe the connections that individuals in the subsample have with individuals not in the subsample. Connections between individuals who are not in the subsample are not observed. Consequently, the *star network* is obtained for each network (Chandrasekhar & Lewis, 2016), see Figures F1a and F1b.

[3] See, for example, Jackson and Wolinsky (1996); Bala and Goyal (2000); Echenique, Fryer Jr, and Kaufman (2006); Jackson et al. (2008); Currarini, Jackson, and Pin (2009); Leung (2015b); Mele (2017); Badev (2018); Sheng (2018); Gualdani (2019).

cases of multiplicity – generally called the *selection mechanism* in the partial identification litearture (Tamer, 2010; Molinari, 2019).[4] The *sharp identified region* contains all network-formation parameters such that the parameterized model is consistent with the observed data. I theoretically characterize the sharp identified region for the network-formation model under partially observed network data using an approach similar to that described in Gualdani (2019) and Molinari (2019). The sharp region cannot, however, be feasibly computed, because doing so would entail checking that $O(2^{2^{\frac{n(n-1)}{2}}})$ moment inequalities are satisfied, with n denoting the number of individuals in the network (in a small classroom with 5 students, that amounts to 10^{308} moment inequalities). In light of this challenge, I make progress by implementing the following steps. I leverage a useful property of the network-formation model called *strategic complementarity* for the purpose of identification.[5] Strategic complementarity implies that the marginal value of a friendship is increasing as more friendships form in the network. As a result, all equilibria belong to an easily characterizable lattice (Miyauchi, 2016), which I call the *admissible lattice*. I prove that the admissible lattice can be computed in no more than $\frac{n(n-1)}{2} - n + 1$ evaluations of a matrix function. This is important for establishing computational feasibility of the model. I also show how to use the admissible lattice for identification of the network-formation model and to obtain bounds on KBC and other centrality measures.

My main theorem establishes an outer region (a set containing the sharp identified region) using these bounds coupled with subnetwork identification (Sheng, 2018). Subnetwork identification bounds the joint probability of a subnetwork and the full characteristic vector that determines friendships by: (1) the probability that the subnetwork is the unique equilibrium; and (2) the probability that the subnetwork is in the set of equilibria. I extend this framework in two ways. First, I take expectations with respect to characteristics of individuals not contained in the subnetwork. This reduces the number of moment inequalities resulting in a feasible method and does not suffer from inference issues related to many moments. Second, I explicitly use the lattice structure of the equilibria set to enhance the feasibility of the model.

To show the applicability of my paper to a general economics context, I present results from a simulation study. My method for identifying the network-formation model is 100 to 1000 times faster than existing methods based on a comparison of run-time between my method

[4]The resulting identification problem is similar to the well known one affecting inference in entry games (see, for example, Ciliberto and Tamer (2009); Beresteanu, Molchanov, and Molinari (2011)).

[5]Strategic complementarity is also referred to as supermodularity in the mathematical economics literature (Tarski et al., 1955; Topkis, 1978).

and the results reported in Sheng (2018). These comparisons are based on subnetworks of size two, and I expect my method to offer an even larger advantage for larger subnetworks. I report results based on subnetworks of up to size five. In the simulation study, I allow for two channels of spillovers in the network-formation process: popularity spillover and mutual friend spillover. With only one spillover channel, I obtain very tight bounds on the network-formation model and on KBC. In particular, the identified region for the popularity spillover is $[0.996, 1.012]$ when the true value is equal to 1. With respect to KBC, worst-case bounds range from 2.850 to 5.532, while the true value is 4.020.[6] In contrast, I obtain bounds on effort equal to $[4.016, 4.239]$ when applying my framework with one channel of spillovers. This is a 12 fold improvement on the worst-case bounds. When I allow for two channels of spillovers, I find fairly wide bounds on the network-formation parameters. However, the bounds on KBC remain informative, ranging from 3.678 to 3.730 when the true value is 3.690.

The rest of this paper is organized as follows. Section 2 describes the related literature. Section 3 details the data requirements. Section 4 provides details on centrality measures and the model. Section 5 discusses identification. Section 6 presents the application and Section 7 concludes. All proofs are relegated to Appendix B. Appendix Table G1 summarizes all relevant notation for this paper.

2 Related Literature

I make contributions to two distinct literatures: (1) estimating network models and statistics with missing network data; and (2) identification, estimation, and computation of partially identified, structural network formation models. To do so, I leverage useful results from games with strategic complementaries.

There is a growing literature on estimating network models with missing network data, ranging from partial network data (e.g., survey data or misclassified links) to completely unobserved networks. My paper belongs to the literature on partial network data. Chandrasekhar and Lewis (2016) is the closest paper to mine. They assume the researcher has network survey data and propose a two-step method for identifying network statistics. The network is generated according to an exogenous process and is applied to a model where economic out-

[6] The worst-case bounds result from imposing no assumptions on the unobserved links. These can be thought of as Manski worst-case bounds (Manski, 1989, 2003). In the case when the marginal return to effort is positive for all individuals, these bounds are obtained by evaluating the unobserved entries in the network to 0 and 1. from which no assumptions are imposed.

comes are a linear function of a network statistic.[7] I extend their framework by endogenizing the network and allowing for strategic network formation. Liu (2013) provides conditions under which the structural parameters of the linear social interactions model are identified with a sampled network. The focus of my paper is on identifying network statistics from a partially observed network in a more general context and not on the linear social interactions model *per se*. The equilibrium level of effort in the social interactions model is equal to KBC. Therefore, my framework easily extends to obtaining bounds on peer effort subject to availability of outcome data and maintaining assumptions outlined in Liu (2013). Partial network data may also result from link misclassification where, for example, two individuals are friends and the friendship is incorrectly recorded as not existing in the data. Lewbel, Qu, and Tang (2019) allow for link misclassification and maintain linear restrictions on the structural parameters of the linear social interactions model to obtain point identification. I assume that the links are correctly reported, but I do not impose linear restrictions on the structural parameters.

There is a separate literature where the network is completely unobserved. Various assumptions have been proposed to obtain point identification for the data generating process governing network formation or the underlying structural parameters for a game played on the network. I do not maintain these assumptions. For example, de Paula, Rasul, and Souza (2018), Rose (2015), Gautier and Rose (2016), and Manresa (2016) identify and estimate the structural parameters of the social interactions model and de Paula et al. also estimate the network links when the network is completely unobserved. These papers assume that the network is sparse and require panel data consisting of multiple draws of outcomes on a fixed network. Boucher and Houndetoungan (2019) assume observability of aggregate network statistics and obtain point identification for the structural parameters of the social interactions model. Battaglini, Patacchini, and Rainone (2019) propose a model where one observes only the outcomes of legislators (legislative effectiveness) and put forward a new network competitive equilibrium concept. This can be thought of as a general market equilibrium where effectiveness is analogous to market-clearing prices. They assume that legislators optimally choose friends while taking effectiveness as given (i.e., they are price takers).

The second literature that I contribute to is on identification and estimation of network-

[7] There is a large literature using exogenous processes to impute networks with misclassified nodes (i.e., individuals) and links, e.g., see Robins, Pattison, and Woolcock (2004); Smith and Moody (2013); Huisman (2014); Krause, Huisman, Steglich, and Sniiders (2018) and references within. These methods are not suitable for social network formation as they do not allow for strategic network formation.

formation models. I assume that the network forms according to a model with complete information and strategic complementarity. There are a variety of strategies that have been proposed to identify the structural parameters for this model, see Graham (2014), Chandrasekhar (2016), de Paula (2017), de Paula (2019), or Graham (2019) for an overview. I use subnetwork identification, which was first proposed by Sheng (2018). I extend this procedure, as mentioned in the introduction, by (1) integrating out characteristics of individuals not in the subnetwork, and (2) bounding the distribution of subnetworks using the admissible lattice (a set defined below that contains all network equilibria). Other identification strategies have also been suggested. For example, Miyauchi (2016) proposes the use of monotone network statistics to partially identify the structural parameters of the model. I can augment my identification procedure with monotone network statistics. However, these statistics may require full knowledge of the network, which is not feasible in my data setting. de Paula, Richards-Shubik, and Tamer (2015) consider a single, growing network. Restrictions are imposed on the richness of unobserved heterogeneity and the number of direct friends (i.e., the network is sparse). I do not impose these restrictions and, in particular, I allow each dyadic pair to receive an i.i.d taste shock. Menzel (2015) also considers a growing network and shows that the asymptotic probability of forming a link is summarized by a conditional inclusive value.[8] Menzel proposes a maximum likelihood estimator based on the asymptotic distribution after imposing additional assumptions for point identification. Mele (2017) proposes that individuals meet sequentially at random and obtains point identification for the network-formation parameters. Specifying the meeting process completes the model and results in a unique equilibrium. Christakis, Fowler, Imbens, and Kalyanaraman (2010), Mele and Zhu (2017), Badev (2018), Boucher (2018), and Hsieh, Lee, and Boucher (2019) also assume a similar meeting processes to obtain point identification. In contrast to these papers, I do not impose restrictions on the selection mechanism and, as a result, my method allows for any meeting process.

There are alternate models of network-formation to the one that I propose that can also capture positive externalities from friendship formation. For example, dynamic models can allow friendship formation to depend on past popularity of individuals in the network (Goldsmith-Pinkham & Imbens, 2013; Graham, 2016; Lee, Fosdick, & McCormick, 2018; Bykhovskaya, 2019). These dynamic models require rich panel network data. They also do not allow for contemporaneous strategic interaction and, as a consequence, admit a unique equilibrium.

[8] The inclusive value is a sufficient statistic for the choice probabilities of the available alternatives in the standard Logit model for multinomial choice (Train, 2009).

Incomplete information games with strategic interaction have also been proposed in the literature (Leung, 2015a; Song & van der Schaar, 2015; de Martí & Zenou, 2015; Ridder & Sheng, 2017). Incomplete information assumes that individuals do not observe taste shocks before choosing friends. One criticism of an incomplete information game is that it suffers from *ex-post* regret, where individuals would like to reevaluate the links after the network has formed. I assume that the observed static network is the equilibrium of a long-run game where individuals have all relevant information available to them. Based on their preferences, individuals are unwilling to remove friends or mutually form new friendships. As a result, my model does not suffer from *ex-post* regret.

I show how to theoretically characterize the sharp identified region for the network-formation model with partially observed network data. This theoretical result builds on Molinari (2019) who shows how to obtain the sharp identified region for the network-formation model that I specify in this paper when the network is fully observed. Under a different network-formation model than the one I consider, Gualdani (2019) shows how to theoretically characterize the sharp identified region for her game. While the sharp identified region cannot be feasibly computed in my model, it is an important ingredient for characterizing an outer region that both can be feasibly computed and is informative about the underlying structural parameters and the network statistic of interest.

Using theory from games with strategic complementarities (Topkis, 1978; Tarski et al., 1955; Milgrom & Roberts, 1990), Miyauchi (2016) provides us with a useful characterization of the set of pairwise stable networks. Similar characterizations have been applied in Nash games, finite level rationalizability, and two-sided matching games (Jia, 2008; Molinari & Rosen, 2008; Uetake & Watanabe, 2012; Nishida, 2014). A key insight of my paper is to use the characterization to achieve computationally tractable bounds for the network formation model and obtain bounds on network statistics, such as KBC, intercentrality, and diffusion centrality.

3 Data

Before diving into the theoretical section of this paper, I discuss data requirements. These are weak in the sense that I only require the observation of a small number of links between a set of individuals to identify the network-formation model (i.e., a subnetwork – see Figure F2a). As a motivating example for how a partially observed network may arise, consider a classroom setting. The researcher collects information about students' social interactions, which are represented by the network adjacency matrix G where $G_{ij} = 1$ if and only if individuals i and j are friends. However, due to financial and time constraints, the researcher subsamples a set of students $\bar{\boldsymbol{n}} \subset \boldsymbol{n}$ and asks who they are friends with. The interviewed students reveal social ties that they have with everyone in the school population. As a result, the connections between individuals in the subsample $\bar{\boldsymbol{n}}$ to individuals in $\boldsymbol{n}$ are revealed (shaded area of Figure F1a). I assume that the network is undirected, so that i is friends with j if and only if j is friends with i.[9] As a result, links from individuals in $\boldsymbol{n}$ to $\bar{\boldsymbol{n}}$ are also revealed and hence the shaded area in Figure F1b is observed by the researcher. I partition the network into two components $G = (G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})$, where $G^{\bar{\boldsymbol{n}}}$ represents information known to the researcher.

More generally, $G^{\bar{\boldsymbol{n}}}$ represents any network with missing links. The minimal requirement is that all links are observed between a small set of individuals in $\bar{\boldsymbol{n}} \subset \boldsymbol{n}$. This is a very weak requirement and holds, for instance, when the researcher asks one student in the classroom to reveal a randomly chosen friend. While links to individuals outside of $\bar{\boldsymbol{n}}$ are not required and the size of $\bar{\boldsymbol{n}}$ can trivially equal two, the bounds on the network statistic of interest are tighter when more information is available. I assume that the researcher observes T partially observed networks $\{G_t^{\bar{\boldsymbol{n}}_t}\}_{t=1}^{T}$, where $G_t^{\bar{\boldsymbol{n}}_t}$ is a random set of links from the network G_t. These can be thought of as a cross-section from many markets. For instance, T may count the number of schools surveyed (Harris, 2009) or the number of remote villages in India (Banerjee et al., 2013). I also assume that the researcher has access to covariates $\boldsymbol{x}_{it}$ for all individuals in $\boldsymbol{n}_t$ (e.g., from a census or school enrollment data).[10]

Assumption 1 (Observational Assumption)**.** *The researcher observes an i.i.d. sequence* $\{G_t^{\bar{\boldsymbol{n}}_t}, \boldsymbol{x}_t\}_{i \in \boldsymbol{n}_t, t=1,\dots,T}$ *with* $T \to \infty$*. The partially observed network* $G_t^{\bar{\boldsymbol{n}}_t}$ *contains a random sample of links from* G_t *(i.e., links are missing at random).*

Example 1 (Complete Survey Data)**.** *Consider the case where the partially observed network* $G_t^{\bar{\boldsymbol{n}}_t}$ *is constructed from a random survey of individuals in a population,* $\bar{\boldsymbol{n}}_t \subset \boldsymbol{n}_t$*. Suppose that all surveyed individuals reveal all of their friendships to the researcher. Assumption 1 holds this type of survey data.*

[9] In certain contexts, we may believe that the network is directed. For example in a classroom, students may form friendships with the athletic star, but the star does not reciprocate. My framework can be easily extended to the case of directed networks by using directed network-formation models, e.g., Gualdani (2019).

[10] My framework can be extended to the case where we only have survey data on $\boldsymbol{x}_t$. The network is reconstructed using the estimated distribution of $\boldsymbol{x}_t$. This, however, will result in wider bounds as I lose information.

4 Model

My framework can be thought of as a two-stage model. In the first stage the network is formed according to a social network-formation model parameterized by $\theta \in \Theta$. In the second stage, a network statistic $d(G)$ results, where $d : \mathcal{G} \to B$ is a known function and $B \subseteq \mathbb{R}^p$. The statistic can be a simple function of the network G, e.g., the number of links, but also an endogenous outcome from a game played on the network. In the linear social interactions game, for example, individuals simultaneously choose effort. The equilibrium level of effort is equal to a statistic that has received much attention in the literature, the Katz-Bonacich Centrality. Other second-stage applications are also considered in the appendix, such as the game described in Battaglini and Patacchini (2018). The goal is to learn about $(\theta, d(G))$ given a distribution of observables (partially observed networks, characteristics, and outcomes) P, under the conditions imposed by the model. I first discuss important centrality measures and I illustrate worst-case bounds for these statistics when no assumptions are imposed on the missing links.

4.1 Network Statistics and Centrality Measures

Centrality measures are network statistics that allow researchers to parsimoniously capture different features of an individual's network position. For example, KBC has been used to characterize the way peer effects impact student outcomes (Calvó-Armengol, Patacchini, & Zenou, 2009); intercentrality can identify the key player in a criminal network (Ballester et al., 2006); and diffusion centrality is a strong predictor of how information will spread across the network (Banerjee et al., 2013). For exposition, I first focus on KBC and then generalize to other centrality measures in Section 4.1.2.

4.1.1 Katz-Bonacich Centrality

KBC was first proposed in Katz (1953) and further developed by Bonacich (1987). It is a weighted measure of all paths leading to a particular individual, with shorter paths receiving larger value. Individuals that have a lot of friends have a higher KBC than those with few or no friends, *ceteris paribus*. Popular individuals that are connected to popular individuals also have a higher KBC than popular individuals connected to isolated individuals.

Definition 1 (Weighted Katz-Bonacich Centrality)**.** Consider a network G and fix a weighting vector $\boldsymbol{w} \in \mathbb{R}^n$ and a decay parameter $\lambda \in \mathbb{R}$. The *Weighted Katz-Bonacich Centrality*

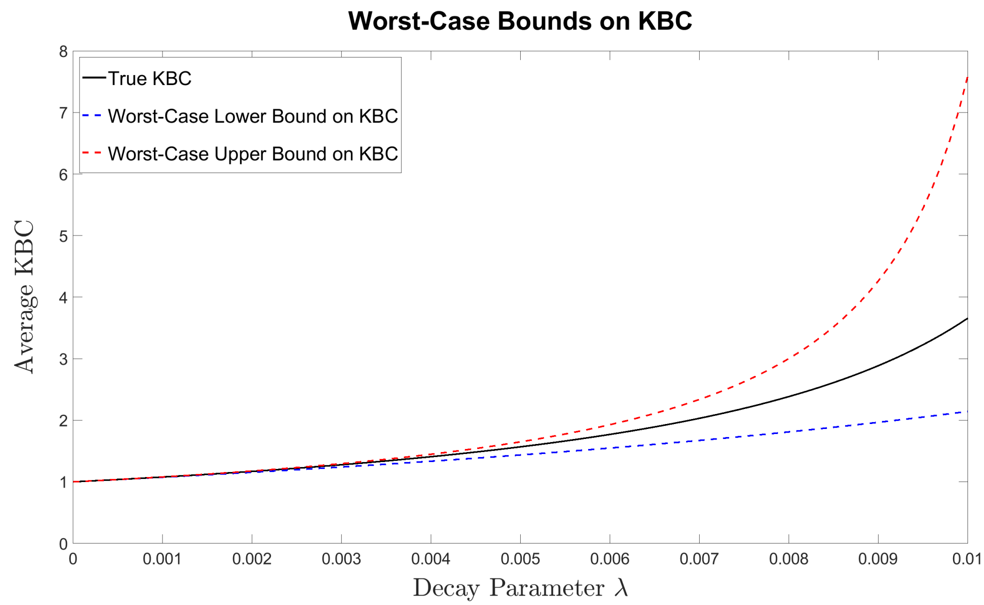


Figure 1: Worst-case bounds on peer effort. The solid line plots the true level of effort as a function of the social multiplier λ, and the dotted lines plot worst-case bounds. Results are reported for a network of size $n = 100$ with $\bar{n} = 25$ observed individuals in a Complete Network Survey. The network is generated according to the model described in Section 6 with $\gamma_1 = \gamma_2 = 0.5$ and $\theta_0 = 0$.

is

$$d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}.$$

The *unweighted Katz-Bonacich Centrality* corresponds to $\boldsymbol{w} = (1, \ldots, 1)'$, so that the expression simplifies to $d^{\text{kbc}}(G; \lambda) = \sum_{k=0}^{\infty} \lambda^k G^k$.

The weighting vector $\boldsymbol{w}$ allows for the possibility that certain individuals have larger influence over their friends' behavior. The weight w_i can be negative in which case the individual is a *negative influencer*. A path is a sequence of friendships between two individuals. The matrix G^k counts all (not necessarily unique) paths of length k. The decay parameter λ controls the importance of longer paths: longer paths contribute less to the value of KBC when λ is small. As $\lambda \to 0$, $d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) \to \boldsymbol{w}$ in which case all paths are irrelevant. Denote the largest eigenvalue of G by $\mu(G)$. If $\lambda\mu(G) < 1$, then KBC is finite and given by

$$d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) = (I - \lambda G)^{-1} \boldsymbol{w}$$

Four economic applications relating to KBC are provided in Examples 8, 9, C1.

