A NETWORK FORMATION MODEL BASED ON SUBGRAPHS

ARUN G. CHANDRASEKHAR[‡] AND MATTHEW O. JACKSON[⋆]

Abstract. We develop a new class of random-graph models for the statistical estimation of network formation—subgraph generated models (SUGMs)—that allow for substantial correlation in links. Various subgraphs (e.g., links, triangles, cliques, stars) are generated and their union results in a network. We show that all SUGMs are identified and, further, establish the consistency and asymptotic distribution of parameter estimates in empirically relevant cases. We show that a simple four-parameter SUGM matches basic patterns in observed networks more closely than four standard models (with many more dimensions): (i) stochastic block models; (ii) models with node-level unobserved heterogeneity; (iii) latent space models; (iv) exponential random graphs. We illustrate the framework's value further via several applications using networks from rural India. We study whether network structure helps enforce risk-sharing, whether cross-caste interactions are more likely to be private, and how the introduction of microcredit changes network formation incentives. We also develop a new central limit theorem for correlated random variables, which is required to prove our results and is of independent interest.

JEL Classification Codes: D85, C51, C01, Z13.

Keywords: Subgraphs, Random Networks, Random Graphs, Exponential Random Graph Models, Exponential Family, Social Networks, Network Formation, Consistency, Central Limit Theorem, Sparse Networks, Multiplex, Multigraphs

---

Date: Revision: March 25, 2021.

This grew out of a paper: "Tractable and Consistent Random Graph Models," (http://arxiv.org/abs/1210.7375), which we have now split into two pieces. This part contains the material on subgraph generation models and includes new results on identification, asymptotic normality, and estimation via minimum distance that were not part of the original paper. We thank Alberto Abadie, Isaiah Andrews, Emily Breza, Aureo de Paula, Paul Goldsmith-Pinkham, Bryan Graham, Han Hong, Guido Imbens, Michael Leung, Shane Lubold, Elena Manresa, Tyler McCormick Angelo Mele, Joe Romano, Elie Tamer, and the referees, as well as seminar participants, for helpful comments and suggestions. We thank Shreya Chaturvedi, Vasu Chaudhary, Shobitha Cherian, Andres Drenik, Anoop Singh Rawat, and Meghna Yadav for valuable research assistance. Chandrasekhar is grateful for support from the NSF Graduate Research Fellowship Program, NSF grant SES-1156182, and the Alfred P. Sloan Foundation. Jackson gratefully acknowledges financial support from the NSF under grants SES-1629446 and SES-2018554 and from grant FA9550-12-1-0411 from the AFOSR and DARPA, and ARO MURI award No. W911NF-12-1-0509.

[‡]Department of Economics, Stanford University; member of the NBER; member of J-PAL.

[⋆]Department of Economics, Stanford University; external faculty member of the Santa Fe Institute.

1. Introduction

Networks of interactions impact many economic behaviors including insuring one's self (e.g., Cai, deJanvry, and Sadoulet (2015)), participating in microfinance (e.g., Banerjee et al. (2013)), educating one's self (e.g., Calvo-Armengol, Patacchini, and Zenou (2009); Carrell, Sacerdote, and West (2013)), and engaging in criminal behavior (e.g., Glaeser, Sacerdote, and Scheinkman (1996); Patacchini and Zenou (2008)). Networks of interactions are also essential to understanding financial contagions (e.g., Gai and Kapadia (2010); Elliott, Golub, and Jackson (2014); Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015)), as well as world trade (e.g., Chaney (2016)), inter-state war (e.g., Jackson and Nei (2015); Koenig, Rohner, Thoenig, and Zilibotti (2015)), and a host of other economic phenomena. As such, the structure that a network takes has profound consequences—changing the possibility of contagions, the decisions that people make, and the beliefs that people hold. All of these applications make it essential to understand and be able to estimate network formation.

Moreover, networks are of interest in all such applications precisely because there are externalities—one agent's behavior impacts the welfare and behaviors of others.[1] This feature means that connections between pairs of agents are not independent, not only in determining behaviors but also in network formation. Thus, appropriate models of network formation must admit correlations in connections.

Despite the importance of network formation in such a wide range of social and economic settings, general, flexible, and tractable econometric models for the estimation of network formation are lacking. This stems from two challenges: the aforementioned dependence in connections and the fact that many studies involve one (large) network. Thus, one is often confronted with estimating a model of formation by taking advantage of the large number of connections, but having them all be dependent observations. Despite the dependence, it is possible that the many relationships in a network still provide rich enough information to consistently estimate the parameters of a network model and test of hypotheses from a single observed network, at least hypothetically. Here we develop a class of models that admit correlations in links and also provide practical techniques of estimating the models, showing that they are easily estimable even if a researcher only has one network, as well as in cases with many networks.

Let us discuss some of the other approaches that are available.

1.1. **Existing Models of Network Formation.** The most basic models are what are known as 'stochastic block models', in which links may depend on node characteristics but are (conditionally) independent of each other. That approach requires correlation between links to be well-approximated by observables, and may not be sufficient for most applications.[2] In particular, stochastic block models are not an option for estimation in many economic applications. In fact, in Section 5 we show that our model (even with only four parameters) models the graph structure of real-world data better than a stochastic block model with even

[1]For detailed discussions see Jackson, Rogers, and Zenou (2016) and Jackson (2019).

[2]A variation on this is community detection where nodes are estimated to belong to certain groups, though this calculation is NP-hard. See Bickel et al. (2011) for a "non-parametric view" of network formation from this perspective.

when the block model admits a rich set of covariates and an extension of this to allow for unobserved node level heterogeneity (fixed effects) (Chatterjee et al., 2010; Graham, 2017).[3] Although there are still some challenges in taking such models to data, they have should be useful in settings that in which links are not formed in a correlated manner once one accounts for all observed and (node-level) unobserved characteristics.

Given this void, a literature spanning several disciplines (sociology, statistics, economics, and computer science) turned to exponential random graph models—henceforth "ERGMs"—to meet these challenges. ERGMs admit link interdependencies and have become the workhorse models for estimating network formation.[4] However, from the onset of the use of these models, people realized that the parameter estimates could be very unstable on all except very small networks. It has been shown that maximum likelihood and Bayesian estimators may not be computationally feasible (the required Gibbs sampler will take exponential time to mix) nor consistent for important classes of such models—effectively the ERGMs that include many link dependencies of interest—and neither parameter estimates nor standard errors can be trusted. For details see Bhamidi, Bresler, and Sly (2008); Shalizi and Rinaldo (2012); Chandrasekhar and Jackson (2012).[5]

A set of models that does allow for link dependencies and are estimable are those based on explicit link formation algorithms (e.g., Barabasi and Albert (1999); Jackson and Watts (2001); Jackson and Rogers (2007); Currarini, Jackson, and Pin (2009, 2010); Christakis, Fowler, Imbens, and Kalyanaraman (2010); Bramoullé, Currarini, Jackson, Pin, and Rogers (2012)). These models can be estimated since the algorithms are particular enough so that one can directly derive how parameters in the model translate into aggregate network statistics, such as the degree distribution or homophily levels. The advantage of such models is that a specific algorithm allows for estimation. The disadvantage is that the specificity of the algorithms also necessarily results in narrow models. Thus, these approaches are useful in some contexts, but they are not designed, nor intended, for general statistical testing of a wide variety of network formation models and hypotheses. For instance, such models cannot generate considerable triadic closure (where links correlated across triples of nodes—so if two people have a friend in common, are they more likely to be friends with each other than if link formation were independent).[6]

[3]In fact, correlations can be viewed as driven by unobserved heterogeneity (Chatterjee, Diaconis, and Sly, 2010),which has links be uncorrelated conditional on all (observed and unobserved) characteristics (as extended by Graham (2017)). See also Charbonneau (2017) for related work in a panel data setting. Such models have been studied in the mathematics and statistics literatures (e.g., Holland and Leinhardt (1981); Park and Newman (2004); Blitzstein and Diaconis (2011)).