KBC requires full knowledge of the network and cannot be computed from a partially observed network $G^{\bar{\boldsymbol{n}}}$. Intuitively, paths of length two – i.e., friends-of-friends – impact the value of KBC. Even in the case that we observe complete survey data (Example 1), not all friends-of-friends are not observed. Worst-case bounds (Manski, 1989, 2003) on KBC are obtained by imposing no restrictions on the unobserved links. Suppose that the weighting vector is positive, $\boldsymbol{w} \in \mathbb{R}^n_+$. Then KBC is monotone in the network and worst-case bounds are obtained by setting the unobserved links to zero or one. Figure 1 illustrates worst-case bounds as a function of the decay parameter λ. When λ is small, long paths have little influence and KBC is driven by the first order term $\lambda^0 G^0 \boldsymbol{w} = \boldsymbol{w}$ (normalized to one in this case). As a result, the discrepancy between G and the worst-case assignments for its unobserved portion is irrelevant and the worst-case bounds' interval is thin. However, when the decay parameter is larger, the discrepancy becomes a serious issue and the worst-case bounds are very wide. This example illustrates the importance of accounting for unobserved links when estimating network statistics, centrality measures, or equilibrium behavior.

While worst-case bounds can be uninformative, they do not rely on any data or model. However, if we have a small amount of data on friendships, I can dramatically shrink the bounds. To accomplish this, I need a network-formation model and a clever computational approach. The network-formation model is discussed in Section 4.2 and the formal identification approach is detailed in Section 5. First, I provide generic bounds on network statistics (e.g., KBC) subject to G belonging to a "nice set" – a lattice.

4.1.2 Bounding Centrality Measures

I later show that, under the assumptions imposed by my network formation model, all possible equilibria networks belong to a lattice. A lattice is a set of networks that are binary and componentwise bounded between two networks $\underline{G}$ and $\overline{G}$. Formally,

$$\mathcal{L}(\underline{G}, \overline{G}) = \{G \in \mathcal{G} : \underline{G} \leq G \leq \overline{G}\}, \quad \text{where} \quad \underline{G} \leq \overline{G}.$$ [11]

The goal is to solve the following problem

$$\min_{G \in \mathcal{L}(\underline{G}, \overline{G})} / \max \, d(G), \tag{1}$$

where $d(G)$ is a network statistic of interest. The solution to Problem (1) improves on the worst-case bounds by using information about the model to restrict the set of possible networks to $\mathcal{L}(\underline{G}, \overline{G})$. Since $\mathcal{L}(\underline{G}, \overline{G})$ belongs to a discrete space, the naïve solution to Problem

[11] The partial order $\leq$ refers to element-wise dominance. That is, $G \leq G'$ if and only if $\forall i, j$, $G_{ij} \leq G'_{ij}$.

(1) is to check all networks in $\mathcal{L}(\underline{G}, \overline{G})$. If, however, $\underline{G}$ and $\overline{G}$ differ by k elements, then this requires 2^k evaluations of $d(G)$ – even for relatively small values of k this is not feasible. Monotonicity is a very useful property of a network statistic that delivers an analytical solution to Problem (1).

Definition 2 (Monotonic Network Statistic)**.** A network statistic $d(G)$ is *monotonically increasing* if and only if $d(G) \leq d(G')$ for all networks $G, G' \in \mathcal{G}$ such that $G \leq G'$. A network statistic $d(G)$ is *monotonically decreasing* if and only if $d(G) \geq d(G')$ for all networks $G, G' \in \mathcal{G}$ such that $G \leq G'$.

Under monotonicity, the solutions to Problem (1) are the extreme points of the lattice.

Lemma 1. *Fix $\underline{G}$ and $\overline{G}$ and consider a network statistic $d(G)$.*

1. *If $d(G)$ is monotonically increasing, then the solutions to Problem (1) are $\underline{G}$ and $\overline{G}$, respectively.*
2. *If $d(G)$ is monotonically decreasing, then the solutions to Problem (1) are $\overline{G}$ and $\underline{G}$, respectively.*

Remark 1. Lemma 1 is sharp in the sense that it delivers the exact solution to Problem (1). That is, if we know from a model and data that the true network G belongs to $\mathcal{L}(\underline{G}, \overline{G})$ and we are unable to further refine this set, then Lemma 1 delivers the tightest possible bounds on the network statistic.

Lemma 1 establishes bounds on many network statistics, including KBC, provided that the sign of w_i is the same for all individuals. I now provide a list of centrality measures that Lemma 1 applies to. Each of these centrality measures are useful in parsimoniously describing various forms of economic activity resulting from social interaction, and are commonly used in many practical applications.[12]

Example 2 (Monotonic Katz-Bonacich Centrality)**.** $d^{kbc}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}$, *where* $\lambda \geq 0$. *Suppose that the largest eigenvalue $\mu(\overline{G})$ of $\overline{G}$ satisfies $\lambda\mu(\overline{G}) < 1$. In addition, suppose that the weights have the same sign: $\boldsymbol{w} \in \mathbb{R}^n_+ \cup \mathbb{R}^n_-$. Then KBC is well defined and*

[12] These centralities are summarized in Bloch et al. (2019). Bloch et al. show that all centrality measures belong to a particular family that is characterized by set of axiom. That is, all centrality measures satisfies a particular set of axioms and, hence, all are related to one another. These axioms, however, do not imply that the statistic is monotonically increasing in the network. Indeed, eigenvector centrality is a counterexample.

is monotonic in the network. An economic application for the monotonic KBC is given in Example C2.

Example 3 (Diffusion Centrality). $d_i^{df}(G;\lambda,K) = \sum_{k=1}^{K}\sum_{j=1}^{n}\lambda^k G_{ij}^k$, *where* $\lambda \geq 0$. *Diffusion centrality is equal to a truncation of the unweighted KBC. In particular, all non-unique paths up to length* K *enter this measure; KBC is the limiting case with* $K \to \infty$. *Diffusion centrality was proposed by Banerjee et al. (2013) and Banerjee, Chandrasekhar, Duflo, and Jackson (2014) to capture the information flow of microfinance uptake in remote rural villages in India.*

Example 4 (Degree Centrality). $d_i^{deg}(G) = \sum_{j=1}^{n} G_{ij}$. *Degree centrality is a measure of popularity. It is useful for understanding the friendship paradox and its implication on the statistical properties of peer effort (Jackson, 2019). Degree centrality has also been used to understand power laws and disease epidemics (Easley & Kleinberg, 2010).*

Example 5 (Closeness Centrality). $d_i^{cc}(G) = \frac{n-1}{\sum_{j=1}^{n}\rho_{ij}(G)}$, *where* $\rho_{ij}(G)$ *is the shortest distance between* i *and* j *in network* G. *If* i *is isolated, then* $d_i(G) = 0$.

A closely related centrality measure is harmonic centrality.

Example 6 (Harmonic Centrality). $d_i^{hc}(G) = \sum_{j=1}^{n}\frac{n-1}{\rho_{ij}(G)}$. *Harmonic and closeness centrality have been proposed to capture the speed at which a message is transmitted through a network (Bavelas, 1950; Sabidussi, 1966).*

Example 7 (Decay Centrality). $d_i^{dc}(G;\lambda) = \sum_{k=1}^{n-1}\sum_{j=1}^{n}\lambda^k \mathbb{1}(\rho_{ij}(G) = k)$. *While diffusion centrality counts all paths up to length* K *and KBC counts all paths, decay centrality only counts shortest paths. The maximum length of the shortest is equal to* $n-1$. *Decay centrality has been proposed for optimal targeting/treatment in a social network (Banerjee et al., 2013; Chatterjee & Dutta, 2016; Tsakas, 2016a, 2016b) and can lead to the greatest amount of information diffusion about, e.g., a new vaccine.*

Proposition 1. *The centrality measures in Examples 2–7 are either monotonically increasing or decreasing in the network. Hence, solutions to Problem (1) for these centrality measures are given by Lemma 1.*

There exist network statistics that are not monotonic in the network. Examples include eigenvalue centrality, targeting centrality (Bramoullé & Genicot, 2018), and the nodal neighborhood statistic in Bloch et al. (2019). Despite this, there still exist computationally feasible bounds to Problem (1) for certain statistics. That is, I find $\underline{d}(\underline{G}, \overline{G})$ and $\overline{d}(\underline{G}, \overline{G})$ that are informative and satisfy

$$\underline{d}(\underline{G}, \overline{G}) \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \quad \text{and} \quad \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \leq \overline{d}(\underline{G}, \overline{G}).$$

These bounds are solved on a case-by-case basis. I now present two important examples, for which I derive informative bounds in Propositions 2 and 3 below.

Example 8 (Non-monotonic Katz-Bonacich Centrality). *Let* $d^{kbc}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}$. *and suppose that the largest eigenvalue of* $\overline{G}$ *denoted* $\mu(\overline{G})$ *satisfies* $\lambda \mu(\overline{G}) < 1$. *However, suppose that the weights do not have the same sign, so that* $w_i > 0$ *and* $w_j < 0$ *for some* i *and* j. *KBC is not monotonic in* G *in this case. As a concrete example for this statistic, consider the linear social interactions model. Taking the network* G *as given, suppose that the students simultaneously choose schooling effort* $y_i \in \mathbb{R}$ *to maximize the following utility function:*

$$v_i(\boldsymbol{y}, G; \alpha_i, \phi) = \alpha_i y_i - \frac{1}{2} y_i^2 + \phi \sum_{j=1}^{n} G_{ij} y_i y_j. \tag{2}$$

The first term $\alpha_i y_i - \frac{1}{2} y_i^2$ *captures the direct benefit of effort. The term* α_i *allows for heterogenity in the marginal returns to effort and is typically modeled as a linear function of observable characteristics* $\boldsymbol{z}$ *and an unobservable idiosyncratic shock. The second term* $\phi \sum_{j=1}^{n} G_{ij} y_i y_j$ *captures local spillovers from direct friends exerting effort. The unique Nash equilibrium is* $\boldsymbol{y}^*(G; \boldsymbol{\alpha}, \phi) = (I - \phi G)^{-1} \boldsymbol{\alpha} = c(G; \boldsymbol{\alpha}, \phi)$. *That is, equilibrium effort is equal to KBC with weights* $\boldsymbol{\alpha}$ *and decay parameter* ϕ. *In this model, the marginal returns to effort* α_i *can be negative or positive. Consequently, KBC is generally not monotonic in the network.*

Example 9 (Intercentrality). *Denote intercentrality by* $d^{int}(G; \phi) \equiv \frac{d_i^{kbc}(\underline{G}; \phi)}{(I - \phi G)_{ii}^{-1}}$. *Intercentrality is a non-linear transformation of KBC and is a network statistic that can provide a guide for determining the Key Player (Ballester et al., 2006; Zenou, 2016) – the individual who if removed would result in the largest decrease in total activity. In a criminal network, the Key Player is the person who law enforcement should target if the goal is to reduce criminal* *activity. The Key Player is the solution to:*

$$\min_{i=1,\ldots,n} \sum_{j=1}^{n} y_i^*(G^{[-i]}),$$

where $y_i^*(G^{[-i]})$ is the equilibrium level of effort after removing individual i from the network. Under the assumptions on preferences detailed in their paper, Ballester et al. (2006) show that the Key Player is the individual with the largest intercentrality.

Proposition 2. *Fix $\boldsymbol{w}, \phi, \underline{G}$, and $\overline{G}$ and order $\boldsymbol{w}$*

$$\boldsymbol{w} = (w_1, \ldots, w_{n_+}, w_{n_+ +1}, \ldots, w_n)$$

such that $\forall i \leq n_+ : w_i \geq 0$ and $\forall i > n_+ : w_i < 0$. Suppose that the largest eigenvalue of $\overline{G}$ denoted $\mu(\overline{G})$ satisfies $\phi\mu(\overline{G}) < 1$. Define mappings $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ and $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ by

$$\underline{d}(\underline{G}, \overline{G}; \phi) = [(I - \phi\underline{G})_1^{-1}, \ldots, (I - \phi\underline{G})_{n_+}^{-1}, (I - \phi\overline{G})_{n_+ +1}^{-1}, \ldots (I - \phi\overline{G})_n^{-1}]$$
$$\overline{d}(\underline{G}, \overline{G}; \phi) = [(I - \phi\overline{G})_1^{-1}, \ldots, (I - \phi\overline{G})_{n_+}^{-1}, (I - \phi\underline{G})_{n_+ +1}^{-1}, \ldots, (I - \phi\underline{G})_n^{-1}]$$

where $(I - \phi G)_j^{-1}$ is the j^{th} column of $(I - \phi G)_1^{-1}$. Then $\forall i = 1, \ldots, n$:

$$[\underline{d}(\underline{G}, \overline{G}; \phi)\boldsymbol{w}]_i \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d_i^{kbc}(G; \boldsymbol{w}, \phi)$$
$$[\overline{d}(\underline{G}, \overline{G}; \phi)\boldsymbol{w}]_i \geq \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d_i^{kbc}(G; \boldsymbol{w}, \phi)$$

Proposition 3. *Fix $\phi, \underline{G}$, and $\overline{G}$. Suppose that the largest eigenvalue of $\overline{G}$ denoted $\mu(\overline{G})$ satisfies $\phi\mu(\overline{G}) < 1$. Then*

$$\frac{d_i^{kbc}(\underline{G}; \phi)}{(I - \phi\overline{G})_{ii}^{-1}} \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d^{int}(G; \phi) \quad \text{and} \quad \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d^{int}(G; \phi) \leq \frac{d_i^{kbc}(\overline{G}; \phi)}{(I - \phi\underline{G})_{ii}^{-1}}$$

Remark 2. There does not always exist a $G \in \mathcal{L}(\underline{G}, \overline{G})$ such that $\overline{d}(\underline{G}, \overline{G}; \phi) = (I - \phi G)^{-1}$, and hence the bounds given in Propositions 2 and 3 are generally not sharp. See Appendix B for details.

I present one final example for non-monotonic statistics in the appendix, see Example C1 and Proposition B1.

When the researcher has access to partial network data, I have shown that worst-case bounds – bounds obtained by imposing no restrictions on the missing links – can be uninformative. I have obtained bounds on a wide range of centrality measures that are commonly used in the applied networks literature subject to the network belonging to a generic lattice. I will now specify a network-formation model that restricts the set of possible networks to a particular lattice. The model can be estimated using the partially observed data to provide informative bounds on the network statistic.

4.2 Network Formation Model

The network-formation model is off-the-shelf and is featured in many theoretical and structural network-related papers (Jackson & Wolinsky, 1996; de Paula et al., 2015; Currarini et al., 2009; Miyauchi, 2016; Sheng, 2018), none of which focus on partially observed networks. Specifically, I assume that individuals play a complete information, pairwise stable game and that links are undirected. One important feature of the model is that it allows for positive network externalities, resulting in individuals strategically choosing their friends.

Let $\boldsymbol{n} = \{1, \cdots, n\}$ be the set of individuals in a population and let $i, j, k, l \in \boldsymbol{n}$ be arbitrary individuals. The network is encoded by the adjacency matrix G where $G_{ij} = 1$ if and only if i and j are linked. The network is undirected, $G_{ij} = G_{ji}$, and there are no self loops, $G_{ii} = 0$. Denote the space of such matrices by

$$\mathcal{G} = \{G \in \mathbb{Z}_2^{n\times n} : G_{ii} = 0,\ G_{ij} = G_{ji} \quad \forall i, j \in \boldsymbol{n}\}.$$

The notation $G + \{ij\}$ denotes the network with the link ij added (i.e., $G + \{ij\}$ has k, l entry equal to G_{kl} for all $kl \neq ij$ and $G_{ij} = 1$). Similarly, $G - \{ij\}$ denotes the network with the link ij deleted. Each individual $i \in \boldsymbol{n}$ is characterized by a vector of observable characteristics $\boldsymbol{x}_i \in \mathcal{X}$ (collect these in the matrix $\boldsymbol{x} = (\boldsymbol{x}_i)_{i\in\boldsymbol{n}}$), a matrix of preference shocks $\varepsilon \equiv (\varepsilon_{ij})_{ij\in\boldsymbol{n}} \in \mathcal{E}$, and a network utility function $\pi_i : \mathcal{G} \times \mathcal{X} \times \mathcal{E} \times \Theta \to \mathbb{R}$. The network utility function $\pi_i(\cdot, \cdot, \cdot; \theta)$ is parameterized by the same value θ_0 for all individuals and it represents the value that individual i places on network $G \in \mathcal{G}$. Marginal utility is a key ingredient in defining an equilibrium. In contrast to standard economic models, utility is a function of the discrete network G. Hence, the marginal utility of a link is defined as the difference in utilities when the link is present and when it is not present.

Definition 3 (Marginal Utility)**.** The marginal utility of individual i over link ij is the mapping $\Pi_{ij} : \mathcal{G} \times \mathcal{X} \times \mathcal{E} \times \Theta \to \mathbb{R}$ defined as:

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \equiv \pi_i(G + \{ij\}, \boldsymbol{x}, \varepsilon; \theta) - \pi_i(G - \{ij\}, \boldsymbol{x}, \varepsilon; \theta).$$

I maintain the assumption of pairwise stability with non-transferable utility as an equilibrium condition.[13] Pairwise stability (Jackson & Wolinsky, 1996) ensures that any two individuals are unwilling to mutually create a link and no individual is willing to sever a link. Other equilibrium concepts have been proposed such as Nash equilibrium and coalition equilibrium. The Nash equilibrium concept is not attractive in network games due to coordination failure and, in particular, because the network with no links is a Nash equilibrium (Myerson, 1991; Calvó-Armengol & İlkılıç, 2009). Coalition equilibrium is a refinement of pairwise stability and ensures that no group of individuals are willing to renegotiate the set of links between them (Myerson, 1977; Jackson & Wolinsky, 1996). I do not impose coalition equilibrium as it is stronger than pairwise stability and I aim to impose minimal assumptions on equilibrium behavior.

Definition 4 (Pairwise Stable Network)**.** Given $\boldsymbol{x}$, ε, and a payoff function π_i, the network G is said to be *pairwise stable* if the following conditions hold:

1. For all $i, j \in \boldsymbol{n}$ such that $G_{ij} = 1$,

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0 \quad \text{and} \quad \Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0,$$

2. For all $i, j \in \boldsymbol{n}$ such that $G_{ij} = 0$,

$$\text{if} \quad \Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) > 0, \quad \text{then} \quad \Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) < 0.$$

A key assumption that I maintain is that marginal utility is monotonically increasing in the network. Intuitively, this assumes that there are positive externalities when individuals form friends and, in particular, allows students in the classroom to receive larger value from connecting to the popular student. This type of assumption is also referred to as strategic complementarity or supermodularity in the microeconomic theory literature (Tarski et al., 1955; Topkis, 1978). The assumption allows me to refine the set of networks to a lattice and

[13] The game with transferable utility is discussed in Sheng (2018), and the game with directed networks is discussed in Gualdani (2019).

guarantees the existence of an equilibrium.

Assumption 2 (Monotonically Increasing)**.** *Marginal utility is monotonically increasing in G: if $G \leq G'$, then*

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \leq \Pi_{ij}(G', \boldsymbol{x}, \varepsilon; \theta) \quad \forall i, j \in \boldsymbol{n}.$$

Monotonically increasing marginal utility is a strong assumption; it rules out competition and cannibalization effects present in many Industrial Organization applications (Berry & Jia, 2008; Jia, 2008; Nishida, 2014).[14] Nonetheless it holds in many applications of interest, such as models of social interaction where individuals tend to form connections with popular individuals.

In the application I assume the following functional form for the utility function.