[4]These grew from work on what were known as Markov models (e.g., Frank and Strauss (1986)) or $p*$ models (e.g., Wasserman and Pattison (1996)). An alternative approach is to work with regression models at the link level, but to allow for dependent error terms, as in the "MRQAP" approach (e.g., see Krackhardt (1988)).

[5]Recent work has made progress on both the speed of convergence of estimation algorithms as well as characterizing the asymptotic distribution of sufficient statistics in some classes of ERGMs that avoid extensive link dependencies (see e.g., Mele (2017a,b); Mele and Zhu (2017)).

[6]The Jackson and Rogers (2007) model does have a parameter that affects triadic closure, but in that model closure cannot be separated from the shape of the degree distribution. So, it is best suited for growing random networks where new nodes are born over time.

Another approach can be thought of as having roots in the spatial statistics literature. Such models organize nodes such that pairs can be evaluated in terms of distance, with linking probabilities decaying in distance. The distance may be latent (unobserved) or in characteristic space (such as geography or demographics). The arguments therein have foundations in the mathematics literature on random geometric graphs (Penrose, 2003)—where nodes are distributed in a latent space according to some Poisson point process and then linking is much more likely among proximate nodes—and has been taken to the statistics literature by work on latent space models such as in Hoff et al. (2002). In such models, the idea is that links between distant enough pairs of nodes are asymptotically (at a fast enough rate) independent and has been adapted to the econometrics literature (e.g., Boucher and Mourifié (2012); Leung (2014)). This approach holds promise for some enormous networks—in which the graph can almost be decomposed into independent pieces.[7]

A curious fact about these latent space (and generally spatial) models is that the geometry of the space the nodes are embedded in dictates and limits strongly the structure of link correlation. To take the simplest possible example, consider a large tree, which is a structure that is certainly of empirical interest in a number of disciplines including economics, computer science, sociology, and biology. The spatial style models, as typically used, cannot generate large trees. This is because an infinite tree cannot be embedded in a finite dimensional Euclidean (or spherical) space.[8] So using these models in a reasonable way may in fact require estimating the unobserved manifold, which presents its own challenges (Lubold et al., 2020). Our model dispenses with these problems in a straightforward way, allowing correlations across nodes but not forcing correlations generated through distances in unobserved or characteristic space. Therefore, we can easily model large trees.

Finally, there is a large literature on the theory of network formation from a strategic perspective (for references, see Jackson (2005, 2008)). Since the first writing of this paper, researchers have started to derive versions of such models that can be taken to data. One approach builds upon the relationship between certain classes of strategic network formation models and potential games (Mele (2017a); Badev (2013); Sheng (2013)). Another derives restrictions on parameters of an observed network under the presumption that it is in equilibrium (pairwise stable) (De Paula, Richards-Shubik, and Tamer (2018)).[9] Although the progress to date requires restrictions on how links can enter agent's payoffs, they provide important first steps in deriving implications of the arsenal of strategic network formation models. Below, we also provide ways to incorporate strategic formation in SUGMs, thus in part bridging our approach here and the strategic formation approach.

1.2. **Our Subgraph Model Approach.** Our approach is distinct from all of the above, both in terms of the fundamentals of the approach (working with subgraphs as the basic building blocks) and the technicalities of allowing nontrivial conditional correlations. We develop a new central limit theorem for non-trivially correlated random variables that moves

[7]McCormick and Zheng (2015) merge the insights from the unobserved heterogeneity and the latent space distance models. Breza, Chandrasekhar, McCormick, and Pan (2017) evaluate its empirical performance.

[8]In fact, these models need to be altered to have distances defined in hyperbolic space instead.

[9]For a recent overview of the recent literature, see de Paula (2015).

away from relying on spatial-style mixing arguments that force decaying dependence in distance. Our contribution is to develop models of network formation that admit considerable and less geometrically restricted interdependency, and have the presence of links be highly correlated – even across distances, but still prove consistency and asymptotic normality of the parameter estimates.

The paucity of flexible models that are computable and can be used across many applications for hypothesis testing and inference is what motivates our work here.

Although the basic ideas behind our models are very simple, we provide four different applications that illustrate how easily such models admit strategic network formation, general covariates, and generate rich network features.

In Section 2 we introduce subgraph generated models (SUGMs), which are a new class of random-graph models for the statistical estimation of network formation that allow for substantial correlation in links. In these models, various subgraphs (e.g., links, triangles, cliques, and stars) are generated directly. For instance, students may form friendships with their roommate(s), members of a study group, teammates, band members, etc.; researchers may form collaborations on writing papers in pairs, or triples, or quadruples, etc; villagers may form specific bilateral or multilateral agreements independently, each to sustain some collection of favors between those individuals involved in the agreement. This results in links and those links are then naturally correlated since they are formed in combinations. The union of all these subgraphs results in a network. In this section, we also introduce four motivating applications to demonstrate how this model could be used: (i) motives for risk-sharing, (ii) descriptively modeling network structure, (iii) incentives to link across social boundaries, and (iv) changes in the incentive to link due to the introduction of microfinance.

The statistical challenge to the researcher is that often only the final network is observed: a survey may ask people to list their friends and acquaintances, or links may be observed on a social platform, or emails or phone calls are observed, and so forth, but the original formation process is often not observed. The challenge that then arises in estimating how the network formed is that subgraphs may overlap and may also incidentally generate new subgraphs, and so the true rate of formation of the subgraphs cannot generally be inferred just by counting their presence in the resulting network.

Despite this, in Section 3 we prove that every subgraph generated model is identified. That is, if we consider a SUGM, a collection of subgraphs that can potentially form together with a set of parameters governing the probabilities of each subgraph forming, any two distinct set of parameters necessarily has two distinct set of distributions over the set of possible networks. Furthermore, we explore specific cases that are of empirical relevance—for instance, links and triangles models—and demonstrate that not only are the distributions generally distinct, but we can find very simple statistics (such as the share of links or triangles that form) that allow us to identify the parameters of interest.