Assumption 3 (Linear Utility)**.** *Utility is given by*

$$\pi_i(G, \boldsymbol{x}, \varepsilon; \theta) \equiv \sum_{j \in \boldsymbol{n}} G_{ij}(u(\boldsymbol{x}_i, \boldsymbol{x}_j; \theta) + \varepsilon_{ij}) + \gamma_1 \frac{1}{n-1} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{ik} G_{jk},$$

so that marginal utility is linear in parameters:

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) = u(\boldsymbol{x}_i, \boldsymbol{x}_j; \theta) + \varepsilon_{ij} + \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk}.$$

This particular specification for the marginal utility function has three components. The first term, $u(\boldsymbol{x}_i, \boldsymbol{x}_j, \theta) + \varepsilon_{ij}$, is the direct benefit that i gets by connecting to individual j. This term allows for homophily – that is, individuals who have similar characteristics are more likely to become friends. Second, the popularity spillover is given by $\gamma_1 \frac{1}{n-1} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{jk}$. The term $\frac{\gamma_1}{n-1}$ is the marginal value that individual i gets from having one more indirect link. Finally, the mutual friend spillover is given by $\gamma_2 \frac{1}{n-2} \sum_{j,k \in \boldsymbol{n}: k \neq i} G_{ij} G_{ik} G_{jk}$. The term $\frac{\gamma_1}{n-2}$ is the marginal value that individual i gets from having one more mutual link. The two spillover terms are normalized by their maximum values $(n-1)$ and $(n-2)$, so that the spillovers takes values in $[0, \gamma_1]$ and $[0, \gamma_2]$, respectively. Under this utility specification marginal utility is monotonically increasing provided that the popularity and mutual friend

[14] Jia (2008) shows that in a two-player network game with competition effects, a simple transformation of the model can be performed to yield entry decisions that satisfy strategic complementarity. This, however, cannot be generalized to markets with more than two players. The pairwise stable network game consists of $n > 2$ players and, therefore, monotonically increasing marginal utility is required.

spillovers are non-negative.

4.2.1 Equilibrium Results

There are three useful results related to the characterization and existence of the equilibrium that are based on games with strategic complementarities (Tarski et al., 1955; Topkis, 1978; Milgrom & Roberts, 1990; Jia, 2008), see Miyauchi (2016) for the network case. The first result characterizes pairwise stability in terms of a fixed-point mapping.

Fixed-point Characterization. *Fix $\boldsymbol{x}$, ε, and a payoff functions π_i. Define the mapping $V : \mathcal{G} \to \mathcal{G}$ by*

$$V_{ij}(G) \equiv \mathbb{1}[\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0]\mathbb{1}[\Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0],$$

where $\mathbb{1}(\cdot)$ is the indicator function. The network G is pairwise stable if and only if $G = V(G)$.

The fixed-point characterization provides a useful algorithm for checking whether G is pairwise stable. General conditions for the existence of a pairwise stable network are given in Jackson and Watts (2001) and Hellmann (2013). Sheng (2018) shows by example that if Assumption 2 does not hold, then there are cases where no pairwise stable network exists. A sufficient condition for the existence of a pairwise stable network with non-transferable utility is that Assumption 2 holds.

Existence of an Equilibrium. *Let Assumption 2 hold. For all $\boldsymbol{x} \in \mathcal{X}$, $\varepsilon \in \mathcal{E}$, there exists at least one pairwise stable network.*

Under monotonicity, the set of pairwise stable equilibria belong to a lattice.

Set of Equilibria. *Let Assumption 2 hold. Given $\boldsymbol{x}, \varepsilon$, there exists networks $\underline{G}(\boldsymbol{x}, \varepsilon, \theta)$ and $\overline{G}(\boldsymbol{x}, \varepsilon, \theta)$ such that:*

1. *$\underline{G}(\boldsymbol{x}, \varepsilon, \theta)$ and $\overline{G}(\boldsymbol{x}, \varepsilon, \theta)$ are pairwise stable; and*
2. *If G is pairwise stable, then $\underline{G}(\boldsymbol{x}, \varepsilon, \theta) \leq G \leq \overline{G}(\boldsymbol{x}, \varepsilon, \theta)$.*

I define any network that belongs to the equilibria lattice to be *admissible.*

Definition 5 (Admissible Set of Networks, Pairwise Stable Set of Network)**.** Let $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$

denote the *Admissible Set of Networks*:

$$\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon) \equiv \{G \in \mathcal{G} : \underline{G}(\boldsymbol{x}, \varepsilon, \theta) \leq G \leq \overline{G}(\boldsymbol{x}, \varepsilon, \theta)\},$$

Similarly, let $\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon)$ denote the *Pairwise Set of Stable Networks*:

$$\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon) \equiv \{G \in \mathcal{G} : G \text{ is pairwise stable}\}.$$

As the following proposition shows, the lattice $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ is very quick to compute while computation of $\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon)$ is infeasible. Typically not all networks in $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ are pairwise stable. Hence, to compute $\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon)$ I need to evaluate $V(G)$ for all networks G in the admissible lattice. If $\underline{G}(\boldsymbol{x}, \varepsilon, \theta)$ and $\overline{G}(\boldsymbol{x}, \varepsilon, \theta)$ differ by k elements, then this requires 2^k evaluations of $V(G)$. This is infeasible even for relatively small values of k.

Proposition 4. *The admissible set of networks $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ can be computed in no more than $\frac{n(n-1)}{2} - n + 1$ evaluations of $V(\cdot)$.*

The proof of Proposition 4 also provides a constructive way of computing the lattice. In particular, the lattice can be computed by iteratively applying $V(\cdot)$ to the network of zeros and ones.

Propositions 1 and 4 as well as the results in this section play an important role in identification of the bounds on the statistics, discussed further in Section 5.2. To understand how, consider the case where we have a candidate value for θ and suppose that all relevant characteristics of network-formation are observed by the researcher. I have shown that all networks belong to an admissible lattice $\mathcal{G}_\theta(\boldsymbol{x})$, which, in this case, depend only on observed characteristics $\boldsymbol{x}$. The lattice $\mathcal{G}_\theta(\boldsymbol{x})$ can be feasibly computed and, provided the model is correctly specified, must contain the observed network $G^{\bar{\boldsymbol{n}}}$. The aforementioned results can then be applied to obtain bounds on a network statistic such as KBC, which is what I set out to achieve. There are two issues with this thought experiment. First, there are typically unobserved characteristics ε that affect the admissible lattice $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$. This is resolved by taking expectations and integrating out the unobserved taste shocks. Second, I do not have a candidate value for θ. I use the observed network data and an identification strategy discussed in the next section to provide a set of candidate values for the network-formation parameter, which can then be used to bound the network statistic of interest.

5 Identification and Estimation

I first discuss identification of the network formation model. Identification is challenging due to multiplicity of pairwise stable equilibria – typically the admissible lattice contains more than one network $|\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)| > 1$. If there were a single pairwise stable network, then the network formation model would yield a single model implied distribution for the network $\boldsymbol{G}$, and one would be able to learn θ by matching the observed distribution $\boldsymbol{P}(\boldsymbol{G} = G_0|\boldsymbol{x})$ with the one implied by the model for all $G_0 \in \mathcal{G}$.[15] Due to multiplicity of equilibria, however, the model implies multiple distributions for the network. This identification problem is further aggravated by the assumption that the networks are only partially observed. That is, the data only reveals information about the distribution over partially observed networks $\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} = G_0^{\bar{\boldsymbol{n}}}|\boldsymbol{x})$. These data limitations and multiplicity of equilibria result in a set identification.[16]

I show two theoretical results in Section 5.1. First, I show how to obtain the sharp identified region $\mathcal{H}_P[\theta]$. This set includes all network-formation parameters θ such that the data-implied distribution $\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} = G_0^{\bar{\boldsymbol{n}}}|\boldsymbol{x})$ is consistent with one of the multiple distributions implied by the model. The sharp identified region, however, cannot be feasibly computed. I propose an outer region $\mathcal{O}_P[\theta]$ that contains $\mathcal{H}_P[\theta]$, which can be computed by only considering subnetworks of $G_0^{\bar{\boldsymbol{n}}}$. By restricting the size of the subnetwork, the outer region can be feasibly computed and remains informative about the underlying parameters of the network-formation model. In Section 5.2, I show how to augment $\mathcal{O}_P[\theta]$ with bounds on the network statistic implied by Lemma 1 to obtain a joint outer region $\mathcal{O}_P[\theta, \beta_\theta]$ for both the network-formation parameter and the network statistic of interest, where $\beta_\theta = E_\theta(d(G))$ (the expectation is taken with respect to one of the multiple distributions for G).

5.1 Identification: The Network-Formation Model

Identification in networks with complete information and externalities is both data intensive and computationally difficult. The sharp identified region, $\mathcal{H}_P[\theta]$, is defined by moment inequalities – functions that are less than or equal to zero for $\theta \in \mathcal{H}_P[\theta]$. The sharp identified

[15] In terms of notation, $\boldsymbol{P}(\cdot)$ is the probability measure and capital bold face letters indicate random variables, such as $\boldsymbol{G}$ and $\boldsymbol{X}$. Realized values for the network are capital non-boldface letters such as G. Realized values for covariates are given by lower-case bold face $\boldsymbol{x}$.

[16] There is an additional concern that the frequency estimator for $\boldsymbol{P}(G^{\bar{\boldsymbol{n}}} = G_0^{\bar{\boldsymbol{n}}}|\boldsymbol{x})$ is not precise unless T, the number of observed networks, is very large. The reason for this is that the space of networks $\mathcal{G}$ is very large, and so the likelihood of observing the exact same network-configuration twice is very small. My identification strategy relies on subcomponents of the network, which are precisely estimated even for relatively small values of T.

region requires evaluating $O(2^{2^{\frac{n(n-1)}{2}}})$ moment inequalities, which is not computationally feasible. In light of this challenge, I use a subnetwork identification strategy similar to Sheng (2018). The idea is to derive moment inequalities defined over small subnetworks of the observed network. These moment inequalities define an outer region for the network-formation parameters $\mathcal{O}_P[\theta]$. This strategy resolves the computational challenge, since the number of moment inequalities implied by subnetworks is equal to $2\sum_{m=1}^{q} 2^{\frac{m(m-1)}{2}}$. The integer q is chosen by the researcher to control the computational burden of the problem and is typically set to a number less than six. Formally, a subnetwork is defined as follows.

Definition 6 (Subnetwork $A^{\boldsymbol{s}}$ and Complement Subnetwork $A^{-\boldsymbol{s}}$)**.** Let $\boldsymbol{s} \subset \boldsymbol{n}$ be a subset of the population set of individuals. The *Subnetwork* denoted $A^{\boldsymbol{s}}$ represents all links between individuals in $\boldsymbol{s}$ with other individuals in $\boldsymbol{s}$. The *Complement Subnetwork* denoted $A^{-\boldsymbol{s}}$ represents all other links in the network. That is, $A^{-\boldsymbol{s}}$ contains all undirected links between individuals in $\boldsymbol{n} - \boldsymbol{s}$ with individuals in $\boldsymbol{n}$. The Subnetwork and Complement Subnetwork fully characterize $G = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$.

The dark shaded area in Figure F2a displays a subnetwork $A^{\boldsymbol{s}}$. I choose $A^{\boldsymbol{s}}$ so that its links are contained in the observed part of the network $G^{\bar{\boldsymbol{n}}}$. I select any set $\boldsymbol{s} \subset \boldsymbol{n}$ in the case of complete survey data. The striped area in Figure F2b is the complement subnetwork and this includes all links between individuals not in $\boldsymbol{s}$. See Example C3 for a matrix representation of a subnetwork. I derive bounds on the joint distribution of a subnetwork $A^{\boldsymbol{s}}$ and the characteristics of the individuals within the subnetwork, $\boldsymbol{x}^{\boldsymbol{s}}$. For that, I require the theoretical distribution of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$.

Proposition 5. *The distribution of* $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ *is given by*

$$\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) = \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \left[\sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \right] dF_{\varepsilon}(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}).$$

In Proposition 5, I integrate out characteristics of individuals not in the subnetwork $\boldsymbol{x}^{-\boldsymbol{s}}$. Moving in, I take expectations with respect to unobservable preference shocks ε. Finally, inside the square brackets is the probability that $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ forms conditional on $\boldsymbol{x}$ and ε. Because the model admits multiple equilibria, I require a rule that selects which network forms in equilibrium. The term $\psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta)$ is the network selection mechanism that determines with what probability $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ occurs in equilibrium. The selection mechanism is formally defined below. It is equal to one if G is the unique equilibrium at $(\boldsymbol{x}, \varepsilon; \theta)$ and zero if G is not in the equilibria set. Otherwise, the selection mechanism is any valid probability measure (Tamer, 2003).

Definition 7 (Network Selection Mechanism)**.** The *Network Selection Mechanism* is a measurable function $\psi(\cdot|\boldsymbol{x}, \varepsilon; \theta)$ satisfying the following conditions

1. $\forall G \notin \mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon),\ \psi(G|\boldsymbol{x}, \varepsilon; \theta) = 0$;
2. $\forall G \in \mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon),\ \psi(G|\boldsymbol{x}, \varepsilon; \theta) \in [0, 1]$; and
3. $\sum_{G \in \mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon)} \psi(G|\boldsymbol{x}, \varepsilon; \theta) = 1$.

Remark 3. The first key distinction between my approach and Sheng (2018) is that I derive bounds on the joint distribution of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ and Sheng (2018) derives bounds for $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$. My method requires a moment inequality for each possible realization of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$. By limiting the size of the subnetwork I can keep the number of moment inequalities in check. In contrast, Sheng (2018) requires a moment inequality for all combinations of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$, where $F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}})$ is a value for the distribution at a support point $\boldsymbol{x}^{-\boldsymbol{s}}$. If the support is sufficiently rich (e.g., discrete but not binary), the number of moment inequalities implied by all combinations of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$ is very large even when $|\boldsymbol{s}|$ is small. The problem persists even after using exchangeability and equivalence classes to limit the number of moment conditions, which is discussed below. There is, however, a cost from averaging over $\boldsymbol{x}^{-\boldsymbol{s}}$ in Proposition 5 in that we may lose information. In particular, the set of moment inequalities for $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ may be satisfied at a particular value of θ, but may fail for those defined by $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}, F_{\boldsymbol{x}^{-\boldsymbol{s}}}(\boldsymbol{x}^{-\boldsymbol{s}}))$.

I partition the support for ε into regions that characterize the equilibrium status of a particular subnetwork $A^{\boldsymbol{s}} \in \mathcal{G}^{\boldsymbol{s}}$. I will use this partition to decompose the probability mass function for each subnetwork.

Definition 8 (Region of Uniqueness, Region of Multiplicity, Region of Admissibility)**.** Denote $\mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}$ to be the *Region of Uniqueness*. If $\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$, then there exists a complement subnetwork $A^{-\boldsymbol{s}}$ with the property that $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ is a pairwise stable equilibrium. Moreover, any pairwise-stable equilibrium G has $A^{\boldsymbol{s}}$ as its subnetwork: $G = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ for some $A^{-\boldsymbol{s}}$. Denote $\mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}$ to be the *Region of Multiplicity*. If $\varepsilon \in \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$, then there exists a complement subnetwork $A^{-\boldsymbol{s}}$ with the property that

$(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ is a pairwise stable equilibrium. Moreover, there exists a pairwise-stable equilibrium G that does not have $A^{\boldsymbol{s}}$ as its subnetwork: $G = (\tilde{A}^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ for some $\tilde{A}^{\boldsymbol{s}} \neq A^{\boldsymbol{s}}$ and $A^{-\boldsymbol{s}}$. Denote $\mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}$ to be the *Region of Admissibility*. If $\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$, then there exists $A^{-\boldsymbol{s}}$ such that $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) \in \mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$, i.e., $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ belongs to the admissible lattice.

Using these definitions, I decompose the probability mass function for the subnetwork into two terms.

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} =& A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) \\ &= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \Bigg[\int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \\ &\quad + \int \mathbb{1}[\varepsilon \in \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) dF_\varepsilon(\varepsilon) \Bigg] \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}),\end{aligned}$$

where $\mathbb{1}[\cdot]$ is the indicator function. The first term involving $\mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})]$ is the integral over the region where the subnetwork is unique. The summation $\sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta)$ drops out of the first term as it is equal to one – the selection mechanism picks a network $(A^{\boldsymbol{s}}, A^{-\boldsymbol{s}})$ with $A^{\boldsymbol{s}}$ as its subnetwork almost surely. The second term involving $\mathbb{1}[\varepsilon \in \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})]$ is the integral over the region where the subnetwork is not unique.[17] The selection mechanism is a measurable function, thus the following bounds are satisfied

$$0 \leq \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \leq \sum_{G \in \mathcal{G}} \psi(\boldsymbol{G} = G | \boldsymbol{x}, \varepsilon; \theta) = 1.$$

Therefore, we can set $\sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \in \{0, 1\}$ in Equation **??** to obtain bounds on the probability mass function:

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) &\geq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) \\ \boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) &\leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}).\end{aligned}$$

Feasibly computing these bounds is an important consideration as I need to augment this with moment inequalities implied by the network statistic and construct projected confidence intervals. Checking whether $\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$ is computationally costly and must be repeated for each simulated draw of ε. Sheng (2018) shows that checking $\varepsilon \in$

[17] This is a similar decomposition to the one applied to the Nash entry game (Ciliberto & Tamer, 2009)

$\mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$ for the Transferable Utility case requires solving an optimization routine. Based on times reported in Sheng (2018), it is approximately 100 to 1000 times faster to compute the admissible lattice and work with the following bounds:

$$\boldsymbol{P}(\boldsymbol{A^s} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x^s}; \theta) \geq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}})$$
$$\boldsymbol{P}(\boldsymbol{A^s} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x^s}; \theta) \leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}),$$

which are valid because $\mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \subset \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})$. To that end, I define the following moment inequality functions:

$$\tilde{m}_1(A^{\boldsymbol{s}}, \boldsymbol{x^s}; \theta) \equiv -\boldsymbol{P}(\boldsymbol{A^s} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x^s}) + \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}})$$
$$\tilde{m}_2(A^{\boldsymbol{s}}, \boldsymbol{x^s}; \theta) \equiv \boldsymbol{P}(\boldsymbol{A^s} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x^s}) - \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}).$$

These moment functions satisfy $\tilde{m}_j(A^{\boldsymbol{s}}, \boldsymbol{x^s}; \theta) \leq 0$ for all $\theta \in \mathcal{H}_P[\theta]$, where $\mathcal{H}_P[\theta]$ is the sharp identified region for θ. Hence, these moment inequality functions define a valid outer region. However, even for small subnetworks the number of moment inequalities can be quite large. One way to reduce the number of moment inequalities is to aggregate subnetworks into isomorphic equivalence classes. Two subnetworks $(A^{\boldsymbol{s}}, \boldsymbol{x^s})$ and $(\tilde{A}^{\boldsymbol{s}}, \tilde{\boldsymbol{x^s}})$ are in the same equivalence class if there is a permutation $\tau(\cdot)$ such that $(A^{\boldsymbol{s}}, \boldsymbol{x^s}) = (\tilde{A}^{\tau(\boldsymbol{s})}, \tilde{\boldsymbol{x}}^{\tau(\boldsymbol{s})})$. Two moment inequalities are required for each equivalence class, which dramatically reduces the number of moment inequalities defining the outer region. I further reduce the computational complexity by summing over equivalence classes of $\boldsymbol{x}^{-\boldsymbol{s}}$. If the covariate is binary and univariate, it is sufficient to count the number of cases such that $x_i = 1$ for $i \in \boldsymbol{n} - \boldsymbol{s}$. Denote the equivalence class of $(A^{\boldsymbol{s}}, \boldsymbol{x^s})$ to be $\mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x^s})$. The moment inequalities that I work with are the following:

$$m_1(A^{\boldsymbol{s}}, \boldsymbol{x^s}; \theta)$$
$$\equiv -\boldsymbol{P}((\boldsymbol{A^s}, \boldsymbol{X^s}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x^s})) + \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}}))$$
$$m_2(A^{\boldsymbol{s}}, \boldsymbol{x^s}; \theta)$$
$$\equiv \boldsymbol{P}((\boldsymbol{A^s}, \boldsymbol{X^s}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x^s})) - \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})). \quad (3)$$

This aggregation can, in some cases, result in a loss of information. That is, the moment inequalities $\tilde{m}_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$ can be violated for some θ, but the corresponding moment inequalities $m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$ might hold. Provided the primitives of the utility function and the selection mechanism satisfy interchangeability (de Finetti, 1929; Chernoff & Teicher, 1958; Kallenberg, 2006; Austin, 2008), the moment inequalities can be aggregated up to equivalence classes without loss of information (Sheng, 2018). These interchangeability conditions, along with a full support assumption on the preference shocks ε, imply that these subnetwork frequency estimators are informative (i.e., bounded away from 0 and 1) as the network grows in size (Sheng, 2018, Proposition 4.2). However, interchangeability is restrictive and implies that the network is dense (Orbanz & Roy, 2014).

I specify below econometric and structural assumptions that I maintain for identification. Under these assumptions as well as Assumptions 1 and 2, I display two theorems that describe the sharp identified region and the outer region based on the above moment inequalities over all subnetworks up to a particular size chosen by the researcher. The sharp identified region does not require Assumption 2.

Assumption 4 (Econometric Assumptions)**.** *There is a sequence of random elements $(\boldsymbol{x}_t, \varepsilon_t)$ such that: (1) $\forall t$, $\boldsymbol{x}_t$ and ε_t are independent, and (2) $\forall i \neq j$ and $\forall t$, ε_{ijt} are i.i.d., supported on $\mathcal{E} \subset \mathbb{R}$, and generated from a continuously differentiable parametric distribution $F(\varepsilon; \theta_\varepsilon)$ that depends on the finite-dimensional parameter $\theta_\varepsilon \in \mathbb{R}^{dim(\theta_\varepsilon)}$; (3) $\boldsymbol{x}$ is discrete.*

Assumption 5 (Structural Assumptions)**.** *Each individual $i \in \boldsymbol{n}_t$ receives utility from a network according to a their payoff function $\pi_i(G, \boldsymbol{x}, \varepsilon; \theta)$. Individuals simultaneously choose friendships with complete information over $(\boldsymbol{x}_t, \varepsilon_t)$ with the restriction that the resulting network G_t is pairwise stable.*

Remark 4. The assumption that $\boldsymbol{x}$ is discrete is non-restrictive, as one can discretize a continuous variable. Sheng (2018) discuses the continuous case. For exposition I discuss the discrete case only.

Theorem 1 (Sharp Identified Region for Network Formation Parameter)**.** *Let Assumptions 1, 4, and 5 hold. Define*

$$\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) = \{G^{\bar{\boldsymbol{n}}} \in \mathcal{G} : \exists G^{-\bar{\boldsymbol{n}}} \ s.t. \ (G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}}) \in \mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon)\}$$

That is, $\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ *is a random set that consists of all partially observed networks* $G^{\bar{\boldsymbol{n}}}$ *that can be completed by a pairwise stable network. The sharp identification region for* θ *is given by*

$$\mathcal{H}_P[\theta] = \{\theta \in \Theta : \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K}|\boldsymbol{x}) \leq \boldsymbol{T}_{\mathcal{G}_\theta^{ps}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) \ \forall \mathcal{K} \subset \mathcal{G}^{\bar{\boldsymbol{n}}}, \boldsymbol{x} - a.s\},$$

where $\boldsymbol{T}_{\mathcal{G}_\theta^{ps}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) \equiv \boldsymbol{P}(\{\mathcal{G}_\theta^{ps}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) \cap \mathcal{K} \neq \emptyset\}, \varepsilon \sim F_\varepsilon)$ *is the capacity functional.*

Theorem 1 characterizes the sharp identification region in a similar way to Molinari (2019, Theorem SIR-3.8). In contrast to what I do, Molinari assumes the researcher observes the complete network. The key difference is the information revealed to the researcher and the random set governing Artstein's inequality (Artstein, 1983). While I do maintain Assumption 2 throughout this paper, Theorem 1 does not technically require it. Suppose that Assumption 2 did not hold so that marginal utility is not monotonically increasing in the network. It is possible that there does not exist an equilibrium for some value of θ in which case $\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon) = \emptyset$ and $\boldsymbol{T}_{\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) = 0$. The sharp identified region in this case would exclude these values of θ.

The sharp identified region cannot be feasibly computed as it requires enumerating a doubly exponential number of compact sets. It is, however, useful for establishing an outer region that is computationally feasible. The next Theorem displays an outer region $\mathcal{O}_P[\theta]$ based on subnetwork moment inequalities that contains the set $\mathcal{H}_P[\theta]$.

Theorem 2 (Outer Region for Network Formation Parameter)**.** *Fix an integer* $q \leq |\bar{\boldsymbol{n}}|$ *and suppose that Assumptions 1, 2, 4, and 5 hold. Define*

$$\mathcal{O}_P[\theta] \equiv \{\theta \in \Theta : m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \ \leq 0 \quad j = 1, 2 \quad \forall A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \ s.t. \ |\boldsymbol{s}| \leq q, \ \boldsymbol{s} \subset \bar{\boldsymbol{n}}\}.$$

where $m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta)$ *and* $m_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta)$ *are given in Equation (3). Then* $\mathcal{H}_P[\theta] \subseteq \mathcal{O}_P[\theta]$.

I compute the outer region $\mathcal{O}_P[\theta]$ using simulated methods, which are discussed in detail in Appendix A. Next, I show how to use Propositions 1, 2, and 3 to identify a wide class of centrality measures, including Katz-Bonacich Centrality, diffusion centrality, and decay centrality. These results naturally extend to identifying endogenous outcomes of a game played on the network, such as the peer effects game.

5.2 Identification: Network Statistics

Consider a network statistic $d(G)$. Ideally, I would like to compute the value of $d(\cdot)$ for one of the observed networks $G^{\bar{\boldsymbol{n}}}$. However, computing $d(G)$ typically requires the full network.[18] Using the model, I can take expectations with respect to $d(G)$ conditional on $G^{\bar{\boldsymbol{n}}}$. Due to multiplicity of equilibria, however, the model implies multiple distributions for the network. Therefore, there is a collection of values of $\beta(G^{\bar{\boldsymbol{n}}}) \equiv \boldsymbol{E}(d(\boldsymbol{G})|G^{\bar{\boldsymbol{n}}};\theta)$ that are consistent with the model. Using Propositions 1, 2, and 3, I can bound these values to provide informative bounds on the network statistic of interest.

A related and easier problem that I address first is obtaining bounds on the unconditional expectation of $d(G)$. These bounds can be applied to out-of-sample networks – networks that I have no prior knowledge over. An out-of-sample network is, for example, a friendship network among students in a school that I have not sampled. As another example, I can use the out-of-sample network bounds to execute counterfactual analysis where I hypothetically remove students from the classroom and allow the remaining students to re-optimize their friendship network according to the model. For exposition, suppose that $d(G)$ is monotonically increasing in the network. I will relax this assumption in the main theorem. Conditional on $\boldsymbol{x}$, the expected value of $d(G)$ is given by

$$\beta \equiv \boldsymbol{E}_\theta(d(\boldsymbol{G})|\boldsymbol{x};\theta) = \int_\varepsilon \sum_{G\in\mathcal{G}} d(G)\psi(\boldsymbol{G}=G|\boldsymbol{x},\varepsilon;\theta)dF_\varepsilon(\varepsilon). \tag{4}$$

The joint sharp identified region for (β_θ,θ) includes all values given by Equation (4) for some $\theta \in \mathcal{H}_P[\theta]$ and for some valid network selection mechanism $\psi(\cdot|\boldsymbol{x},\varepsilon;\theta)$.[19] Formally,

$$\mathcal{H}_P[\beta,\theta] \equiv \left\{ \begin{array}{ll} (\beta,\theta)\in B\times\mathcal{H}_P[\theta]: & \exists \text{ network selection mechanism } \psi(\cdot) \\ & \text{s.t. } \beta \text{ satisfies Equation (4)} \end{array} \right\}. \tag{5}$$

Applying Lemma 1, monotonically increasing network statistics are maximized by $\overline{G}(\boldsymbol{x},\varepsilon,\theta)$

[18] In the case of degree centrality, only the direct friends of an individual is required. With complete survey data, the degree centrality is computable for surveyed individuals. However, this information is not available with arbitrary random missing link data.

[19] This sharp identified region can also be derived from the Aumann Expectation of the random set containing admissible values of $d(G)$. The Aumann Expectation is a correspondence from a random set to a set of values that are consistent with one of the multiple underlying distributions (Aumann, 1965; Molchanov, 2005; Molchanov & Molinari, 2018). In this case, it is easier (and isomorphic) to work with the selection mechanism.

on $\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$ from which it follows that

$$\beta \leq \int_\varepsilon d(\overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_\varepsilon(\varepsilon).$$

Similarly, the lower bound is achieved by loading all of the selection mechanism's mass on $\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta})$ to obtain

$$\beta \geq \int_\varepsilon d(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_\varepsilon(\varepsilon).$$

Define the following moment inequalities:

$$\begin{aligned} m_3(\boldsymbol{x}; \beta, \theta) &\equiv -\beta + \int_\varepsilon d(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_\varepsilon(\varepsilon) \\ m_4(\boldsymbol{x}; \beta, \theta) &\equiv \beta - \int_\varepsilon d(\overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_\varepsilon(\varepsilon). \end{aligned} \tag{6}$$

For non-monotone statistics, the moment inequalities are replaced by

$$\int_\varepsilon \underline{d}(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_\varepsilon(\varepsilon) \quad \text{and} \quad \int_\varepsilon \overline{d}(\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}))dF_\varepsilon(\varepsilon),$$

where $\underline{d}(\cdot, \cdot)$ and $\overline{d}(\cdot, \cdot)$ are bounds on Problem (1). The next theorem displays the outer region for (β, θ) for an out-of-sample network G that I have no information on.

Theorem 3 (Outer Region for a Network Statistic)**.** *Fix an integer* $q \leq |\bar{\boldsymbol{n}}|$ *and suppose that Assumptions 1, 2, 4, and 5. Let* $d(G)$ *be a network statistic. Assume that there are mappings* $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *and* $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *such that*

$$\underline{d}(\underline{G}, \overline{G}) \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \quad \textit{and} \quad \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \leq \overline{d}(\underline{G}, \overline{G})$$

for all network lattices $\mathcal{L}(\underline{G}, \overline{G}) = \{G : \underline{G} \leq G \leq \overline{G}\}$*. Define*

$$\mathcal{O}_P[\beta, \theta] \equiv \left\{ \begin{array}{ll} \beta \in B, \theta \in \Theta : & m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0 \quad j = 1, 2 \quad \forall A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \textit{ s.t. } |\boldsymbol{s}| \leq q, \ \boldsymbol{s} \subset \bar{\boldsymbol{n}} \\ & m_j(\boldsymbol{x}; \beta, \theta) \leq 0, \quad j = 3, 4 \quad \boldsymbol{x} - a.s. \end{array} \right\},$$

where

$$m_3(\boldsymbol{x};\beta,\theta) \equiv -\beta + \int_\varepsilon \underline{d}(\underline{G}(\boldsymbol{x},\varepsilon,\boldsymbol{\theta}),\overline{G}(\boldsymbol{x},\varepsilon,\boldsymbol{\theta}))dF_\varepsilon(\varepsilon)$$
$$m_4(\boldsymbol{x};\beta,\theta) \equiv \beta - \int_\varepsilon \overline{d}(\underline{G}(\boldsymbol{x},\varepsilon,\boldsymbol{\theta}),\overline{G}(\boldsymbol{x},\varepsilon,\boldsymbol{\theta}))dF_\varepsilon(\varepsilon).$$

Then $\mathcal{H}_P[\beta,\theta] \subseteq \mathcal{O}_P[\beta,\theta]$.

Remark 5. The above theorem is useful for analyzing policy interventions where we do not observe the network. Consider for example a policy where resources are injected into all schools in a country, but the researcher lacks network data on all of the schools. The researcher has, however, estimated the network-formation model using network data from a sample of students from a sample of schools. Extrapolating the model to non-sampled schools, I can obtain bounds on a centrality measures based on observed characteristics of the students. The centrality measure can then be used to infer how the policy affects outcomes on the unobserved network.

I now turn to obtaining bounds on a network statistic for an in-sample network that is partially observed. Conditional on the realized network $G^{\bar{\boldsymbol{n}}}$, I require that $(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})$ is pairwise stable. This is true if only if the subgame played on links in $G^{-\bar{\boldsymbol{n}}}$ is pairwise stable conditional on $G^{\bar{\boldsymbol{n}}}$. That is, for all links with $G^{-\bar{\boldsymbol{n}}}_{ij} = 1$, it must be the case that the marginal utility over $(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})$ is positive for both i and j. If $G^{-\bar{\boldsymbol{n}}}_{ij} = 0$, then either i has negative marginal utility over j or j has negative marginal utility over i. The sharp identified region $\mathcal{H}_P[\beta(G^{\bar{\boldsymbol{n}}}),\theta]$ is analogous to the one displayed in Equation (5) with the restriction that $\beta(G^{\bar{\boldsymbol{n}}})$ satisfies

$$\beta(G^{\bar{\boldsymbol{n}}}) = \int_\varepsilon \sum_{G^{-\bar{\boldsymbol{n}}}} d(G)\psi(\boldsymbol{G} = (G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}})|G^{\bar{\boldsymbol{n}}},\boldsymbol{x},\varepsilon;\theta)dF_\varepsilon(\varepsilon).$$

I apply Topkis (1978) to this subgame to obtain the conditional admissible lattice denoted

$$\mathcal{G}_\theta(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon) = \{G : \underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon;\theta) \leq G \leq \overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon;\theta)\},$$

where $\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)$ and $\overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)$ are obtained by applying the modified mapping

$$\tilde{V}_{ij}(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}}) = \begin{cases} G^{\bar{\boldsymbol{n}}}_{ij} & \text{if } ij \in G^{\bar{\boldsymbol{n}}} \\ V_{ij}(G^{\bar{\boldsymbol{n}}}, G^{-\bar{\boldsymbol{n}}}) & \text{if } ij \in G^{-\bar{\boldsymbol{n}}} \end{cases}$$

to $(G^{-\bar{\boldsymbol{n}}}, \mathbf{0}^{-\bar{\boldsymbol{n}}})$ and $(G^{-\bar{\boldsymbol{n}}}, \mathbf{1}^{-\bar{\boldsymbol{n}}})$, where $\mathbf{0}^{-\bar{\boldsymbol{n}}}$ and $\mathbf{1}^{-\bar{\boldsymbol{n}}}$ are the zero and one submatrices with dimension equal to $G^{-\bar{\boldsymbol{n}}}$. From this, I obtain the following bounds on a monotonically increasing network statistic $d(G)$:

$$\begin{aligned} \beta(G^{\bar{\boldsymbol{n}}}) &\leq \int_{\varepsilon} d(\overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon) \\ \beta(G^{\bar{\boldsymbol{n}}}) &\geq \int_{\varepsilon} d(\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon). \end{aligned} \tag{7}$$

Consequently, I obtain an outer region similar to the one characterized in Theorem 3.

Theorem 4 (Outer Region for a Network Statistic)**.** *Fix an integer $q \leq |\bar{\boldsymbol{n}}|$ and suppose that Assumptions 1, 2, 4, 5 hold. Let $d(G)$ be a network statistic. Suppose that a partially observed network $G^{\bar{\boldsymbol{n}}}$ is specified. Assume that there are mappings $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ and $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ such that*

$$\underline{d}(\underline{G}, \overline{G}) \leq \min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \quad \textit{and} \quad \max_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G) \leq \overline{d}(\underline{G}, \overline{G})$$

for all network lattices $\mathcal{L}(\underline{G}, \overline{G}) = \{G : \underline{G} \leq G \leq \overline{G}\}$. Define

$$\begin{aligned} &\mathcal{O}_P[\beta(G^{\bar{\boldsymbol{n}}}), \theta] \\ &\equiv \left\{ \begin{matrix} \beta(G^{\bar{\boldsymbol{n}}}) \in B, \theta \in \Theta : & m_j(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0 \quad j = 1, 2 \quad \forall A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \ s.t. \ |\boldsymbol{s}| \leq A, \ \boldsymbol{s} \subset \bar{\boldsymbol{n}} \\ & m_j(\boldsymbol{x}; \beta(G^{\bar{\boldsymbol{n}}}), \theta) \leq 0, \quad j = 3, 4 \quad \boldsymbol{x} - a.s. \end{matrix} \right\}, \end{aligned}$$

where

$$\begin{aligned} m_3(\boldsymbol{x}; \beta(G^{\bar{\boldsymbol{n}}}), \theta) &\equiv -\beta(G^{\bar{\boldsymbol{n}}}) + \int_{\varepsilon} \underline{d}(\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta), \overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon) \\ m_4(\boldsymbol{x}; \beta(G^{\bar{\boldsymbol{n}}}), \theta) &\equiv \beta(G^{\bar{\boldsymbol{n}}}) - \int_{\varepsilon} \overline{d}(\underline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta), \overline{H}(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon; \theta)) dF_{\varepsilon}(\varepsilon). \end{aligned}$$

Then $\mathcal{H}_P[\beta(G^{\bar{\boldsymbol{n}}}), \theta] \subset \mathcal{O}_P[\beta(G^{\bar{\boldsymbol{n}}}), \theta]$.

Theorem 4 is the key result of my paper. It provides us with a joint set of parameters for both the network statistic of interest as well as the network formation parameters that are consistent with the partially observed network $G^{\bar{\boldsymbol{n}}}$. I can leverage this outer region to answer many interesting questions, including: who is the most influential player, who should we optimally target to spread information about a new vaccine, and who is the key player in a criminal network? I use calibrated projection Kaido, Molinari, and Stoye (2019) as implemented by Kaido, Molinari, Stoye, and Thirkettle (2017) to construct confidence intervals. I conclude with a simulation study to show the applicability of my framework.

6 Application

I apply my framework to the Katz-Bonacich Centrality. I can also apply my framework to other centrality measures that are monotonic in the network, including Decay Centrality and Diffusion Centrality. Decay and Diffusion Centralities are especially relevant as they describe how fast information about a new microfinance program spreads through remote villages in India (Banerjee et al., 2013, 2014). I specify the marginal utility of links as a linear function of three terms:

$$\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) = \theta_0|x_i - x_j| + \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk} + \varepsilon_{ij},$$

The first term $\theta_0|x_i - x_j|$ allows for homophily. This is the concept that individuals are more likely to form friends with individuals who have similar characteristics. I set $\theta_0 < 0$ so that if $x_i \neq x_j$, then the marginal utility over the link is smaller and hence a friendship is less likely to occur. I assume x_i is i.i.d, binary (e.g., the individual's sex), and $\boldsymbol{P}(x_i = 1) = 0.5$. This model allows for two channels of spillovers: popularity spillover $\gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk}$ and mutual friend spillover $\gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk}$.[20] I also assume that the additive error ε_{ij} is distributed i.i.d $N(0,1)$ with $\varepsilon_{ij} = \varepsilon_{ji}$, and $\varepsilon_{ij} \perp\!\!\!\perp x_k$ for all i, j, k. I report bounds on the unweighted Katz-Bonacich centrality, $d(G) = (I - \lambda G)^{-1}\mathbf{1}$. The decay parameter λ is in $[0, \bar{\lambda}]$, where $\bar{\lambda}$ is selected to ensure that the network statistic is well defined.

I first restrict the model to one spillover with no homophily in order to understand the performance of the model and the variation required to pin down the set of structural parameters. I show that this parsimonious model is nearly point identified and provides tight bounds on KBC.

[20] The simulations in Sheng (2018) is a variation of the model I present. In her paper, $\gamma_1 = 0$ (i.e., no popularity spillovers) and utility is transferable.

6.1 Popularity Spillover Only

To understand the performance of the model, I first shut down observed characteristics and set $\theta_0 = 0$ and $\gamma_2 = 0$. I assume that this is known to the researcher. Marginal utility is given by

$$\Pi_{ij}(G, \varepsilon; \theta) = \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \varepsilon_{ij}. \tag{8}$$

I first report the size and density of the admissible lattice. I vary the size of the spillover $\gamma_1 \in [0, 3]$ and set the population size equal to $n = 50$. Figure D1c displays the expected fraction of links in $\underline{G}$ (solid line) and $\overline{G}$ (dotted line). The vertical distance between $\overline{G}$ and $\underline{G}$ is proportional to the size of the admissible lattice and counts the fractional difference in the number of links between $\overline{G}$ and $\underline{G}$. On average, the lattice is narrow: for a fixed value of γ_1, there is little variation in the overall density of the equilibrium network. Increasing the size of the spillover results in a denser network. As γ_1 approaches 3, the network becomes a complete network where all individuals are friends. Next, I increase the population size to $n = 100$ in Figure D1d. For every value of γ_1, the lattice becomes more dense. The intuition here is that individuals now have a larger set of potential friends and experience larger spillovers. In addition, the difference in the density between $\overline{G}$ and $\underline{G}$ also declines. What could be happening here is that the lattice is converging to the complete network as n increases.