Next we turn to estimation of the parameters in Section 4. We are interested in when we can consistently estimate the parameters and the asymptotic distribution of the estimates so we can conduct inference. There are two situations a researcher may face. In the first case, the researcher may have access to "many networks". This could be because they have collected network data from numerous schools, many villages, or so on. In such a case, we think about the network size as fixed, imagine that the researcher has a large number of independent networks, and demonstrate using standard results that the parameters governing the SUGM can consistently be estimated with MLE estimators that are asymptotically normally distributed. For some empirically relevant classes of models, we demonstrate that there are computationally simple estimators which satisfy consistency and asymptotic normality.

The second case the researcher may be in is where they have one (or just a few) "single large network". This could be because they have collected very rich network data with resource constraints in just a few communities, or because they are looking at a single market, or because they are looking at one social media platform, etc. In this case, the asymptotics are more technically challenging for two reasons. First, for consistency now there will be rate requirements for the parameters in order for information to accumulate and there not to be too much interference across the emergence of distinct subgraphs. Second, existing central limit theorems from the spatial and time-series econometrics literatures do not apply to our setting, as we need to allow subgraphs to form on arbitrary groups of nodes, which then results in correlation patterns across all links in the network. We overcome this problem by developing a new central limit theorem and thereby characterize when certain classes of SUGMs have estimators that are consistent and asymptotically normally distributed.[10]

With the statistical properties established, we turn to our empirical applications in Section 5. In each application we use the detailed network data we collected in 75 villages in Karnataka, India (Banerjee et al., 2014). We begin by comparing SUGMs to four archetypical models from the literature in terms of how well they model real-world data. Specifically, we fit each model to the data and then draw from the distribution at the estimated parameters for each model. We are interested in a variety of economically relevant network features (none of which are directly used to estimate any of the models). We find that across the board a four parameter SUGM outperforms a stochastic block model with flexible covariates, a model of unobserved heterogeneity at the node level as well as rich covariates, a latent space model with unobserved locations and heterogeneity and covariates, and an exponential random graph model with rich covariates. Only the SUGM comes close to capturing the average path length, homophily, maximal eigenvalue, size of the giant component, isolates, and clustering. Having established this, the second example turns to whether the structure of the networks are consistent with the idea that there are stronger incentives to have supported relationships for risk sharing links rather than informational links (Jackson et al., 2012) and we find evidence consistent with this. The third example explores whether linking across social boundaries—here links between upper caste and lower caste (Dalit communities)—is more likely to form in private (bilateral) rather than group (triadic) settings and we find exactly this. Together, these examples demonstrate the utility of our general framework.

[10]An interesting consideration for future work is to employ the techniques in Bhattacharyya et al. (2015), who develop a bootstrapping method to estimate the empirical counts of different subgraphs in enormous networks.

In Section 6 we return to the central limit theorem used prior, which is of independent interest. The standard arguments exploiting strong mixing of random variables (e.g., Bolthausen (1982)) do not apply for SUGMs since there is no sense in which the random variables (nodes, links, triangles) we are concerned with begin to become arbitrarily far from each other, and therefore essentially uncorrelated. Thus, we use a powerful lemma from Stein (1986) in order to prove a new central limit theorem for correlated random variables that provides for more general and permissive results (avoiding a certain kind of geometric restriction) than previously available for our setting.[11] This establishes asymptotic normality for our estimators, and should be useful beyond our network setting. Our results have a connection to the study of central limit theorems for random variables described by dependency graphs (Baldi and Rinott (1989); Goldstein and Rinott (1996); Chen and Shao (2004)), though are less restrictive in the correlation structure of the random variables of interest.

2. A Model of Network Formation via Subgraphs

2.1. **Networks.** $n \geq 3$ is the number of nodes on which a network is formed. Nodes may have characteristics, such as age, profession, gender, race, caste, etc., that we denote by the vector X_i for a generic $i \in \{1, \ldots, n\}$. In what follows we assume that the X_i have finite support. As such nodes can be classified by a finite set of types.[12]

We denote a network by g, the collection of subsets of $\{1, \ldots, n\}$ of size 2 that lists the edges or links that are present in its graph. So, $g = \{\{1, 3\}, \{2, 5\}\}$ indicates the network that has links between nodes 1 and 3 and between nodes 2 and 5. For notational ease, we simply write $g = \{13, 25\}$, and write $ij \in g$ to denote that link ij is present in network g. In general our model easily accommodates directed graphs, and all of the definitions below extend directly, in which case instead of pairs of nodes, these would be ordered pairs so that ij and ji would differ. However, for ease of exposition, most of the examples and discussion refer to the undirected case. $\mathcal{G}^n$ denotes the set of all networks on n nodes.

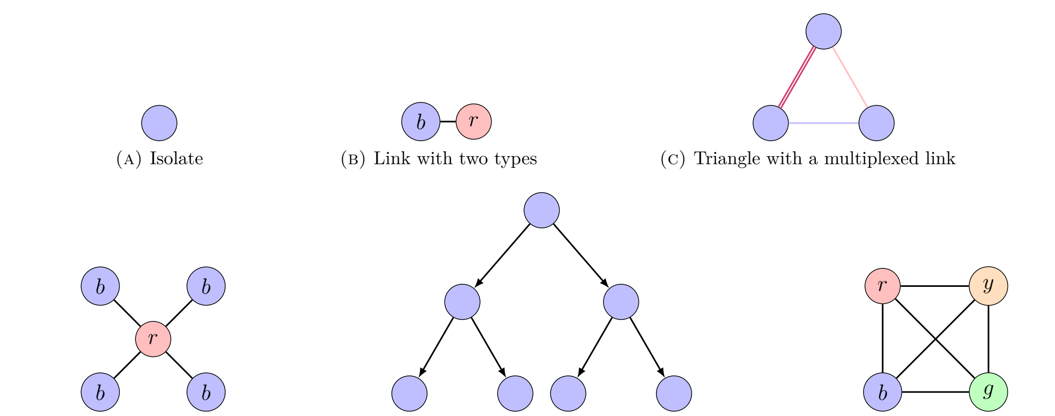
2.2. **Subgraphs and SUGMs.** In a subgraph generation model, henceforth SUGM, subgraphs are directly generated, and then the resulting network is the union of all of the links in all of the subgraphs. Degenerate examples of this are Erdos-Renyi random networks, and the generalization of that model, stochastic-block models, in which links are formed with probabilities based on nodes' attributes. The more interesting classes of SUGMs include richer subgraphs, and hence involve dependencies in link formation. It might be that people of the same caste meet more frequently or are more likely to form a relationship when

[11]This lemma and precursor work in Stein (1972) have been used to derive central limit theorems in two literatures: time-series/spatial statistics and dependency graphs. For instance the oft-used Bolthausen (1982) central limit theorem, crucial in time-series and spatial econometrics, uses a lemma from Stein (1972) to show normality. In time-series and spatial econometrics, a non-exhaustive but illustrative list of papers using Bolthausen (1982) include Conley (1999), Jenish and Prucha (2009), Bester, Conley, and Hansen (2011), among others. However, the arguments of Stein (1972), and therefore Bolthausen (1982), do not apply to our setting in which we need to allow for much richer dependencies than are admitted in previous theorems.