Table E1 reports the outer region for γ_1. The first column reports the identified region using subnetworks up to size $q = 2$. The second column sets $q = 3$ and the third sets $q = 4$. The rows report results varying the number of individuals in the network $n \in \{50, 100\}$ and the true value for the popularity spiller $\gamma_1 \in \{0, 0.5, 1\}$. The model is effectively point identified when $\gamma_1 = 0$ and is partially identified with $\gamma_1 > 0$. In all cases the identified set is narrow. For example, the identified set is $[0.985, 1.018]$ when $\gamma_1 = 1$. As expected, the identified set tightens as I increase the maximum size of the subnetwork from $q = 2$ to $q = 4$. The identified set is tighter when the population size increases to $n = 100$ as well. Reflecting on Figure D1d, this makes sense as the admissible lattice tightens when n increases, implying that bounds on the distribution of subnetworks also tighten.

Table E3 reports the outer region on the average level of Katz-Bonacich Centrality using Theorem 4. I vary both the size of the popularity spillover γ_1 as well as the decay parameter λ. I report bounds using maximum subnetwork sizes ranging from $q = 2$ to $q = 4$. I also report the true average level of KBC and the worst-case bounds. When the popularity spillovers are zero (i.e., $\gamma_1 = 0$) the model is point identified and equivalent to an exponential random graph model. In this case, the outer region on average KBC is a single point and is equal to the true value of KBC, while worst-case bounds are fairly wide. Increasing the size of the popularity spillovers result in wider bounds on average KBC, but remain tight. In the least preferable case, the worst-case bounds are $[2.85, 5.53]$, which are 12 times wider than the bounds that my model predicts, $[4.02, 4.24]$. Using my framework, therefore, I can obtain informative bounds on the network statistic of interest.

I also report results for the case where only the mutual-friend spillover is present, see Appendix E and D. These are largely in line with what I find for the model with only popularity spillovers. I next consider the model with two channels of spillovers.

6.2 Both Spillovers

Suppose now that marginal utility contains both channels of spillovers, but no homophily:

$$\Pi_{ij}(G, \varepsilon, \theta) = \gamma_1 \frac{1}{n-1} \sum_{k \in \boldsymbol{n}: k \neq i} G_{jk} + \gamma_2 \frac{1}{n-2} \sum_{k \in \boldsymbol{n}: k \neq i} G_{ik} G_{jk} + \varepsilon_{ij}.$$

This is a DGP that has not been explored in previous simulation studies – simulation studies in prior literature restrict attention to models with one spillover. Figure D2 reports the density and the size of the admissible lattice where $(\gamma_1, \gamma_2) \in [0, 2] \times [0, 2]$. The first two panels display the fraction of links in $\underline{G}$ and $\overline{G}$, respectively. Similar to what I find for the single-spillover model, the density approaches 1 as the size of the spillovers increase. The third panel displays the fractional difference between $\overline{G}$ and $\underline{G}$. The maximal difference is only 1%.

Tables E4 and E5 report the projected outer region for (γ_1, γ_2). In contrast to the previous results, the outer region is wide, especially for small subnetworks. For example, when $(\gamma_1, \gamma_2) = (0.5, 0.5)$ and setting $q = 2$, I obtain projected outer regions equal to $[0, 0.9]$ for γ_1 and $[0, 1.16]$ for γ_2. The size of the outer region rapidly contracts with the maximum subnetwork size. One particularly remarkable case is $(\gamma_1, \gamma_2) = (0, 0.5)$ with $q = 4$, where the projected outer regions are $[0, 0.03]$ and $[0.44, 0.50]$ for γ_1 and γ_2, respectively. In reflection of Figure D2 this is not surprising. I only find a 1% difference in the density of $\overline{G}$ and $\underline{G}$, indicating that the theoretical sharp identified set based on the full network is narrow. I lose information by only considering subnetworks with restricted size. Since my model remains tractable for relativity large subnetwork sizes, I am able to feasibly obtain tighter bounds on this particular model.

While the bounds on the preference parameters are wide, the bound on KBC is narrow. Table E6 and E7 reports identified regions for KBC when I allow for two channels of spillovers in the network-formation process. I set $\lambda = 0.01$ in Table E6 and $\lambda = 0.02$ in Table E7. I vary $(\gamma_1, \gamma_2) \in \{0, 0.25, 0.5\} \times \{0, 0.25, 0.5\}$. Despite the fact that the identified region for the preference parameters are typically wide and seemingly uninformative, I obtain precise bounds on network statistic of interest. The outer region on average KBC is $[3.725, 3.792]$ when $(\gamma_1, \gamma_2) = (0.5, 0.5)$ and $\lambda = 0.02$. In contrast, the worst-case bounds are $[2.376, 5.328]$. As a result my bounds are 44 times narrower than the worst-case bounds. Overall, I obtain informative bounds on KBC.

7 Conclusion

Social network data with links missing at random is very common in the social sciences. Applied researchers are often faced with the dilemma of estimating a network statistic when the social network is partially observed and formed endogenously. With cross-sectional network data, the literature either: (1) ignores the fact that the network is partially observed in which case the reported network statistic is incorrect; or (2) the missing links are filled in with an exogenous network-formation process. However, in practice, social networks are formed strategically. For example, students in a classroom choose their friends based on the popularity of others. I provide a framework to obtain informative bounds on network statistics in a partially observed network whose formation I explicitly model. I assume that individuals endogenously form links according to the standard social-network formation model of complete information and pairwise stability. I assume that the researcher has access to cross-sectional data from multiple partially observed networks, where the links are missing at random. I obtain a computationally tractable method to obtain bounds on both the preferences determining network formation processes and network statistics. In a simulation study on the Katz-Bonacich centrality measure, I dramatically reduce the worst-case bounds, which do not use the network formation model. I obtain from my procedure bounds that are 12 times narrower than the worst-case bounds.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, *80*(5), 1977–2016.

Artstein, Z. (1983). Distributions of random sets and random selections. *Israel Journal of Mathematics*, *46*(4), 313–324.

Aumann, R. J. (1965). Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications*, *12*(1), 1–12.

Austin, T. (2008). On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probability Surveys*, *5*, 80–145.

Badev, A. (2018). Nash equilibria on (un)stable networks. *arXiv preprint arXiv:1901.00373*.

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

Ballester, C., Calvó-Armengol, A., & Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, *74*(5), 1403–1417.

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

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2014). *Gossip: Identifying central individuals in a social network* (Tech. Rep.). National Bureau of Economic Research.

Battaglini, M., & Patacchini, E. (2018). Influencing connected legislators. *Journal of Political Economy*, *126*(6), 2277–2322.

Battaglini, M., Patacchini, E., & Rainone, E. (2019). *Endogenous social connections in legislatures* (Tech. Rep.). National Bureau of Economic Research.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730.

Beresteanu, A., Molchanov, I., & Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica*, *79*(6), 1785–1821.

Berry, S., & Jia, P. (2008). Tracing the woes: An empirical analysis of the airline industry.

Bloch, F., Jackson, M. O., & Tebaldi, P. (2019). Centrality measures in networks. *Available at SSRN 2749124*.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, *92*(5), 1170–1182.

Boucher, V. (2018). Equilibrium homophily in networks. *Available at SSRN 3264071*.

Boucher, V., & Houndetoungan, A. (2019). Estimating peer effects using partial network data. *Draft avaliable at houndetoungan.wixsite.com/aristide/research*.

Bramoullé, Y., & Genicot, G. (2018). Diffusion centrality: Foundations and extensions.

Bykhovskaya, A. (2019). Evolution of networks: Prediction and estimation. *Available at SSRN 3351454*.

Calvó-Armengol, A., & İlkılıç, R. (2009). Pairwise-stability and nash equilibria in network formation. *International Journal of Game Theory*, *38*(1), 51–79.

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

Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford Handbook of the Economics of Networks*, 303–357.

Chandrasekhar, A., & Lewis, R. (2016). Econometrics of sampled networks. *Unpublished manuscript, MIT.[422]*.

Chatterjee, K., & Dutta, B. (2016). Credibility and strategic learning in networks. *International Economic Review*, *57*(3), 759–786.

Chernoff, H., & Teicher, H. (1958). A central limit theorem for sums of interchangeable random variables. *The Annals of Mathematical Statistics*, *29*(1), 118–130.

Christakis, N. A., Fowler, J. H., Imbens, G. W., & Kalyanaraman, K. (2010). *An empirical model for strategic network formation* (Tech. Rep.). National Bureau of Economic Research.

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, *77*(6), 1791–1828.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, *77*(4), 1003–1045.

de Finetti, B. (1929). Funzione caratteristica di un fenomeno aleatorio. In *Atti del congresso internazionale dei matematici: Bologna del 3 al 10 de settembre di 1928* (pp. 179–190).

de Martí, J., & Zenou, Y. (2015). Network games with incomplete information. *Journal of Mathematical Economics*, *61*, 221–240.

de Paula, A. (2017). Econometrics of network models. In *Advances in economics and econometrics: Theory and applications, eleventh world congress* (pp. 268–323).

de Paula, A. (2019). Econometric models of network formation. arXiv:1910.07781.

de Paula, A., Rasul, I., & Souza, P. (2018). Recovering social networks from panel data: identification, simulations and an application.

de Paula, A., Richards-Shubik, S., & Tamer, E. T. (2015). Identification of preferences in network formation games. *Available at SSRN 2577410*.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets* (Vol. 8). Cambridge university press Cambridge.

Echenique, F., Fryer Jr, R. G., & Kaufman, A. (2006). Is school segregation good or bad? *American Economic Review*, *96*(2), 265–269.

Gautier, E., & Rose, C. (2016). *Inference in social effects when the network is sparse and unknown.* preparation.

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

Graham, B. S. (2014). Methods of identification in social networks.

Graham, B. S. (2016). *Homophily and transitivity in dynamic network formation* (Tech. Rep.). National Bureau of Economic Research.

Graham, B. S. (2019). Network data. , *7A*.

Gualdani, C. (2019). An econometric model of network formation with an application to board interlocks between firms. *Forthcoming at Journal of Econometrics*.

Harris, K. M. (2009). The national longitudinal study of adolescent to adult health (add health), waves i & ii, 1994–1996; wave iii, 2001–2002; wave iv, 2007-2009. *Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill*.

Hellmann, T. (2013). On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, *42*(1), 211–237.

Hsieh, C.-S., Lee, L.-F., & Boucher, V. (2019). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Available at SSRN 2967867*.

Huisman, M. (2014). Imputation of missing network data: some simple procedures. *Encyclopedia of Social Network Analysis and Mining*, 707–715.

Jackson, M. O. (2019). The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, *127*(2), 777–818.

Jackson, M. O., et al. (2008). *Social and economic networks* (Vol. 3). Princeton university press Princeton.

Jackson, M. O., & Watts, A. (2001). The existence of pairwise stable networks.

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

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, *76*(6), 1263–1316.

Kaido, H., Molinari, F., & Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, *87*(4), 1397–1432.

Kaido, H., Molinari, F., Stoye, J., & Thirkettle, M. (2017). Calibrated projection in matlab: Users' manual. *arXiv preprint arXiv:1710.09707*.

Kallenberg, O. (2006). *Probabilistic symmetries and invariance principles.* Springer Science & Business Media.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39–43.

Krause, R. W., Huisman, M., Steglich, C., & Sniiders, T. A. (2018). Missing network data a comparison of different imputation methods. In *2018 ieee/acm international conference on advances in social networks analysis and mining (asonam)* (pp. 159–163).

Lee, W., Fosdick, B., & McCormick, T. (2018). Inferring social structure from continuous-time interaction data. *Applied stochastic models in business and industry*, *34*(2), 87–104.

Leung, M. (2015a). A random-field approach to inference in large models of network formation. *Available at SSRN 2520272*.

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

Lewbel, A., Qu, X., & Tang, X. (2019). Social networks with misclassified or unobserved links.

Liu, X. (2013). Estimation of a local-aggregate network model with sampled networks. *Economics Letters*, *118*(1), 243–246.

Manresa, E. (2016). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, 343–360.

Manski, C. F. (2003). *Partial identification of probability distributions.* Springer Science & Business Media.

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

Mele, A., & Zhu, L. (2017). Approximate variational estimation for a model of network formation. *arXiv preprint arXiv:1702.00308*.

Menzel, K. (2015). *Strategic network formation with many agents* (Tech. Rep.). Working papers, NYU.

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

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

Molchanov, I. (2005). *Theory of random sets* (Vol. 19) (No. 2). Springer.

Molchanov, I., & Molinari, F. (2018). *Random sets in econometrics* (Vol. 60). Cambridge University Press.

Molinari, F. (2019). Econometrics with partial identification. *CeMMAP Working Paper CWP25/19*.

Molinari, F., & Rosen, A. (2008). The identification power of equilibrium in games: The supermodular case. *Journal of Business & Economic Statistics*, *26*(3), 297–302.

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of operations research*, *2*(3), 225–229.

Myerson, R. B. (1991). Game theory: analysis of conflict. *The President and Fellows of Harvard College, USA*.

Nishida, M. (2014). Estimating a model of strategic network choice: The convenience-store industry in okinawa. *Marketing Science*, *34*(1), 20–38.

Orbanz, P., & Roy, D. M. (2014). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence*, *37*(2), 437–461.

Ridder, G., & Sheng, S. (2017). *Estimation of large network formation games* (Tech. Rep.). Working papers, UCLA.

Robins, G., Pattison, P., & Woolcock, J. (2004). Missing data in networks: exponential random graph (p*) models for networks with non-respondents. *Social networks*, *26*(3), 257–283.

Rose, C. (2015). *Essays in applied microeconometrics* (Unpublished doctoral dissertation). University of Bristol.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603.

Sheng, S. (2018). *A structural econometric analysis of network formation games* (Tech. Rep.). Conditional Acceptance at Econometrica.

Smith, J. A., & Moody, J. (2013). Structural effects of network sampling coverage i: Nodes missing at random. *Social networks*, *35*(4), 652–668.

Song, Y., & van der Schaar, M. (2015). Dynamic network formation with incomplete information. *Economic Theory*, *59*(2), 301–331.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, *70*(1), 147–165.

Tamer, E. (2010). Partial identification in econometrics. *Annu. Rev. Econ.*, *2*(1), 167–195.

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

Tonelli, L. (1909). Sull'integrazione per parti. *Rend. Acc. Naz. Lincei*, *5*(18), 246–253.

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations research*, *26*(2), 305–321.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

Tsakas, N. (2016a). On decay centrality. *The BE Journal of Theoretical Economics*, *19*(1).

Tsakas, N. (2016b). Optimal influence under observational learning. *Available at SSRN 2449420*.

Uetake, K., & Watanabe, Y. (2012). A note on estimation of two-sided matching models. *Economics Letters*, *116*(3), 535–537.

Zenou, Y. (2016). Key players. *The Oxford Handbook of the Economics of Networks*, *11*.

Appendices

A Outer Region Computation

This section details how to construct the outer identified region for the network-formation parameter, $\mathcal{O}_P[\theta]$, and the joint outer identified region for the target network statistic $\mathcal{O}_P[\beta_\theta, \theta]$. I detail how to construct equivalence classes and also how to simulate the model-implied probability a subnetwork forms.

A.1 Network Formation Moment Inequalities

Recall that the moment inequalities for the network-formation model that I work with are:

$$\begin{aligned} m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) &\equiv -\boldsymbol{P}((\boldsymbol{A}^{\boldsymbol{s}}, \boldsymbol{X}^{\boldsymbol{s}}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) + \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int_{\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})) \\ m_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) &\equiv \boldsymbol{P}((\boldsymbol{A}^{\boldsymbol{s}}, \boldsymbol{X}) \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) - \sum_{\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})} \int_{\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})). \end{aligned} \tag{9}$$

The first term in Equation (9) is estimated from data:

$$\hat{\boldsymbol{P}}(\boldsymbol{A}^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(A^{\boldsymbol{s}} \sim A_t^{\boldsymbol{s}}, \boldsymbol{x} \sim \boldsymbol{x}_t^{\boldsymbol{s}}),$$

where $(A^{\boldsymbol{s}} \sim A_t^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}} \sim \boldsymbol{x}_t^{\boldsymbol{s}})$ denotes an isomorphism. That is, there exists a permutation of $\boldsymbol{s}$, say $\tau(\boldsymbol{s})$, such that $A^{\boldsymbol{s}} = G_t^{\tau(\boldsymbol{s})}$ and $\boldsymbol{x}^{\boldsymbol{s}} = \boldsymbol{x}_t^{\tau(\boldsymbol{s})}$. The first computational problem is checking for isomorphisms, which is expensive. I elect to enumerate all possible values of $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ and construct all equivalence classes outside of the main optimization program. This is feasible provided that I restrict $|\boldsymbol{s}|$ to less than or equal to six. I use the `isisomorphic` function from the Graph and Network Algorithm package in MatLab's base toolbox to check whether two colored networks $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$ and $(\tilde{A}^{\boldsymbol{s}}, \tilde{\boldsymbol{x}^{\boldsymbol{s}}})$ are isomorphic. I compare the realized value of $(A_t^{\boldsymbol{s}}, \boldsymbol{x}_t^{\boldsymbol{s}})$ to the enumerated list of colored networks to obtain its equivalence class number. Last, $\hat{\boldsymbol{P}}(\boldsymbol{A}^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}))$ is computed by summing over all realized networks belonging to class $\mathcal{C}(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$.[21]

[21] An alternative method is to check whether $(A_t^{\boldsymbol{s}}, \boldsymbol{x}_t^{\boldsymbol{s}})$ is isomorphic to the representative colored network in each equivalence class. This alternate method does not work for computing isomorphic dominance, which

The second term in Equation (9) is computed by simulation and from data. The term $\boldsymbol{P}(\boldsymbol{X}^{\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}))$ is estimated from the data. I maintain the assumption that $\boldsymbol{x}_i$ is independent of $\boldsymbol{x}_j$, hence $\boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) = \prod_{i \in \boldsymbol{n}-\boldsymbol{s}} \boldsymbol{P}(\boldsymbol{x}_i)$. In the case of binary data $\mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})$ is characterized by the number of instances that $\boldsymbol{x}_i = 1$. Let $n_1(\boldsymbol{x}^{-\boldsymbol{s}}) = \sum_{i \in \boldsymbol{n}-\boldsymbol{s}} \mathbb{1}(\boldsymbol{x}_i = 1)$. Then

$$\boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} \in \mathcal{C}(\boldsymbol{x}^{-\boldsymbol{s}})) = p^{n_1(\boldsymbol{x}^{-\boldsymbol{s}})}(1-p)^{n-|\boldsymbol{s}|-n_1(\boldsymbol{x}^{-\boldsymbol{s}})},$$

where p is the probability that $\boldsymbol{x}_i = 1$. The sum over equivalence classes is then equivalent to summing over all possible values for $n_1(\boldsymbol{x}^{-\boldsymbol{s}})$, i.e., 1 to $n - |\boldsymbol{s}|$. In a richer setting where $\boldsymbol{x}$ is discrete I can simulate values of $\boldsymbol{x}^{-\boldsymbol{s}}$ from its estimated distribution rather than summing up over all possible values.

The heart of the problem is computing $\int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon)$ and $\int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon)$. I simulate these integrals. First, draw S realizations of ε_s from $F_\varepsilon(\varepsilon; \theta_\varepsilon)$ using an importance sampler or other simulation bias-reducing method. Next, given a pre-defined subnetwork $\boldsymbol{s}$, covariate $\boldsymbol{x}$, and parameter θ, I compute the lattice of admissible networks using Algorithm 1 below, which is based on Proposition 4 and similar to the algorithm in Jia (2008) and Miyauchi (2016). Define $I_1(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) \equiv \int \mathbb{1}[\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon)$ and $I_2(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) \equiv \int \mathbb{1}[\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})] dF_\varepsilon(\varepsilon)$. The respective simulated integrals are:

$$\hat{I}_1(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}(\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \sim A^{\boldsymbol{s}} \sim \overline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta))$$
$$\hat{I}_2(A^{\boldsymbol{s}}, \boldsymbol{x}; \theta) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}(\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \preceq A^{\boldsymbol{s}} \preceq \overline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta)),$$

where the notation $\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \preceq A^{\boldsymbol{s}}$ means that there exists a permutation of $\boldsymbol{s}$, say $\tau(\boldsymbol{s})$, such that $\underline{G}^{\tau(\boldsymbol{s})}(\boldsymbol{x}, \varepsilon_s, \theta) \leq A^{\boldsymbol{s}}$. I define this condition to be *isomorphic dominance*.

Unfortunately there does not exist a function that checks for isomorphic dominance. I use the isomorphic classes and the enumerated list of networks to check whether two networks are isomorphically dominated. The procedure works as follows. Take two equivalence classes $\mathcal{C}(\tilde{A}^{\boldsymbol{s}})$ and $\mathcal{C}(A^{\boldsymbol{s}})$. For all subnetworks $A^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}})$, I check whether $\tilde{A}^{\boldsymbol{s}} \leq A^{\boldsymbol{s}}$. If there exists a $A^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}})$ such that $\tilde{A}^{\boldsymbol{s}} \leq A^{\boldsymbol{s}}$, then all subnetworks in the equivalence class $\mathcal{C}(\tilde{A}^{\boldsymbol{s}})$ are isomorphically dominated by those in $A^{\boldsymbol{s}} \in \mathcal{C}(A^{\boldsymbol{s}})$. I execute this procedure outside of the main algorithm and record in a binary, square matrix with dimension equal to the number of equivalence classes. Using this binary matrix, I can evaluate the term

I will discuss shortly.

$\mathbb{1}(\underline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta) \preceq A^{\boldsymbol{s}} \preceq \overline{G}^{\boldsymbol{s}}(\boldsymbol{x}, \varepsilon_s, \theta))$ inside the main algorithm.

The following algorithm demonstrates how to compute the equilibria lattice.

Algorithm 1 (Admissible Lattice)**.** *The input is* $(\boldsymbol{x}, \varepsilon, \theta)$ *and the output is* $\underline{G}(\boldsymbol{x}, \varepsilon; \theta)$ *and* $\overline{G}(\boldsymbol{x}, \varepsilon; \theta)$*. Execute the following steps.*

0. *Initialize: Create the* $n \times n$ *matrices* G_0 *and* G_1 *as defined in the proof of Proposition 4. Define the function*

$$V_{ij}(G) \equiv \mathbb{1}[\Pi_{ij}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0]\mathbb{1}[\Pi_{ji}(G, \boldsymbol{x}, \varepsilon; \theta) \geq 0].$$

   Set $k = 1$*. Set* $V_{ij}^0(G) \equiv G_0$*.*

1. *Compute* $V_{ij}^k(G_0) \equiv V(V_{ij}^{k-1}(G_0))$*.*
2. *If* $V_{ij}^k(G_0) = V_{ij}^{k-1}(G_0)$*, set*

$$\underline{G}(\boldsymbol{x}, \varepsilon; \theta) = V_{ij}^k(G_0),$$

   set $k = 1$*, and go to step 3. Otherwise, set* $k = k + 1$ *and go to step 1.*

3. *Compute* $V_{ij}^k(G_1) \equiv V(V_{ij}^{k-1}(G_1))$*.*
4. *If* $V_{ij}^k(G_1) = V_{ij}^{k-1}(G_1)$*, set*

$$\overline{G}(\boldsymbol{x}, \varepsilon; \theta) = V_{ij}^k(G_1)),$$

   and return $\underline{G}(\boldsymbol{x}, \varepsilon; \theta)$ *and* $\overline{G}(\boldsymbol{x}, \varepsilon; \theta)$*. Otherwise, set* $k = k + 1$ *and go to step 3.*

A.2 Smoothing Moments

The simulated moments are stepwise and hence the gradient of the moment functions are locally flat. There are many ways of smoothing this problem. I elect to use Kriging. Kriging is an interpolation method that smooths a function $f : \mathbb{R}^Q \to \mathbb{R}^W$ using a Gaussian process governed by prior covariances. I use the DACE MatLab package for this purpose. I first draw M structural network parameters using Latin hypercube sampling. I compute the moment functions using the simulated methods described above for each of these M points. The moments are interpolated using Kriging. Finally, I solve $\min/\max_\theta \theta_j$ subject to the interpolated moment inequalities. The interpolated moment inequalities are smooth and this problem can be solved using any standard gradient based algorithm such as `SQP`. The same Kriging procedure can be applied to get calibrated projected confidence intervals.

B Theorems and Proofs

Proof of Lemma 1. I argue that the solution to

$$\min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G)$$

is equal to $\underline{G}$ when $d(G)$ is monotonically increasing in G. The remaining three cases follow a similar logic. Pick any $G \in \mathcal{L}(\underline{G}, \overline{G})$. By the definition of the lattice, $\underline{G} \leq G$. Moreover, $d(\underline{G}) \leq d(G)$ by the definition of a monotonically increasing statistic. Since G is chosen arbitrarily from $\mathcal{L}(\underline{G}, \overline{G})$ and $\underline{G} \in \mathcal{L}(\underline{G}, \overline{G})$ it follows that $\underline{G} = \arg\min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d(G)$, as required. □

The following Lemma is useful for showing Propositions 1 and 2.

Lemma B1. *Consider the two binary networks G and G'. If $G \leq G'$, then $G^k \leq G'^k$ for all k.*

Proof. I show this by mathematical induction. For the first step

$$\begin{aligned}(G^2)_{ij} &= \sum_{k=1}^{n} G_{ik} G_{kj} \\ &\leq \sum_{k=1}^{n} G'_{ik} G'_{kj} \\ &= (G'^2)_{ij}.\end{aligned}$$

So the base step holds. Now suppose that $G^k \leq G'^k$ holds true and consider the ij^{th} element of G^{k+1}

$$\begin{aligned}(G^{k+1})_{ij} &= \sum_{k=1}^{n} G_{ik} (G^k)_{kj} \\ &\leq \sum_{k=1}^{n} G'_{ik} (G'^k)_{kj} \\ &= (G'^{k+1})_{ij}.\end{aligned}$$

Therefore, the claim holds by mathematical induction. □

Proof of Proposition 1. I show that the centrality measures in Examples 2 - 7 are monotone in the network.

Monotonic Katz-Bonacich Centrality: $d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) \equiv \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w}$. Consider the case when $\boldsymbol{w} \geq 0$. Pick any G, G' such that $G \leq G'$ and consider the i^{th} element of $G^k \boldsymbol{w}$:

$$(G^k \boldsymbol{w})_i = \sum_{j=1}^{n} (G^k)_{ij} w_j \leq \sum_{j=1}^{n} (G'^k)_{ij} w_j = (G'^k \boldsymbol{w})_i.$$

The inequality follows because $w_j \geq 0$ for all j, $G \leq G'$, and by the result in Lemma B1. Thus, for all k, $G^k \boldsymbol{w} \leq G'^k \boldsymbol{w}$. It follows that

$$d^{\text{kbc}}(G; \boldsymbol{w}, \lambda) = \sum_{k=0}^{\infty} \lambda^k G^k \boldsymbol{w} \leq \sum_{k=0}^{\infty} \lambda^k G'^k \boldsymbol{w} = d^{\text{kbc}}(G'; \boldsymbol{w}, \lambda),$$

as required.

Diffusion Centrality: $d_i(G; \lambda, K) = \sum_{k=1}^{K} \sum_{j=1}^{n} \lambda^k G_{ij}^k$. This follows the same argument for KBC – diffusion centrality is equal to a truncated KBC.

Degree Centrality: $d_i(G) = \sum_{j=1}^{n} G_{ij}$. Let $G \leq G'$ and let $\mathcal{N}_i(G) = \{j : G_{ij} = 1\}$. It is clear that $\mathcal{N}_i(G) \subset \mathcal{N}_i(G')$. Hence,

$$d_i(G) = \sum_{j \in \mathcal{N}_i(G)} G_{ij} \leq \sum_{j \in \mathcal{N}_i(G)} G_{ij} + \sum_{j \in \mathcal{N}_i(G') - \mathcal{N}_i(G)} G'_{ij} = d_i(G').$$

Closeness Centrality: $d_i(G) = \frac{n-1}{\sum_{j=1}^{n} \rho_{ij}(G)}$. Recall that $\rho_{ij}(G)$ is the length of the shortest path between individuals i and j in network G. This term does not have an analytical expression. However, it is clear from the definition that $\rho_{ij}(G) \geq \rho_{ij}(G')$ if and only if $G \leq G'$ – adding links can only shorten the shortest path between individuals. Therefore $d_i(G) \leq d_i(G')$.

Harmonic Centrality: $d_i(G) = \sum_{j=1}^{n} \frac{n-1}{\rho_{ij}(G)}$. This follows from the same argument for Closeness Centrality.

Decay Centrality: $d_i(G; \lambda) = \sum_{k=1}^{n-1} \sum_{j=1}^{n} \lambda^k \mathbb{1}(\rho_{ij}(G) = k)$. Consider any two networks satisfying $G \leq G'$. Suppose that $\lambda \in [0, 1]$ and that G is fully connected. Then there is a unique number l_{ij} such that $\mathbb{1}(\rho_{ij}(G) = l_{ij}) = 1$. Adding links can only shorten the shortest path, so there exists another unique number $l'_{ij} \leq l_{ij}$ such that $\mathbb{1}(\rho_{ij}(G) = l'_{ij}) = 1$. Using this notation,

$$d_i(G;\lambda) = \sum_{j=1}^{n} \lambda^{l_{ij}} \mathbb{1}(\rho_{ij}(G) = l_{ij}) \leq \sum_{j=1}^{n} \lambda^{l'_{ij}} \mathbb{1}(\rho_{ij}(G) = l'_{ij}) = d_i(G';\lambda).$$

The inequality follows because $\lambda^{l_{ij}} \leq \lambda^{l'_{ij}}$ when $\lambda \in [0,1]$. The reverse holds true when $\lambda > 1$. The argument is also true for non-connected networks in which case $\mathbb{1}(\rho_{ij}(G) = k) = 0$ for all k for individuals i and j in different communities. □

Proof of Proposition 2. I can re-write the expression for $\underline{d}(\underline{G}, \overline{G}; \lambda)$ as follows:

$$\underline{d}(\underline{G}, \overline{G}; \lambda) = \left[\left(\sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_1, \ldots, \left(\sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_{n_+}, \left(\sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{n_+ + 1}, \ldots, \left(\sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{n} \right].$$

Consider the i^{th} element of $\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}$:

$$\begin{aligned}
&[\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i \\
&= \left\{ \left[\left(\sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_1, \ldots, \left(\sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_{n_+}, \left(\sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{n_+ + 1}, \ldots, \left(\sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{n} \right] \boldsymbol{w} \right\}_i \\
&= \sum_{j=1}^{n_+} \left(\sum_{k=0}^{\infty} \lambda^k \underline{G}^k \right)_{ij} w_j + \sum_{j=n_+ + 1}^{n} \left(\sum_{k=0}^{\infty} \lambda^k \overline{G}^k \right)_{ij} w_j \\
&= \sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (\underline{G}^k)_{ij} w_j + \sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (\overline{G}^k)_{ij} w_j.
\end{aligned}$$

Now consider any $G \in \mathcal{L}(\underline{G}, \overline{G})$. By Lemma B1, for all k: $\underline{G}^k \leq G^k$ and by assumption $w_j \geq 0$ for $j \leq n_+$, hence:

$$\sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (\underline{G}^k)_{ij} w_j \leq \sum_{j=1}^{n_+} \sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j.$$

Similarly, since $w_j < 0$ for $j > n_+$

$$\sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (\overline{G}^k)_{ij} w_j \leq \sum_{j=n_+ + 1}^{n} \sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j.$$

It therefore follows that

$$\begin{aligned}
[\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i &= \sum_{j=1}^{n_+}\sum_{k=0}^{\infty} \lambda^k (\underline{G}^k)_{ij} w_j + \sum_{j=n_++1}^{n}\sum_{k=0}^{\infty} \lambda^k (\overline{G}^k)_{ij} w_j \\
&\leq \sum_{j=1}^{n_+}\sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j + \sum_{j=n_++1}^{n}\sum_{k=0}^{\infty} \lambda^k (G^k)_{ij} w_j \\
&= d_i^{\text{kbc}}(G; \boldsymbol{w}, \lambda)
\end{aligned}$$

Since G and i are arbitrary, it follows that $[\underline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i \leq \arg\min_{G \in \mathcal{L}(\underline{G}, \overline{G})} d_i^{\text{kbc}}(G; \boldsymbol{w}, \lambda)$. A similar argument also shows that $[\overline{d}(\underline{G}, \overline{G}; \lambda)\boldsymbol{w}]_i \geq \arg\max_{G \in \mathcal{L}(\underline{G}, \overline{G})}$. □

Proof of Proposition 3. This follows from the fact that $d_i^{\text{kbc}}(G; \lambda)$ and $(I - \lambda \underline{G})^{-1}_{ii}$ are monotone increasing in G. □

Proof of Proposition 4. Define G_0, G_1 by

$$(G_0)_{ij} \equiv 0 \quad \forall i, j \in \boldsymbol{n}$$
$$(G_1)_{ij} \equiv \begin{cases} 1 & \forall i \neq j \\ 0 & i = j \end{cases}.$$

Apply the mapping $V(\cdot)$ to G_0 k times. Under Assumption 2, $V^k(G_0) \leq V^{k+1}(G_0)$. Hence after a finite $\underline{k}$ number of iterations we obtain $V^{\underline{k}}(G_0) = V^{\underline{k}+1}(G_0)$. Note that $\underline{k} \leq \frac{n(n-1)}{2} - n$, since at most $\frac{n(n-1)}{2} - n$ elements of G_0 can change and at least one element must change each time $V(\cdot)$ is applied before a fixed point is obtained.

It remains to show that $V^{\underline{k}}(G_0) = \underline{G}$. For the purpose of obtaining a contradiction, suppose $V^{\underline{k}}(G_0) \neq \underline{G}$. The network $V^{\underline{k}}(G_0)$ is a fixed point of the mapping $V(\cdot)$, so it is pairwise stable. All pairwise stable networks belong to the admissible lattice, hence $\underline{G} < V^{\underline{k}}(G_0)$. Apply the increasing function $V(\cdot)$ $\underline{k}$ times to both sides of the inequality $G_0 \leq \underline{G}$ to obtain

$$V^{\underline{k}}(G_0) \leq V^{\underline{k}}(\underline{G}) = \underline{G} < V^{\underline{k}}(G_0),$$

a contradiction. Therefore $V^{\underline{k}}(G_0) = \underline{G}$. A symmetric argument can be applied to G_1 to show that there exists a $\overline{k}$ such that $V^{\overline{k}}(G_1) = \overline{G}$.

The term $\underline{k} + \overline{k}$ is bounded by $\frac{n(n-1)}{2} - n + 1$. In the worse case scenario $\underline{G} = \overline{G}$, which requires $\underline{k} = \sum_{i \neq j} \mathbb{1}(\underline{G}_{ij} = 1)$ applications of $V(\cdot)$ to G_0 plus one additional application to ensure that fixed point is obtained, and $\overline{k} = \sum_{i \neq j} \mathbb{1}(\overline{G}_{ij} = 0)$ applications of $V(\cdot)$ to G_1. In this case,

$$\underline{k} + \overline{k} = \sum_{i \neq j} \mathbb{1}(\underline{G}_{ij} = 1) + \sum_{i \neq j} \mathbb{1}(\overline{G}_{ij} = 1) + 1 = \frac{n(n-1)}{2} - n + 1.$$

□

Formal explanation of Remark 2. Let $\overline{D} \equiv \overline{d}(\underline{G}, \overline{G}; \lambda)$. The claim is that there does not always exist a $G \in \mathcal{L}(\underline{G}, \overline{G})$ such that $\overline{D} = (I - \lambda G)^{-1}$, and hence the bounds given in Theorem 2 are not necessarily sharp. For sake of argument, suppose that there did exist a G such that $\overline{D} = (I - \lambda G)^{-1}$. Then

$$\begin{aligned}
&\overline{D} = (I - \lambda G)^{-1} = \sum_{k=0}^{\infty} \lambda^k G^k && \iff \\
&I = (I - \lambda G)\overline{D} && \iff \\
&\boldsymbol{e}_j = \begin{cases} (I - \lambda G)\left(\sum_{k=0}^{\infty} \lambda^k (\overline{G})_j^k\right) & \text{if } j \leq n_+ \\ (I - \lambda G)\left(\sum_{k=0}^{\infty} \lambda^k (\underline{G})_j^k\right) & \text{else} \end{cases},
\end{aligned}$$

where $\boldsymbol{e}_j$ is the j^{th} basis vector. Rewriting this we obtain

$$\begin{aligned}
\sum_{k=0}^{\infty} \lambda^k (\overline{G})_j^k &= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G (\overline{G})_j^k \quad \forall j \leq n_+ \\
\sum_{k=0}^{\infty} \lambda^k (\underline{G})_j^k &= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G (\underline{G})_j^k \quad \forall j > n_+
\end{aligned} \tag{B.1}$$

Setting $G = \overline{G}$ satisfies the first condition in Equation (B.1), since:

$$\begin{aligned} \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G(\overline{G})_j^k &= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} \overline{G}(\overline{G})_j^k \\ &= \boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} (\overline{G})_j^{k+1} \\ &= \boldsymbol{e}_j + \sum_{k=1}^{\infty} \lambda^{k} (\overline{G})_j^{k} \\ &= \sum_{k=0}^{\infty} \lambda^{k} (\overline{G})_j^{k}. \end{aligned}$$

Of course, $G = \overline{G}$ fails the second condition when $j > n_+$ for cases where $\overline{G} \neq \underline{G}$. It follows that $G \neq \overline{G}$ and $G \neq \underline{G}$. I show that no such G exists satisfying Equation (B.1) in general. For this purpose, maintain the following assumptions: $\lambda > 0$ and $\overline{G}$ is fully connected so that for all i, l there exists a k such that $(\overline{G}^k)_{il} \neq 0$.