[12]We conjecture that under sensible conditions our results extend to allow for continuous covariates as well, though that requires specifying parametric functions for the probability of subgraphs as a function of covariates and so remains beyond the scope of this paper.

they do meet, as in a stochastic block model, but it could also be that groups of three (or more) meet and can decide whether to form a triangle, with the meeting probability and decision potentially driven by their castes and/or other characteristics. The model can then be described by a list of probabilities, one for each type of subgraph, where subgraphs can be based on the subgraph shape as well as the nodes' characteristics.

SUGMs are formally defined as follows. There are finitely many types of nonempty subgraphs, indexed by $\ell \in \{1, \ldots, k\}$, on which the model is based – for instance in the links and triangles case $\ell \in \{L, T\}$.[13] The k subgraph types are denoted by $(G_\ell)_{\ell \in \{1,\ldots,k\}}$, where each $G_\ell \subset \mathcal{G}^n$ is a set of possible subgraphs on $m_\ell \leq n$ nodes. Each pair of subgraphs in $g' \in G_\ell$ and $g'' \in G_\ell$ are such that there exists a bijection π on $\{1, \ldots, n\}$ for which $ij \in g'$ if and only if $\pi(i)\pi(j) \in g''$. The definitions of the subgraph types can have restrictions based on node characteristics, for instance, requiring that the characteristics X_i and $X_{\pi(i)}$ be the same – e.g., G_ℓ for some ℓ could be the set of "triangles that involve one child and two adult nodes". As an example, the set G_ℓ for some ℓ could be all stars with one central node and four other nodes, and another ℓ could be all of the links that involve people of different castes, and so forth. These could also be directed subgraphs in the case of a directed network. A few examples are pictured in Figure 1.



(A) Isolate

(B) Link with two types

(C) Triangle with a multiplexed link

(D) 4-star with differing types

(E) Tree of all blues

(F) 4-clique with differing types

FIGURE 1. Examples of subgraphs. Links could be directed or undirected or even multiplexed (take on multiple edge types) and nodes can have different characteristic combinations (denoted by node colors and labels).

The probability that various subgraphs form is described by a vector of parameters, denoted $\beta \in \mathcal{B}$, where $\mathcal{B}$ is (unless otherwise noted) a compact subset of $[0,1]^k$. [14] For instance, $\beta = (\beta_L, \beta_T) \in \mathcal{B} \subset [0,1]^2$ in a links and triangles example.[15]

[13]This definition does not admit isolates since we define subgraphs to be nonempty and connected, but isolates are easily admitted with notational complications, and are illustrated in some of our supplementary material and examples.

[14]We treat vectors as row or column vectors as is convenient in what follows.

[15]In some examples below, we expand this demonstrating how β can have entries that are monotone functions of preference parameters (or equilibrium behavior), which allows us to study certain economic questions. Estimating β allows us to either recover the parameters or behavior of interest in some cases or conduct loose hypothesis testing using our estimates of β.

A network g on n nodes is randomly formed as follows:

(1) Each of the possible subnetworks $g_\ell \in G_\ell$ forms independently with probability β_ℓ.
(2) The resulting network, g, is the union of all the links that appear in any of the generated subgraphs.

2.3. **An Example with Node Characteristics.** Suppose that nodes come in two colors: blue and red (for instance different genders, age groups, religions, etc., and clearly this extends directly to more than two colors). In our example of links and triangles, there are now three types of links: (blue, blue), (blue, red), (red, red); and four types of triangles (blue,blue,blue), (blue,blue,red), (blue,red,red), (red,red,red) which comprise the set of subgraphs indexed by ℓ.

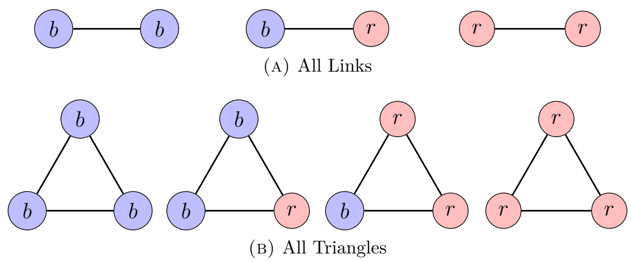


FIGURE 2. Panel (A) shows all possible links and Panel (B) shows all possible triangles when a node has characteristic $X_i \in \{red,\ blue\}$.

Thus, in this example the sets of subgraphs are

$$G_{(blue,blue)} = \{ij : X_i = blue, X_j = blue\}$$

and

$$G_{(blue,blue,red)} = \{ijk : X_i = blue, X_j = blue, X_k = red\},$$

and so forth, as depicted in Figure 2. The parameters

$$\{\beta_{(blue,blue)}, \beta_{(blue,red)}, \beta_{(red,red)}, \beta_{(blue,blue,blue)}, \beta_{(blue,blue,red)}, \beta_{(blue,red,red)}, \beta_{(red,red,red)}\},$$

are the probabilities that the corresponding subgraphs form.

One could restrict or enrich the model by having simpler or more complex sets of parameters – for instance requiring that $\beta_{(blue,blue)} = \beta_{(red,red)}$, or by having preference parameters that govern the probabilities of various subgraphs forming, as we discuss below.

2.4. **Uses of SUGMs.** The SUGM perspective is useful for a number of purposes. First, purely as a statistical modeling tool, simple SUGMs—even ones with just links and triangles—generate higher-order features of empirically observed social networks that link-based models (even those accounting for characteristics, unobserved characteristics, geography, and latent locations) do not. It is important for a formation model to capture realistic features of empirical network data for prediction and also, for example, if a researcher observes only part of a network, having a reasonable model of network formation is important to interpolate over the missing data in a sensible way (e.g., see Chandrasekhar and Lewis (2013) or Breza et al. (2017)). Or, if one is interested in generating networks under a hypothetical policy, a model is only useful if it can generates networks that are likely to occur at a variety parameter values. As we demonstrate, our model outperforms stochastic block models, models with node-level fixed effects, latent space models, and ERGMs in generating realistic distributions of networks even with considerably fewer parameters (e.g., 4 parameter SUGMs versus over 200 (or even 400) parameters in some alternatives).

Second, it can be used to explore incentives for linking in a reduced form way. There are many theories (e.g., Coleman (1988); Jackson, Rodriguez-Barraquer, and Tan (2012)) explaining why triangles and other cliques play special roles in maintaining cooperation in favor exchange. In order to test such theories, we need a statistical model that allows us to test whether cliques appear significantly more often than being randomly generated by links, and whether they appear in configurations that would be predicted by the game theory.

Third, they can be used for structural estimation. There are parsimonious microfoundations – simple models of mutual consent or search – that give rise to SUGMs. Structural parameters may have intrinsic value of their own for welfare analyses, and also aid counterfactuals or policy evaluation. Such parameters may be invertible from SUGM parameters.

We setup four examples and then estimate SUGMs in each of these in Section 5. Our first example shows that SUGMs model myriad network features much better than other models in the literature. The next three examples build, in increasing complexity, models of network formation to address specific economic questions in several domains. They are purposefully nested in structure, with the final model allowing for fully endogenous decisions governing not only mutual consent in linking but also equilibrium efforts invested in generating matching opportunities. In all cases, as will become clear, the equilibrium network is a random draw from a SUGM with interpretable parameters.

2.4.1. *Example 1: Matching Empirical Network Data.* A challenge for network formation models has been to capture more than one or two observed features of social networks at a time. For instance, many observed social networks are sparse but clustered, which motivates developing models that reflect this (Watts and Strogatz, 1998). They also have a variety of differing degree distributions ((Barabasi and Albert, 1999; Jackson and Rogers, 2007) and exhibit high levels of homophily (McPherson, Smith-Lovin, and Cook, 2001; Currarini, Jackson, and Pin, 2009, 2010), which can lead to poverty traps and differences in employment between races (Calvo-Armengol and Jackson, 2007). There are also features such as the expansion properties of a network that are described by maximal eigenvalue of the adjacency matrix speaks to the speed of a diffusion process on the network (Bollobas (2001)). The depth of the max flow min cut speaks to several things such as consensus time in a social learning process Golub and Jackson (2012) as well as the degree of cooperation sustainable (Karlan, Mobius, Rosenblat, and Szeidl, 2009).

We show below that a simple links and triangles SUGM that only has four parameters (estimated from data) captures a number of these features all within a simple model: average distance, the maximal eigenvalue, the cut (homophily), clustering, degrees, among other things) and does so better than a conditional edge independent model (a block model) with numerous parameters that can flexibly depend on a rich set of covariates even when allowing for unobserved heterogeneity for every node, as well as latent space models and also ERGMs.

2.4.2. *Example 2: Do incentives for Risk Sharing Drive Network Formation?*

A model of mutual consent. Consider a simple model in which individuals get utility from being in bilateral relationships ("links") denoted by L, as well as trilateral relationships denoted by T. The value of a partner j to i in a bilateral relationship is a function of their demographics (given by vector X_i) is given by u_i^L:

$$u^L\left(X_i; X_j\right) = X_i'\gamma_{L1} + X_j'\gamma_{L2} + \gamma_{L3}d(X_i, X_j) - \epsilon_{ij},$$

where $d(X_i, X_j)$ is a distance or other metric comparing the demographics - for instance to allow for homophily. Similarly, the value of a triangle of relationships jk to i is given by u_i^T:

$$u^T\left(X_i; X_j, X_k\right) = X_i'\gamma_{T1} + X_j'\gamma_{T2} + X_k'\gamma_{T3} + \gamma_{T4}d(X_i; X_j, X_k) - \epsilon_{ijk}.$$

The value of the relationships depend on the characteristics of the people involved, as well as some idiosyncratic values to the relationships, $-\epsilon_{ij}$ and $-\epsilon_{ijk}$, which may capture personalities, compatibilities, etc., distributed according to some distributions F_L and F_T respectively.

Forming relationships requires mutual consent (e.g., the pairwise stability of Jackson and Wolinsky (1996)), so the net utility must be positive to all agents. The probability that a subgraph ij forms is

$$\beta_L\left(X_{ij}, \gamma_L\right) = F_L\left(X_i'\gamma_{L1} + X_j'\gamma_{L2} + \gamma_{L3}d(X_i, X_j)\right) \times F_L\left(X_j'\gamma_{L1} + X_i'\gamma_{L2} + \gamma_{L3}d(X_j, X_i)\right)$$

and similarly the probability that subgraph ijk forms is

$$\begin{aligned}\beta_T\left(X_{ijk}, \gamma_T\right) = {} & F_T\left(X_i'\gamma_{T1} + X_j'\gamma_{T2} + X_k'\gamma_{T3} + \gamma_{T4}d(X_i; X_j, X_k)\right) \\ & \times F_T\left(X_j'\gamma_{T1} + X_i'\gamma_{T2} + X_k'\gamma_{T3} + \gamma_{T4}d(X_j; X_i, X_k)\right) \\ & \times F_T\left(X_k'\gamma_{T1} + X_i'\gamma_{T2} + X_j'\gamma_{T3} + \gamma_{T4}d(X_k; X_i, X_j)\right).\end{aligned}$$

The products are due to a link requiring two consents and a triangle requiring three consents.

By estimating the probabilities of subgraphs forming ($\beta_T(\cdot)$ and $\beta_L(\cdot)$), under suitable assumptions as described below, one can recover the marginal effects of changes in covariates on preferences for being in various configurations (γ_T and γ_L). Since we have finite support for covariates, we simply call the subgraph formation probabilities β_{T,X_T} and β_{L,X_L} for pair and node covariate combination X_T and X_L respectively. We provide examples below.

Incentives for Risk-Sharing. Jackson, Rodriguez-Barraquer, and Tan (2012) show that whether or not a link is supported plays an important role in maintaining favor exchange. It characterizes renegotiation proof robust pairwise stable networks and shows that, in the homogenous parameter case all such networks are quilts (a union of cliques with no cycle involving more than the minimal clique-size number of nodes), and in the inhomogenous parameter case every link must be supported (if i, j are linked then there exists k such that $g_{ik} = g_{jk} = 1$). Of course, the model of Jackson, Rodriguez-Barraquer, and Tan (2012) is not meant to be taken literally and doesn't lend itself to a simple econometric framework.

Consider a variation on the aforementioned mutual consent model wherein now there are multiple link types: favors and information, and for simplicity we do not consider the interaction of these links. We can use this to study the question raised by Jackson, Rodriguez-Barraquer, and Tan (2012). To make this simple assume there are no covariates, so all nodes are identical. Preferences are described by a random utility framework (McFadden, 1973). In this case the value of a link between i and j to i is given by

$$u_i^{L,favor}(i) = \gamma_{L,favor} - \epsilon_{ij,favor},\ u_i^{L,info}(i) = \gamma_{L,info} - \epsilon_{ij,info}$$

and the value of a triangle is given by

$$u_i^{T,favor}(jk) = \gamma_{T,favor} - \epsilon_{ijk,favor},\ u_i^{T,info}(jk) = \gamma_{T,info} - \epsilon_{ijk,info}.$$

In this case, due to mutual consent, $\beta_{L,favor} = F(\gamma_{L,favor})^2$ and $\beta_{T,favor} = F(\gamma_{T,favor})^3$. It is analogous for information. By the arguments of Jackson, Rodriguez-Barraquer, and Tan (2012), we expect that fraction of links that are supported should be higher in favor exchange than in information links. In the language of this model, with a simple calculation it is easy to see that one expects the following.[16]

Lemma **1.** *Assume the above. Then, a test for* $\frac{\gamma_{T,favor}}{\gamma_{L,favor}} > \frac{\gamma_{T,info}}{\gamma_{L,info}}$ *corresponds to* $\frac{\beta_{T,favor}/\beta_{L,favor}^{3/2}}{\beta_{T,info}/\beta_{L,info}^{3/2}} > 1$.

All proofs are in the Appendices.