Since IK know $G \neq \overline{G}$, there exists an i and l such that $0 = G_{il} \neq \overline{G}_{il} = 1$. Fix any $j \leq n_+$ and consider ij component for the expression in Equation (B.1):

$$\begin{aligned} \sum_{k=0}^{\infty} \lambda^k (\overline{G})_j^k &= \left[\boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} G(\overline{G})_j^k \right]_i \\ &= \mathbb{1}(i=j) + \sum_{k=0}^{\infty} \lambda^{k+1} \sum_{l=1}^{n} G_{il} (\overline{G})_{lj}^k \\ &< \mathbb{1}(i=j) + \sum_{k=0}^{\infty} \lambda^{k+1} \sum_{l=1} \overline{G}_{il} (\overline{G})_{lj}^k \quad \text{(since } \lambda > 0,\ G \leq \overline{G},\ G_{il} < \overline{G}_{il}, \text{ and } (\overline{G}^k)_{lj} \neq 0 \text{ for some k)} \\ &= \left[\boldsymbol{e}_j + \sum_{k=0}^{\infty} \lambda^{k+1} (\overline{G})_j^{k+1} \right]_i \\ &= \sum_{k=0}^{\infty} \lambda^k (\overline{G})_j^k. \end{aligned}$$

I arrive to a contradiction and hence I conclude that there does not exist a G such that $\overline{D} = (I - \lambda G)^{-1}$ in general □

Proposition B1. *Consider the strategic campaign donation with peer effects model in Example C1. Suppose that the utility of money is given by $v_i(s) = \kappa \ln(s)$ for all legislators.*

Define mappings $\underline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *and* $\overline{d} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^n$ *by*

$$\underline{d}(\underline{G}, \overline{G}; \lambda, \phi) = \begin{bmatrix} (I - 2\lambda\phi\overline{G})_1^{-1} \\ \vdots \\ (I - 2\lambda\phi\overline{G})_{i-1}^{-1} \\ (I - 2\lambda\phi\underline{G})_i^{-1} \\ (I - 2\lambda\phi\overline{G})_{i+1}^{-1} \\ \vdots \\ (I - 2\lambda\phi\overline{G})_{i+1}^{-1} \end{bmatrix} \quad \textit{and} \quad \overline{d}(\underline{G}, \overline{G}; \lambda, \phi) = \begin{bmatrix} (I - 2\lambda\phi\underline{G})_1^{-1} \\ \vdots \\ (I - 2\lambda\phi\underline{G})_{i-1}^{-1} \\ (I - 2\lambda\phi\overline{G})_i^{-1} \\ (I - 2\lambda\phi\underline{G})_{i+1}^{-1} \\ \vdots \\ (I - 2\lambda\phi\underline{G})_{i+1}^{-1}. \end{bmatrix}$$

Then

$$\frac{(\underline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_i W}{\sum_j (\underline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_j} \leq s_i^* \leq \frac{(\overline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_i W}{\sum_j (\overline{d}(\underline{G}, \overline{G}; \lambda, \phi)\mathbf{1})_j}$$

Proof of Proposition B1. Let $\underline{D} \equiv \underline{d}(\underline{G}, \overline{G}; \lambda, \phi)$ and $\overline{D} \equiv \overline{d}(\underline{G}, \overline{G}; \lambda, \phi)$. The following statements are equivalent.

$$\begin{aligned} &\frac{(\underline{D}\mathbf{1})_i W}{\sum_j (\underline{D}\mathbf{1})_j} \leq s_i^* && \iff \\ &\frac{(\underline{D}\mathbf{1})_i}{\sum_j (\underline{D}\mathbf{1})_j} \leq \frac{d_i^{\text{kbc}}(G'; 2\lambda\phi)W}{\sum_j d_j^{\text{kbc}}(G'; 2\lambda\phi)} && \iff \\ &(\underline{D}\mathbf{1})_i \sum_j d_j^{\text{kbc}}(G'; 2\lambda\phi) \leq d_i^{\text{kbc}}(G'; 2\lambda\phi) \sum_j (\underline{D}\mathbf{1})_j && \iff \\ &(\underline{D}\mathbf{1})_i \sum_{j \neq i} d_j^{\text{kbc}}(G'; 2\lambda\phi) \leq d_i^{\text{kbc}}(G'; 2\lambda\phi) \sum_{j \neq i} (\underline{D}\mathbf{1})_j \end{aligned}$$

This inequality holds, since

$$(\underline{D}\mathbf{1})_i \leq d_i^{\text{kbc}}(\underline{D}; 2\lambda\phi) \leq (\overline{D}\mathbf{1})_i$$

and

$$\forall j \neq i : \quad (\underline{D}\mathbf{1})_j \geq d_j^{\text{kbc}}(\underline{D}; 2\lambda\phi) \geq (\overline{D}\mathbf{1})_j,$$

by construction of $\underline{D}$ and $\overline{D}$. A similar argument can be applied to the upper bound, as required. □

Proof of Proposition 5. The distribution the $(\boldsymbol{A^s}, \boldsymbol{X^s})$ is equal to the following marginalized distribution of $(\boldsymbol{G}, \boldsymbol{X})$

$$\begin{aligned}
\boldsymbol{P}(\boldsymbol{A^s} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x^s}; \theta) &= \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}), \boldsymbol{X} = (\boldsymbol{x^s}, \boldsymbol{x}^{-\boldsymbol{s}}); \theta) \\
&= \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{X} = (\boldsymbol{x^s}, \boldsymbol{x}^{-\boldsymbol{s}}); \theta) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \left(\int \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) dF_\varepsilon(\varepsilon) \right) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \int \left[\sum_{A^{-\boldsymbol{s}} \in \mathcal{G}^{-\boldsymbol{s}}} \psi(\boldsymbol{G} = (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) | \boldsymbol{x}, \varepsilon; \theta) \right] dF_\varepsilon(\varepsilon) \boldsymbol{P}(\boldsymbol{X}^{-\boldsymbol{s}} = \boldsymbol{x}^{-\boldsymbol{s}}),
\end{aligned}$$

where the fourth equality follows from Tonelli's Theorem (Tonelli, 1909). □

Before proving Theorem 1, I require the following two tools from random set theory (Molchanov, 2005; Molchanov & Molinari, 2018; Molinari, 2019).

Definition B1 (Measurable Selection)**.** Let $\mathcal{Z}$ be a random set. A *measurable selection* of $\mathcal{Z}$ is a random element $\boldsymbol{z} \in \mathbb{R}^d$ such that $\boldsymbol{z}(\omega) \in \mathcal{Z}(\omega)$ almost surely. The set of all selections from $\mathcal{Z}$ is denoted $\mathrm{Sel}(\mathcal{Z})$.

The following Lemma is due to (Artstein, 1983).

Lemma B2 (Artstein's Inequality)**.** *A probability distribution μ on $\mathbb{R}^d$ is the distribution of a selection of a random closed set $\mathcal{Z}$ in $\mathbb{R}^d$ iff*

$$\mu(\mathcal{K}) \leq \boldsymbol{T}_{\mathcal{Z}}(\mathcal{K}) \equiv \boldsymbol{P}\{\mathcal{Z} \cap \mathcal{K} \neq \emptyset\}.$$

for all compact sets $\mathcal{K} \subset \mathbb{R}^d$.

Proof of Theorem 1. Under Assumptions 1, the conditional distribution of $G^{\bar{\boldsymbol{n}}}$ for each $G^{\bar{\boldsymbol{n}}} \in \mathcal{G}^{\bar{\boldsymbol{n}}}$ is revealed. The parameter θ is in $\mathcal{H}_P[\theta]$ if and only if the model-implied distribution coupled with the selection mechanism yields the same conditional distribution of $G^{\bar{\boldsymbol{n}}}$ in the data.

Start by considering any $\theta \in \mathcal{H}_P[\theta]$. By the definition of $\mathcal{H}_P[\theta]$, according to Lemma B2 the terms $G^{\bar{\boldsymbol{n}}}$ and $\mathcal{G}^{\mathrm{ps}}_\theta(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ can be realized on the same probability space as the random element $G^{\bar{\boldsymbol{n}}'} =^d G^{\bar{\boldsymbol{n}}}$ and random set $\mathcal{G}^{ps'}_\theta(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) =^d \mathcal{G}^{\mathrm{ps}}_\theta(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ with the property that

$G^{\bar{\boldsymbol{n}}'} \in \text{Sel}(\mathcal{G}_\theta^{ps'}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}))$. Since $\text{Sel}(\mathcal{G}_\theta^{ps'}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}))$ includes all measurable selections, I can choose the one that assigns probability one to $G^{\bar{\boldsymbol{n}}'}$, which is exactly the selection mechanism needed that yields the same distribution observed in the data.

Now suppose that $\theta \in \Theta$ is such that a valid selection mechanism exists with the property that the model-implied distribution coupled with the selection mechanism yields the same distribution of $G^{\bar{\boldsymbol{n}}}$ observed in the data. It follows that there is a selection of $\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})$ with the same distribution as the selection mechanism whose conditional distribution is $\boldsymbol{P}(G^{\bar{\boldsymbol{n}}}|\boldsymbol{x}), \boldsymbol{x} - a.s.$; hence $\theta \in \mathcal{H}_P[\theta]$. □

Proof of Theorem 2. Consider any $\theta \in \mathcal{H}_P[\theta]$, $A^{\boldsymbol{s}}$, and $\boldsymbol{x}^{\boldsymbol{s}}$ such that $|\boldsymbol{s}| \leq q$ and $\boldsymbol{s} \subset \bar{\boldsymbol{n}}$, and fix $\mathcal{K}$ such that $\forall G^{\bar{\boldsymbol{n}}}, \tilde{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K}$ with $A^{\boldsymbol{s}} = \tilde{G}^{\boldsymbol{s}}$, otherwise $G \in \mathcal{K}$ is unrestricted. Let $G^{\bar{\boldsymbol{n}}-\boldsymbol{s}}$ denote the links between individuals in $\boldsymbol{s}$ with those in $\bar{\boldsymbol{n}} - \boldsymbol{s}$. Observe that

$$
\begin{aligned}
\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}} = A^{\boldsymbol{s}}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{s}}; \theta) &= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \sum_{G^{\bar{\boldsymbol{n}}-\boldsymbol{s}} \in \mathcal{G}^{\bar{\boldsymbol{n}}-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} = (A^{\boldsymbol{s}}, G^{\bar{\boldsymbol{n}}-\boldsymbol{s}})|\boldsymbol{x}; \theta)\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}} \in \mathcal{K}|\boldsymbol{x})\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}) \\
&\leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}} \boldsymbol{T}_{\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon)\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}),
\end{aligned}
$$

since $\theta \in \mathcal{H}_P[\theta]$. Since $\mathcal{K}$ imposes no restrictions on G_{ij} with $i \in \boldsymbol{s}$ and $j \in \bar{\boldsymbol{n}} - \boldsymbol{s}$,

$$
\begin{aligned}
\boldsymbol{T}_{\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}})}(\mathcal{K}; F_\varepsilon) &= \boldsymbol{P}(\{\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon, \bar{\boldsymbol{n}}) \cap \mathcal{K} \neq \emptyset\}, \varepsilon \sim F_\varepsilon) \\
&= \boldsymbol{P}(\text{there exists a } A^{-\boldsymbol{s}} \text{ such that } (A^{\boldsymbol{s}}, A^{-\boldsymbol{s}}) \text{ is pairwise stable}, \varepsilon \sim F_\varepsilon) \\
&= \int_{\varepsilon \in \mathcal{E}_u(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta}) \cup \mathcal{E}_m(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon) \\
&\leq \int_{\varepsilon \in \mathcal{E}_a(A^{\boldsymbol{s}}, \boldsymbol{x}; \boldsymbol{\theta})} dF_\varepsilon(\varepsilon).
\end{aligned}
$$

Therefore $m_2(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$ for all $(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}})$. In order to obtain the lower bound, i.e. $m_1(A^{\boldsymbol{s}}, \boldsymbol{x}^{\boldsymbol{s}}; \theta) \leq 0$, I consider the complement of $\mathcal{K}$ and use the containment functional to relate these inequalities to the capacity functional. Defining $\mathcal{K}$ as above,

$$
\begin{aligned}
1-\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}}=A^{\boldsymbol{s}},\boldsymbol{X}=\boldsymbol{x}^{\boldsymbol{s}};\theta) &= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\sum_{G^{\bar{\boldsymbol{n}}-\boldsymbol{s}}\in\mathcal{G}^{\bar{\boldsymbol{n}}-\boldsymbol{s}}}\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}}=(A^{\boldsymbol{s}},G^{\bar{\boldsymbol{n}}-\boldsymbol{s}})|\boldsymbol{x};\theta)\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}})\\
&= 1-\sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}}\in\mathcal{K}|\boldsymbol{x})\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}})\\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\left(1-\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}}\in\mathcal{K}|\boldsymbol{x})\right)\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}})\\
&= \sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\boldsymbol{P}(\boldsymbol{G}^{\bar{\boldsymbol{n}}}\in\mathcal{K}^c|\boldsymbol{x})\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}})\\
&\leq \sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\boldsymbol{T}_{\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})}(\mathcal{K}^c;F_\varepsilon)\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}})\\
&\leq 1-\sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\boldsymbol{P}(\mathcal{G}_\theta^{\mathrm{ps}}(\boldsymbol{x},\varepsilon,\bar{\boldsymbol{n}})\subset\mathcal{K};F_\varepsilon)\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}})\\
&= 1-\sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\boldsymbol{P}\begin{pmatrix}\text{all pairwise stable networks contain}\\ (A^{\boldsymbol{s}},\boldsymbol{x})\text{ as a subnetwork},\varepsilon\sim F_\varepsilon\end{pmatrix}\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}})\\
&= 1-\sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\int_{\varepsilon\in\mathcal{E}_u(A^{\boldsymbol{s}},\boldsymbol{x};\boldsymbol{\theta})}dF_\varepsilon(\varepsilon)\boldsymbol{P}(\boldsymbol{x}^{-\boldsymbol{s}}).
\end{aligned}
$$

It follows that $\boldsymbol{P}(\boldsymbol{A}^{\boldsymbol{s}}=A^{\boldsymbol{s}},\boldsymbol{X}^{\boldsymbol{s}}=\boldsymbol{x}^{\boldsymbol{s}};\theta)\geq\sum_{\boldsymbol{x}^{-\boldsymbol{s}}}\int_{\varepsilon\in\mathcal{E}_u(A^{\boldsymbol{s}},\boldsymbol{x};\boldsymbol{\theta})}dF_\varepsilon(\varepsilon)\boldsymbol{P}(\boldsymbol{x}^{\boldsymbol{s}})$ and hence $m_1(A^{\boldsymbol{s}},\boldsymbol{x}^{\boldsymbol{s}};\theta)\leq 0$. I conclude that $\theta\in\mathcal{O}_P[\theta]$ and so $\mathcal{O}_P[\theta]$ is a valid outer region. □

Proof of Theorem 3. This follows from a similar argument to Theorem 2. The additional step is to show that the network statistic also belongs to the sharp identified set in Equation (5). This follows from the fact that moment inequalities for $\boldsymbol{E}_\theta(d(G)|\boldsymbol{x})$ are constructed from the solutions to Problem 1.

Formally, consider any $(\beta,\theta)\in\mathcal{H}_P[\beta,\theta]$. Following similar logic to Theorem 2, $m_j(A^{\boldsymbol{s}},\boldsymbol{x}^{\boldsymbol{s}};\theta)\leq 0$ for $j=1,2$ for all $(A^{\boldsymbol{s}},\boldsymbol{x}^{\boldsymbol{s}})$. It remains to show that $m_j(\boldsymbol{x};\beta,\theta)\leq 0$ for $j=3,4$. Since

$(\beta, \theta) \in \mathcal{H}_P[\beta, \theta]$ it follows that there exists a selection mechanism $\psi(\cdot)$ such that

$$\begin{aligned}
\beta &= \int_\varepsilon \sum_{G \in \mathcal{G}} d(G)\psi(\boldsymbol{G} = G|\boldsymbol{x}, \varepsilon; \theta) dF_\varepsilon(\varepsilon) \\
&\leq \int_\varepsilon \sum_{G \in \mathcal{G}} \overline{d}(\underline{G}, \overline{G})\psi(\boldsymbol{G} = G|\boldsymbol{x}, \varepsilon; \theta) dF_\varepsilon(\varepsilon) \\
&= \int_\varepsilon \overline{d}(\underline{G}, \overline{G}) \sum_{G \in \mathcal{G}} \psi(\boldsymbol{G} = G|\boldsymbol{x}, \varepsilon; \theta) dF_\varepsilon(\varepsilon) \\
&= \int_\varepsilon \overline{d}(\underline{G}, \overline{G}) dF_\varepsilon(\varepsilon),
\end{aligned}$$

where the last equality follows from Definition 7 where the selection mechanism integrates to one. A similar argument shows that

$$\beta \geq \int_\varepsilon \underline{d}(\underline{G}, \overline{G}) dF_\varepsilon(\varepsilon).$$

Hence $m_j(\boldsymbol{x}; \beta, \theta) \leq 0$ for $j = 3, 4$ and $(\beta, \theta) \in \mathcal{O}_P[\beta, \theta]$, as required. □

Proof of Theorem 4. The proof parallels Theorem 3 with the modification that the network statistics are evaluated at the bounds implied by the subgame admissible lattice. □

C Examples

Example C1 (Strategic Campaign Donations with Peer Effects)**.** *Battaglini and Patacchini (2018) develop a model where legislators vote based on a utility function that is linear in four factors: (1) donations from a special interest group; (2) their friends' voting decision; (3) an idiosyncratic unobserved shock that has bounded support; and (4) her preference over whether the policy is approved or not, which could be negative, positive, or zero. For exposition, I focus on the case where the legislator does not care about whether the policy is approved. There are two special interest groups – one group would like the policy to be approved (group a) and the other would like the policy to not be approved and status quo to hold (group b). The special interest groups each have wealth W to allocate between the n legislators. Group a allocates donations to maximize the probability that the policy is passed conditional on group b's allocation and legislators' voting policy, and group b allocates donations to minimize this probability. The optimal allocation is symmetric and is given by:*

$$\boldsymbol{s}^* = \underset{\boldsymbol{s}\in\mathbb{R}_+}{\arg\max} \sum_{j=1}^{n} d_j^{kbc}(G'; 2\lambda\phi) v_j(s_j) \quad s.t. \quad \sum_{j=1}^{n} s_j \leq W,$$

where s_j is the allocation to legislator j, $v_j(s_j)$ is the utility from donations (factor (1) from above), λ is the social multiplier (factor (2)), and ϕ is the length of the support for the idiosyncratic shock (factor (3)). The function $v_j(\cdot)$ is assumed to satisfy the Inada conditions. For example, if $v_j(s) = \kappa \ln(s)$ for all j, then

$$s_i^* = \frac{d_i^{kbc}(G'; 2\lambda\psi) W}{\sum_j d_j^{kbc}(G'; 2\lambda\phi)},$$

so that legislators with high KBC (i.e., influential legislators) receive a larger share of the donations relative to legislators with low KBC.

Example C2 (Production Equilibrium with an Intersectoral Network, (Acemoglu, Carvalho, Ozdaglar, & Tahbaz-Salehi, 2012))**.** *Consider an economy where each firm or sector i produces good x_i using a technology that use labor input l_i and output of sector j, x_{ij}, as inputs. For example, General Motors uses labor and steel sheets as inputs for the production of automobiles. Let G denote the weighted and directed input-output intersectoral network. Component G_{ij} is positive if and only if sector j is an input supplier to sector i. With Cobb-Douglas preferences and a Cobb-Douglas production function, the logarithm of* *real value added is given by:*

$$\boldsymbol{y} = \boldsymbol{v}'\boldsymbol{e},$$

where e_i i.i.d log productivity shocks and

$$\boldsymbol{v} = \frac{\alpha}{n}(I - (1-\alpha)G')^{-1}\mathbf{1}$$

is the influence vector, which is also the Katz-Bonacich centrality measure with weights $\mathbf{1}$ *and social multiplier* $(1-\alpha)$*, Here* α *is share of labor hired by each sector. The influence vector plays an important roll for the asymptotic distribution of real value added output.*

Example C3. *As an example, suppose* $\boldsymbol{n} = \{1,2,3,4\}$ *and* $\boldsymbol{s} = \{1,2\}$ *and*

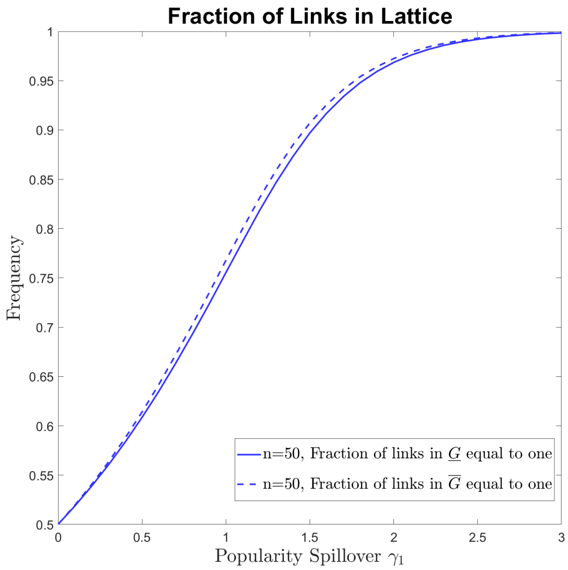
$$G = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

Then

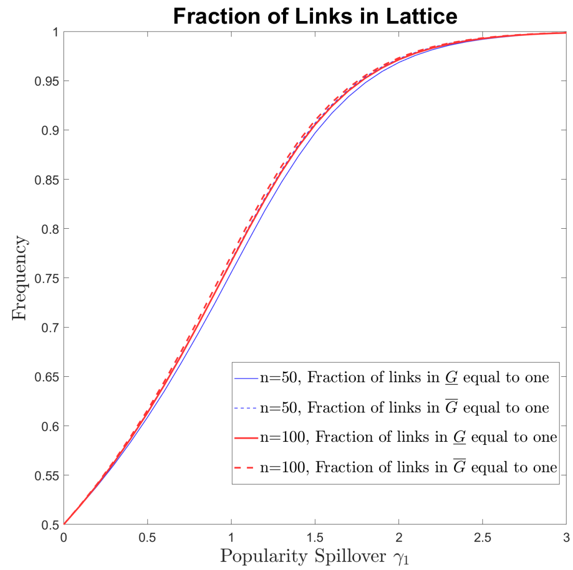
$$A^{\boldsymbol{s}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad A^{-\boldsymbol{s}} = \begin{bmatrix} \cdot & \cdot & 0 & 1 \\ \cdot & \cdot & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

D Simulated Application Results: Figures

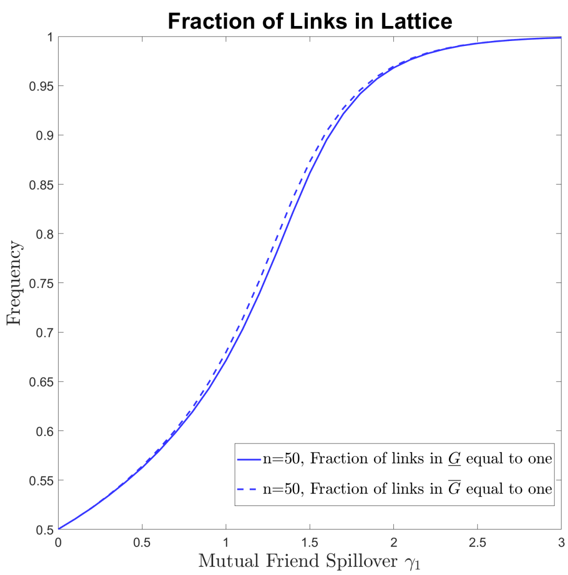
Figure D1: Admissible Lattice Size



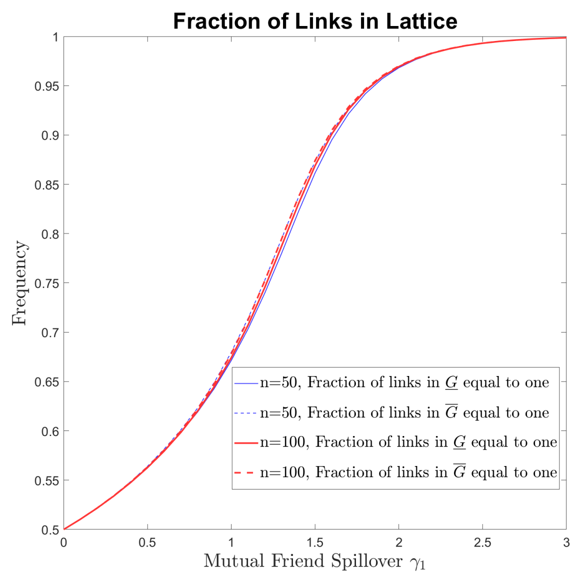
(a) Expected fraction of links in the admissible lattice with popularity spillover only. The dotted line represents the fraction of links in $\overline{G}$ and the solid line represents the fraction of links in $\underline{G}$. The vertical distance between the two is proportional to size of the admissible lattice. These results are reported for $n = 50$.