Given that triangles can be incidentally generated, one cannot test this simply by examining the ratio of supported links to unsupported ones. If $\gamma_{L,info}$ was very high, then it could be that there are many incidentally generated information triangles, and fewer links remain unsupported. By estimating a link and triangle SUGM, one can estimate the parameters and test this hypothesis, as we do in Section 5.3.

2.4.3. *Example 3: Links across Social Boundaries.* Our next example shows how a SUGM can be used to investigate whether there are norms governing link-formation across different social groups. Identities can lead to strong social norms – prescriptions and proscriptions – concerning interactions across groups. For instance, in much of India there are strong forces that influence if and when individuals can form relationships across castes, particularly among "upper caste" Hindus and the "lower caste" communities, comprised of Dalits (Scheduled castes, SC) and scheduled tribes (ST). The SC and ST communities are those defined by the Indian government as being disadvantaged. This is a fundamental distinction over which the strongest cultural forces are likely to focus. Additional norms are at work with finer caste or subcaste distinctions, but those norms are more varied depending on the particular castes in question while this provides a clear barrier(Munshi and Rosenzweig, 2006).

Among many, one natural question concerns the norms around forming public versus private cross-caste group relationships. Namely, are members of upper and lower caste more likely to form cross-group relationships when those links are unsupported (without any friends in common) compared to when those links are supported with at least one friend in common (and thus have a witness to the relationship)?

To answer this we need models that account for link dependencies; cliques of three or more may exhibit greater adherence to a norm prohibiting certain inter-caste relationships, while the norm may be circumvented in isolated bilateral relationships. We can test whether the relative frequency of triangles compared to links is higher when the relationships are within caste than across caste.

[16]It is without loss of generality to take $F(\gamma) = \gamma$ which is just a bijection and is convenient to work with.

This example is instructive because it is more subtle than that in Section 2.4.2 and it demonstrates that a SUGM can be used for a hypothesis test even when preference parameters per se are not identifiable without overly restrictive assumptions. Consider a process in which individuals may meet in pairs or triples and then decide whether to form a given link or triangle. The link is formed if and only if both individuals prefer to form the link, and a triangle is formed if and only if all three individuals prefer to form it. This minimally complicates an independent-link model enough to require modeling link interdependencies.

Individuals' probabilities to have opportunities to form links or triads can depend of course on the composition of castes of those involved. So let $\pi_L(diff), \pi_L(same)$ denote the probabilities that a given link has an *opportunity* to form (i.e., the pair meets and can choose to form the relationship) that depend on the pair of individuals being of different castes or of the same caste, respectively. Analogously define $\pi_T(diff), \pi_T(same)$. Notice these are unlikely to be observed by the researcher.

As noted above, individual i's utility of having a relationship with j can by influenced by whether they share caste (x_{ij} a dummy variable for same caste) and is given by

$$u_i^L(j) = \alpha_{0,L} + \gamma_{0,L} x_{ij} - \epsilon_{L,ij}$$

and similarly for a triad,

$$u_i^T(jk) = \alpha_{0,L} + \gamma_{0,T} x_{ijk} - \epsilon_{T,i,jk},$$

where x_{ijk} is a dummy for whether all three individuals are members of the same caste.[17] The probability of an individual consenting to a subgraph of type $z \in \{L, T\}$ among the m_z nodes is

$$p_{z,same} = F(\alpha_{0,z} + \gamma_{0,z}) \text{ and } p_{z,diff} = F(\alpha_{0,z}).$$

The hypothesis that we explore is that

$$\frac{p_{T,diff}}{p_{T,same}} < \frac{p_{L,diff}}{p_{L,same}}$$

so that people are more reluctant to involve themselves in cross-caste relationships when those are "public" in the sense that other individuals observe those relationships.

The researcher does not observe either the meeting probabilities nor the probabilities within the mutual consent process. Rather, the researcher observes the compositions β_ℓ for $\ell \in \{L, T\} \times \{same, diff\}$ which are precisely SUGM parameters:

(1) $\beta_{L,same} = p_{L,same}^2 \pi_L(same)$ and $\beta_{L,diff} = p_{L,diff}^2 \pi_L(diff)$, and
(2) $\beta_{T,same} = p_{T,same}^3 \pi_T(same)$ and $\beta_{T,diff} = p_{T,diff}^3 \pi_T(diff)$.

There are two challenges. Recall the difference in the exponents reflects that it is more difficult to get a triangle to form than a link. Hence, to perform a proper test, we have to adjust for the exponents as otherwise we would just uncover a natural bias due to the exponent that would end up favoring cross-caste links. Further, identifying a preference bias is confounded by the meeting bias. Thus, we first model the meeting process $\pi_z(x)$ more explicitly and show that we still have identification as the meeting bias makes triangles relatively more likely to be cross-caste than links.

Consider a meeting process where people spend a fraction f of their time mixing in the community that is predominantly of their own types and a fraction $1 - f$ of their time mixing in the other

[17] This is a simplified model for illustration, but one can clearly consider preferences conditional on any string of covariates. This extends a model such as that of Currarini, Jackson, and Pin (2009, 2010) to allow for additional link dependencies. We could also be interested in higher order relationships.

caste's community. Then at any given snapshot in time, a community would have f of its own types present and $1-f$ of the other type present.[18]

This immediately generates a conservative test in the sense that if we find cross-caste links relatively more likely, that is evidence for a (strong) preference bias.

LEMMA **2.** *A sufficient condition for* $\frac{p_{T,diff}}{p_{T,same}} < \frac{p_{L,diff}}{p_{L,same}}$ *is that* $\frac{\beta_{T,diff}}{\beta_{T,same}} < \left(\frac{\beta_{L,diff}}{\beta_{L,same}}\right)^{3/2}$.

2.4.4. *Example 4: Network Formation and the Impact of Microfinance.* Finally, we provide an example how a SUGM can be used to estimate an equilibrium model of network formation used in prior empirical work (Banerjee et al., 2021), that includes payoff externalities and a global congestion externality. The model is consistent with data that cannot be reconciled by most other models of network formation (see the discussion in Banerjee et al. (2021)). Here we show that the model maps to a SUGM, and how to use link and triangle frequencies to infer equilibrium efforts and structural parameters.[19]

In Banerjee et al. (2021), we studied how the introduction of microfinance to 43 out of 75 villages in rural India changed the networks in those 43 villages. In these communities, prior to the entry of the microfinance institution, there was little formal credit. Therefore, agents relied on each other for informal credit and insurance. Once some villagers gained access to microfinance, their incentives to maintain borrowing and lending relationships with other villagers changed, and this affected the entire equilibrium network structure.

There are two types of households (the nodes) in a village. Node i is either highly likely—denoted by $\theta_i = H$—to join microfinance if available, or not likely—denoted by $\theta_i = N$—to join microfinance if available. For example, H types are those who have a female of eligible age in the household because that is a necessary condition to be eligible, and have other education and financial demographics that make them likely candidates to take out a formal loan. N households are missing some of these key characteristics. The reason for distinguishing between these two types of nodes is because it helps us identify and estimate the externalities in link and triangle formation (which are welfare relevant).