(b) Expected fraction of links in the admissible lattice with popularity spillover only and $n = 100$.

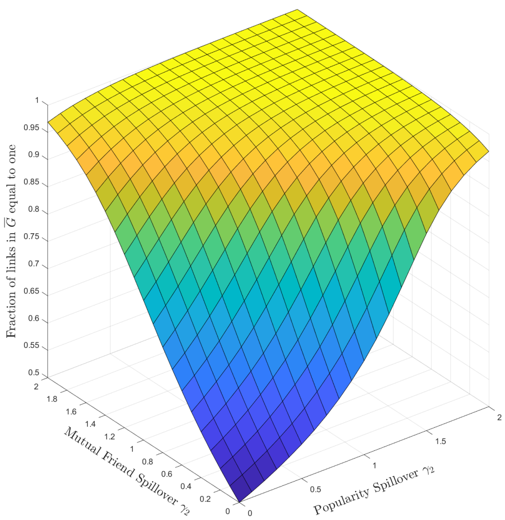


(c) Expected fraction of links in the admissible lattice with mutual friend spillover only. The dotted line represents the fraction of links in $\overline{G}$ and the solid line represents the fraction of links in $\underline{G}$. The vertical distance between the two is proportional to size of the admissible lattice. These results are reported for $n = 50$.

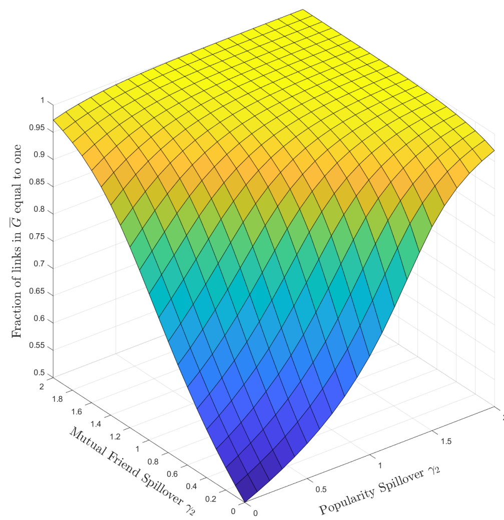


(d) Expected fraction of links in the admissible lattice with mutual friend spillover only and $n = 100$.

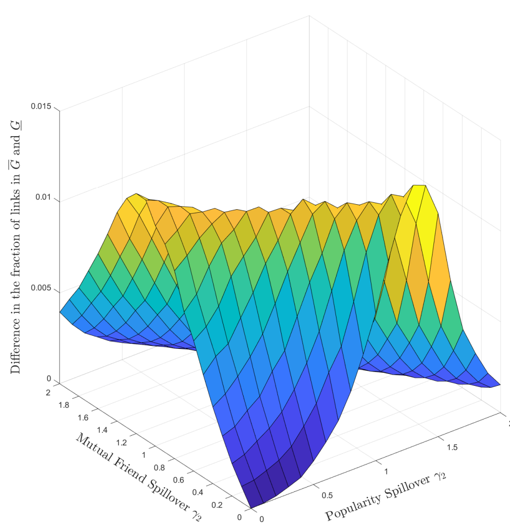
Figure D2: Admissible Lattice Size with Two Channels of Spillovers



(a) Expected fraction of links in for the lower bound of the admissible lattice, $\underline{G}$, with both channels of spillovers. These results are reported for $n = 50$.



(b) Expected fraction of links in for the upper bound of the admissible lattice, $\overline{G}$, with both channels of spillovers. These results are reported for $n = 50$.



(c) Expected fractional difference between the upper and lower bound of the addmissible lattice with both channels of spillovers. These results are reported for $n = 50$.

E Simulated Application Results: Tables

Population Size	**Popularity Spillover**, γ_1	**Max Subnetwork Size**, s $q=2$	$q=3$	$q=4$
50	0	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
	0.5	[0.491,0.512]	[0.491,0.512]	[0.491,0.510]
	1	[0.979,1.018]	[0.984,1.018]	[0.985,1.018]
100	0	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
	0.5	[0.494,0.504]	[0.494,0.503]	[0.500,0.503]
	1	[0.991,1.009]	[0.991,1.006]	[0.992,1.000]

Table E1: Identified set for the popularity spillover with no homophily and one channel of spillovers. The identified set is reported in the square brackets.

Pop Size	**Mutual Friend Spillover,** γ_2	**Max Subnetwork Size** $q=2$	$q=3$	$q=4$
50	0	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
	0.5	[0.495,0.515]	[0.499,0.501]	[0.499,0.500]
	1	[0.988,1.014]	[0.988,1.012]	[0.996,1.012]
100	0	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
	0.5	[0.498,0.503]	[0.500,0.502]	[0.500,0.501]
	1	[0.993,1.006]	[0.994,1.002]	[0.995,1.002]

Table E2: Identified set for the popularity friend spillover with no homophily and one channel of spillovers. The identified set is reported in the square brackets.

λ	**Pop. Spillover**	**Max Subnetwork Size** $q = 2$	$q = 3$	$q = 4$	**True Effort**	**Worst-Case Bounds**
0.01	0	[1.324,1.324]	[1.324,1.324]	[1.324,1.324]	1.324	[1.300,1.356]
	0.5	[1.421,1.426]	[1.421,1.426]	[1.421,1.426]	1.422	[1.383,1.462]
	1	[1.577,1.592]	[1.579,1.592]	[1.579,1.592]	1.580	[1.509,1.626]
0.02	0	[1.972,1.972]	[1.972,1.972]	[1.972,1.972]	1.972	[1.803,2.383]
	0.5	[2.527,2.562]	[2.528,2.560]	[2.528,2.558]	2.528	[2.152,3.190]
	1	[3.991,4.239]	[4.008,4.239]	[4.016,4.239]	4.020	[2.850,5.532]

Table E3: Bounds on KBC using my framework. I restrict to one channel of spillovers (popularity) and set $n = 50$ in each network. I assume that 25 individuals are sampled.

		Max Subnetwork Size		
γ_1	γ_2	$q = 2$	$q = 3$	$q = 4$
0	0	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
0.5	0	[0.000,0.510]	[0.054,0.510]	[0.227,0.510]
0	0.5	[0.000,0.312]	[0.000,0.309]	[0.000,0.0308]
0.5	0.5	[0.000,0.904]	[0.000,0.904]	[0.0233,0.883]

Table E4: Identified region for the popularity spillover γ_1 when both channels of spillovers are present. The popularity spillover is γ_1 and the mutual friend spillover is γ_2. Population size $N = 50$.

		Max Subnetwork Size		
γ_1	γ_2	$q = 2$	$q = 3$	$q = 4$
0	0	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
0.5	0	[0.000,0.768]	[0.000,0.673]	[0.000,0.410]
0	0.5	[0.000,0.502]	[0.000,0.502]	[0.437,0.502]
0.5	0.5	[0.000,1.164]	[0.000,1.144]	[0.0233,1.010]

Table E5: Identified region for the mutual friend spillover γ_2 when both channels of spillovers are present. The popularity spillover is γ_1 and the mutual friend spillover is γ_2. Population size $N = 50$.

Pop. Spillover	Mutual Spillover	Max Subnetwork Size $q = 2$	$q = 3$	$q = 4$	True KBC	Worst-Case Bounds
0	0	[1.336, 1.336]	[1.336, 1.336]	[1.336, 1.336]	1.336	[1.241, 1.473]
0	0.25	[1.363, 1.364]	[1.363, 1.364]	[1.363, 1.364]	1.363	[1.259, 1.497]
0	0.5	[1.396, 1.398]	[1.397, 1.398]	[1.398, 1.398]	1.398	[1.282, 1.526]
0.25	0	[1.386, 1.387]	[1.386, 1.387]	[1.386, 1.387]	1.386	[1.274, 1.517]
0.25	0.5	[1.428, 1.431]	[1.429, 1.431]	[1.429, 1.430]	1.430	[1.304, 1.556]
0.25	1	[1.478, 1.483]	[1.478, 1.482]	[1.479, 1.482]	1.481	[1.335, 1.598]
0.5	0	[1.452, 1.455]	[1.452, 1.455]	[1.452, 1.454]	1.453	[1.319, 1.576]
0.5	0.5	[1.502, 1.507]	[1.502, 1.507]	[1.502, 1.506]	1.504	[1.350, 1.618]
0.5	1	[1.573, 1.580]	[1.573, 1.579]	[1.573, 1.579]	1.576	[1.395, 1.676]

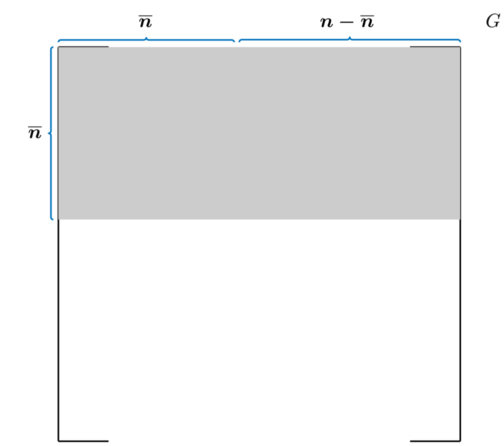
Table E6: Bounds on KBC using my framework. I allow for both channels spillovers (popularity) and set $n = 50$ in each network. There are 25 individuals that are sampled in each network. The decay parameter $\lambda = 0.01$.

Pop. Spillover	Mutual Spillover	Max Subnetwork Size $q = 2$	$q = 3$	$q = 4$	True KBC	Worst-Case Bounds
0	0	[1.336,1.336]	[1.336,1.336]	[1.336,1.336]	1.336	[1.241,1.473]
0	0.25	[1.363,1.364]	[1.363,1.364]	[1.363,1.364]	1.364	[1.259 ,1.497]
0	0.5	[1.396,1.398]	[1.397,1.398]	[1.398,1.398]	1.398	[1.282, 1.526]
0.25	0	[1.386,1.387]	[1.386,1.387]	[1.386,1.387]	1.386	[1.275, 1.517]
0.25	0.25	[2.512,2.535]	[2.519,2.533]	[2.521,2.532]	2.527	[1.905 , 3.623]
0.25	0.5	[2.857,2.895]	[2.860,2.889]	[2.863,2.888]	2.879	[2.053, 4.108]
0.5	0	[2.670,2.695]	[2.670,2.691]	[2.670,2.690]	2.679	[1.975, 3.839]
0.5	0.25	[3.044,3.090]	[3.046,3.087]	[3.048,3.080]	3.065	[2.127, 4.368]
0.5	0.5	[3.718,3.800]	[3.721,3.797]	[3.725,3.792]	3.762	[2.376, 5.328]

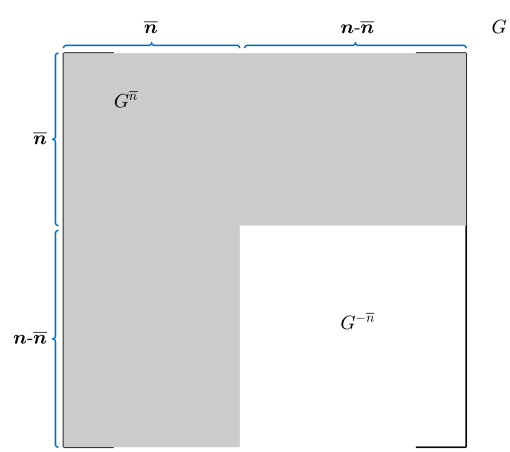
Table E7: Bounds on KBC using my framework. I allow for both channels spillovers (popularity) and set $n = 50$ in each network. There are 25 individuals that are sampled in each network. The decay parameter $\lambda = 0.02$.

F Partially Observed Network and Subnetworks

Figure F1: Partially Observed Network

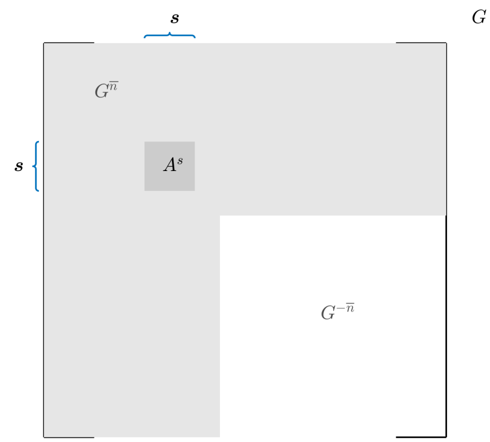


(a) The researcher samples $\bar{\boldsymbol{n}}$ individuals and these individuals reveal all of their direct connections. The revealed connections is represented by the shaded area.

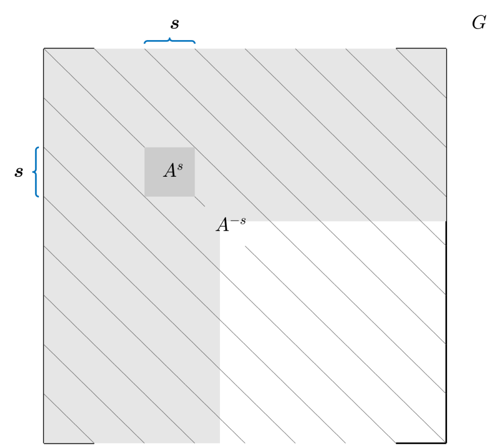


(b) The network is assumed to be symmetric. All links between individuals in $\boldsymbol{n} - \bar{\boldsymbol{n}}$ to $\bar{\boldsymbol{n}}$ are also observed. As a result, the observed network is represented by the shaded area. The observed portion of the network is $G^{\bar{\boldsymbol{n}}}$ and the unobserved portion is $G^{-\bar{\boldsymbol{n}}}$.

Figure F2: Subnetworks and the Partially Observed Network



(a) A subnetwork is constructed from a set of individuals $\boldsymbol{s} \subset \bar{\boldsymbol{n}}$. It contains all links between individuals in $\boldsymbol{s}$. The subnetwork $A^{\boldsymbol{s}}$ is represented by the dark shaded area.



(b) The completion to a subnetwork denoted $A^{-\boldsymbol{s}}$ include all links between individuals not in $A^{\boldsymbol{s}}$. The compleition to the subnetwork is represented by the striped area.

G Table of Notation

Table G1: Table of Notation

$\boldsymbol{n}_t$	Set of individuals in environment t.
$\bar{\boldsymbol{n}}_t$	Set of individuals interviewed about social connections in environment t.
G, G_{ij}	Adjacency matrix where the ij element is equal to one iff i and j are linked.
$A^{\boldsymbol{s}}$, $A^{-\boldsymbol{s}}$	Adjacency matrix for the subnetwork with individuals in $\boldsymbol{s} \subset \{1, \ldots, n\}$ and its completion.
$\boldsymbol{x}$	Network-formation observable characteristic.
ε	Network-formation unobservable characteristic.
$\boldsymbol{G}$, $\boldsymbol{G}^{\bar{\boldsymbol{n}}}$, $\boldsymbol{X}$, $\boldsymbol{A}^{\boldsymbol{s}}$, $\boldsymbol{A}^{-\boldsymbol{s}}$	Random elements of the above realized values.
$G^{\bar{\boldsymbol{n}}}$, $G^{-\bar{\boldsymbol{n}}}$	Partially observed network and its completion.
$G + \{ij\}$, $G - \{ij\}$	Network G with link ij added and deleted, respectively
$\underline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta}), \overline{G}(\boldsymbol{x}, \varepsilon, \boldsymbol{\theta})$	Networks that define the admissible lattice.
θ, θ_0	The network formation parameter and its true value.
$\pi_i(\boldsymbol{g}, \boldsymbol{x}, \varepsilon)$	Agent i's payoff function, where $\boldsymbol{x}$ are observable, ε unobservable.
$\Pi_{ij}(\boldsymbol{g}, \boldsymbol{x}, \varepsilon)$	Agent i's marginal payoff function over link ij.
$\mathcal{G}_\theta^{\text{ps}}(\boldsymbol{x}, \varepsilon)$	Collection of pairwise stable networks.
$\mathcal{G}_\theta(\boldsymbol{x}, \varepsilon)$	Collection of admissible networks.
$\mathcal{G}_\theta(\boldsymbol{x}, G^{\bar{\boldsymbol{n}}}, \varepsilon)$	Collection of conditional admissible networks.
$\mathcal{L}(\underline{G}, \overline{G})$	Generic network lattice containing networks G such that $\underline{G} \leq G \leq \overline{G}$.
$\mathcal{G}$	Collection of undirected networks with no self loops.
$\psi(G \vert \boldsymbol{x}, \varepsilon)$	Selection mechanism for network G.
$d(G)$	Generic network statistic.
$d^{\text{kbc}}(G; \boldsymbol{w}, \lambda)$	Katz-Bonacich Centrality.
$\boldsymbol{P}(\cdot)$	Probability measure.
$m_j(\cdot)$	Moment inequality
P	Joint distribution of observables.
$\mathcal{H}_P[.]$	Sharp identification region for parameter(s) in the square bracket (function of P)
$\mathcal{O}_P[.]$	Outer region for parameter(s) in the square bracket (function of P)
$\mathbb{R}_+^n$, $\mathbb{R}_-^n$	Positive and negative quadrant of the n-dimensional Euclidean space.
$\mathbb{Z}$	Integers.

An *admissible network* is one that is bounded between the smallest and largest pairwise stable network.