The network formation process has two types of externalities. The first is a *payoff externality*, which is the standard sort of externality: the value of a bilateral relationship ij to i may be different from the value i receives if ij is part of a triad between i, j, k. So, the value of the relationship between i and j could depend on whether they have friend in common (e.g., recall Example 2). This can generate networks with more triadic closure—friends of friends are more likely to be friends—than would appear in a network based only on direct link formation. Indeed, this externality is needed to explain some of the patterns in the networks observed in the villages. However, this externality alone is not enough to account for how the networks change in the villages that obtain microfinance.

In a model with only payoff externalities, the relationships that would be changed due to microfinance would involve people who got loans: so HH or HN links would change more than NN links; and similarly HHH, HHN, and HNN triangles would respond more than NNN triangles. For example, in these villages (where relending turns out to be low), there should be a dissolution of links involving those who get microfinance—links and triangles with Hs—but not those that

[18]Variations on this sort of biased meeting process appear in Currarini, Jackson, and Pin (2009, 2010); Bramoullé, Currarini, Jackson, Pin, and Rogers (2012).

[19]Later, we discuss why the model would satisfy regularity conditions for consistent and asymptotically normally distributed parameter estimates, and illustrate it by fitting it to the data.

exclusively involve Ns, who if anything now have more incentive to borrow from each other. In contrast, however, the data exhibit the opposite. The most impacted relationships are NN links and NNN triads. Nodes that are least exposed to microcredit face the largest effects due to the village's exposure.

This motivates a second externality in the model: a global externality in the network formation. In particular, forming relationships (either links or triangles) requires some efforts of all involved to find each other. If nobody else is socializing, then it is not worthwhile for any given person to socialize. These complementarities generate a global externality. For example in the context of microfinance, once Hs have access to credit, that makes them less likely to want to form relationships with Ns. The fact that Hs are no longer willing to form as many relationships with Ns makes the general value of socializing drop for the Ns. This can — and in fact does — lead to larger drops in links and triangles among NNs and NNNs than any mixture of Hs and Ns.[20]

Let us show that the network formation model from that paper is in fact a SUGM with formation parameters $(\beta_\ell)_{\ell=1}^k$ that are readily interpretable.

Agents care about forming links, triangles, and other cliques. These are groups, for instance, in which people pool risks (ROSCAs) or gather socially or work together, etc. Assume the value of an clique of m_ℓ nodes is $v_\ell(\theta; Z)$ where θ is the node's type and Z denotes the number of the other $m_\ell - 1$ nodes that are of type H. For instance $v_L(H; 0)$ is the base value of a link between an H type and an N type to the H type, and $v_T(N, 2)$ is the base value to an N type of a triad together with two Hs. In this way, payoff externalities are encoded. We note a foundation from Banerjee et al. (2021) below.

When actually meeting some group, the utility that an agent would get from forming that group would be

$$v_\ell(\theta_i; Z) - \varepsilon,$$

where the ε is an idiosyncratic term drawn uniformly on $[0, 1]$, drawn independently for all agents and groups that they meet and have a chance to form.

Taking $v_\ell(\theta_i; Z)$ to be in $[0, 1]$, then the probability that i consents to forming the group is the probability that the base utility $v_\ell(\theta_i; Z)$ exceeds the idiosyncratic term, which given the uniform distribution is simply $v_\ell(\theta_i; Z)$.[21]

In addition to consenting to form a group, the group has to meet and have the potential to form. This depends on the composition of the group and the efforts that each of the agents puts into socializing. The probability of some clique C of m_ℓ agents meeting is given by

$$\phi_\ell \times_{i \in C} e_i.$$

Here ϕ_ℓ denotes the fraction of potential matches that i encounters. Whether $\phi_\ell = 1$ or $\phi_\ell < 1$ depends on application.[22] This could be, for instance, because the agent has a time budget and can only attempt to socialize and invest in relationships a limited number of times. $e_i \in [0, 1]$ is an effort choice, and the groups meet in proportion to agents effort choices.

[20] Moreover, in Banerjee et al. (2021), we also check that this is robust by replicating our empirical findings from Karnataka with an independent randomized controlled trial introducing microcredit in Hyderabad, and we find again that relationships between only N types are at least as likely to dissolve as relationships between Ns and Hs.

[21] More generally, if the errors have some distribution F, then this would just be $F(v_\ell(\theta_i; Z))$, which simply complicates notation, but is an easy extension.

[22] In fact, this rate could depend on n when we consider asymptotics later.

The expected utility for an agent conditional on a group forming is then

$$v_\ell(\theta_i; Z)^+ = \mathrm{E}[v_\ell(\theta_i; Z) - \varepsilon_{i,\ell} \mid v_\ell(\theta_i; Z) - \varepsilon_{i,\ell} \geq 0],$$

and $y_i := 1_{\{\theta_i = H\}}$, n_H the number of H types, and $n_N = n - n_H$ the number of N types. The expected utility as a function of effort for an agent i is then

$$U(\theta_i, e_i) = \sum_{\ell=1}^{h} \sum_{Z=0}^{m_\ell - 1} \underbrace{v_\ell(\theta_i; Z)^+}_{\text{value}} \underbrace{v_\ell(\theta_i; Z)(v_\ell(H; Z-1+y_i))^Z (v_\ell(N; Z+y_i))^{m_\ell - Z - 1}}_{\text{mutual consent}} \underbrace{\binom{n_H - y_i}{Z}\binom{n_N - (1-y_i)}{m_\ell - 1 - Z}}_{\#\text{ of such groups}} \underbrace{\phi_\ell e_i e_H^Z e_N^{m_\ell - 1 - Z}}_{\text{probability of meeting the group}}$$
$$- \underbrace{\frac{1}{2} c e_i^2}_{\text{effort cost}} .$$

The equilibrium efforts by type, $e_H^\star$ and $e_N^\star$, then deliver the probabilities of subgraphs, as summarized in the following lemma (which is direct and so offered without proof).[23]

LEMMA **3.** *Let X denote the number of H types in a clique of m_ℓ nodes. The equilibrium results in a SUGM with parameters*

$$\beta_{\ell,X} = (e_H^\star)^X (e_N^\star)^{m_\ell - X} (v_{l,H;X-1})^X (v_{l,N;X})^{m_\ell - X} \phi_\ell. \tag{2.1}$$

So the model, though complex enough to capture a general equilibrium mechanism of network change in a microcredit setting, reduces to a simple SUGM. The frequencies of subgraphs of various types are of interest both in their own right and in how they map to structural parameters. It is worth noting that the structural parameters, such as the efforts, payoff parameters, and meeting frequencies, can be backed out, as we show below. Certainly that map requires assumptions that may depend on context.

We operationalize the model following Banerjee et al. (2021). There the thought experiment is that nodes may have a need to informally borrow from or an opportunity to informally lend to their network neighbors. This can depend on the type of the node—high or low probability of joining microcredit if it enters the village. The type is estimated via random forest in Banerjee et al. (2021), which we take as a given here.

For an agent i, the need to borrow occurs with probability γ_{θ_i} and the opportunity to lend occurs with probability α_{θ_i}. The return to lending informally is r and the return to borrowing informally is b. In our analysis, $(\alpha_\theta, \gamma_\theta)_{\theta \in \{H,N\}}$ and (r, b) are all taken from data. Nodes can form bilateral and trilateral relationships, corresponding to a links and triangles SUGM.

For this application, with just links and triangles, we simplify the notation of the base utility to a type θ from a link to a type θ' by $v_{\theta;\theta'}$, and correspondingly $v_{\theta;\theta'\theta''}$ for a triangle with types $\theta'\theta''$.

The base payoff for a potential bilateral relationship is

$$v_{\theta_i;\theta_j} = \alpha_{\theta_i}\gamma_{\theta_j} r + \alpha_{\theta_j}\gamma_{\theta_i} b.$$

This leads to a probability of mutual consent of $v_{\theta_i\theta_j} v_{\theta_j\theta_i}$. Then the expected payoff of a formed bilateral relationship is, simplifying our notation $v^+_{\theta;\theta'} := \mathrm{E}[v_{\theta;\theta'} - \epsilon | \epsilon < \alpha_\theta \gamma_{\theta'} r + \alpha_{\theta'}\gamma_\theta b]$.

Turning to the case of a triad, we are motivated by the idea of the value of supported links as in Jackson et al. (2012) and the base payoff is

$$v_{\theta_i;\theta_j\theta_k} = \alpha_{\theta_i}\left(\gamma_{\theta_j} + \gamma_{\theta_k}\right) r + \left(\alpha_{\theta_j} + \alpha_{\theta_k}\right)\gamma_{\theta_i} b + s.$$

[23] For the sake of space, we do not provide conditions for existence of equilibrium or discuss uniqueness, but those issues are well-studied in the literature that considers effort in network formation (e.g., see Cabrales, Calvó-Armengol, and Zenou (2011); Canen, Jackson, and Trebbi (2020)).

Under mutual consent, $v_{\theta_i;\theta_j\theta_k}v_{\theta_j;\theta_i\theta_k}v_{\theta_k;\theta_i\theta_j}$ is the probability of the agents agreeing to form a triad conditional upon meeting. Here s is a support factor. If $s > 0$ it means that the value of a triad is more than the value of simply two links for node i. To see this, it is useful to consider a case where all nodes have the same type. It is clear that if the third node k had no capacity to borrow or lend, we are left in the same case as if i and j are bilaterally linked, but with the support factor s.

In Banerjee et al. (2021) we note that if microfinance crowds-out the value of relationships to Hs by reducing the value of borrowing from others and sharing risks with others, then the value of subgraphs that an H can participate in—$v_{H;H}, v_{H;L}, v_{H;NN}, v_{H;NH}, v_{H;HH}$— decline.

Finally, as noted by Feigenberg et al. (2010), microfinance has the property that it changes the meeting technology directly between H types. Specifically, Hs are required to attend center meetings regularly (e.g., weekly) and as a consequence the latent opportunities of interaction themselves may shift. So when we estimate the model in the collection of villages where microfinance was introduce, we allow for an extra multiplicative factor ζ that scales the meeting opportunities. So, for example, this yields

$$\beta^{MF}_{L,HH} = \left(e^{MF,\star}_{H}\right)^2 \left(v^{MF}_{HH}\right)^2 \zeta^2 \text{ and } \beta^{MF}_{T,HHH} = \left(e^{MF,\star}_{H}\right)^3 \left(v^{MF}_{HHH}\right)^3 \phi_T \zeta^3$$

where we have normalized $\phi_L = 1$.

So, we are interested in estimating $\left(e^{NoMF,\star}_{\theta}, e^{MF,\star}_{\theta}\right)_{\theta\in\{H,N\}}$, and seeing how these change due to the introduction of formal credit. We are also interested in estimating the value of support (s), the rate of change in the meeting technology for those who join microcredit (ζ), and we also recover the relative meeting bias ϕ_T.

The analysis of this SUGM will help us analyze the externalities that are present as a result of microfinance introduction, which end up affecting the whole network even though only some people are exposed to microfinance (see the discussion in Banerjee et al. (2021)). The introduction of microcredit has no direct effect on capacity to lend or need to borrow for N types— $\alpha^{MF}_N = \alpha^{NoMF}_N$ and $\gamma^{MF}_N = \gamma^{NoMF}_N$. However, $\alpha^{MF}_H < \alpha^{No\ MF}_H$ and $\gamma^{MF}_H > \gamma^{No\ MF}_H$, so that microcredit reduces the H types' willingness to consent to relationships with N types. There is also a feedback effect since H types are less likely to consent to relationships, implies that the returns to an N type's effort is lower, so that they invest less in relationship building, which itself has a feedback effect. This can lead Ns lose more relationships than Hs and, in fact, see a large drop in equilibrium effort among Ns, as seen in the data, and as we estimate in more detail below.

2.5. **Links and Triangles as Our Leading Example.** The bulk of our illustrations and applications are based on link and triangle SUGMs, though other subgraphs can be included and are covered by our general results (e.g., Theorems 1, 2, and 3). Our illustrations focus on links and triangles for two reasons: first, this case is simple to understand and illustrates the main points since it exhibits correlated links and incidental generation, second, the link and triangle model already matches the moments that are of interest in many research projects (larger cliques are rare). In fact, as we show below, simply looking at a links and triangle SUGM tagged with whether the nodes involved are homogenous or heterogeneous in demographics (e.g., just a 4 parameter model), replicates real-world network features far better than far richer models. Still, We leave further specification to the researcher as it will depend on their context and the phenomenon being modeled. If there are other the types of subgraphs that are hypothesized to arise in some particular context, then that model can be constructed and estimated in the ways we outline and are covered by our general results.[24]

3. Identification

3.1. **The Challenge of Identification.** The researcher's goal is to use the observed data – from one or more networks – to recover the parameters of interest, for example, the (β_L, β_T) in a SUGM of links and triangles. If the researcher observed the links and triangles that were formed directly, then estimation would be straightforward. Indeed, in some instances a researcher has direct information on all the various groups a given individual is involved in: for instance in the case of a co-authorship network, the researcher may observe all the papers a researcher has written and thus observes papers with two authors, three authors, and so forth. Instead, for instance, it may be that there are groups of three people who commonly share favors and risks together – who really form a triangle, but the researcher only has information from a survey asking which pairs of agents are 'friends' based a survey (as in networks derived from the Add Health data set as in Currarini et al. (2009)), or who borrows from whom and who lends kerosene and rice to whom and other bilateral nominations (as in our Indian village data Banerjee et al. (2013)), or from observing that they are friends on a social platform (as in Facebook network data as in Bailey et al. (2016)), or from observing that two people phone each other or remit payments to each other (as in many such mobile phone data sets Blumenstock et al. (2011)).

Thus, the general problem is that the formation of the subgraphs is not directly observed, and so must be inferred in order to estimate the parameters of interest. The observed network g is a projection of all directly generated links and triangles on top of each other. For example, if three links are observed between i, j, and k, is it the case that ijk formed as a triangle, or that ij, jk and ik formed as links, or that ij and jk formed as links and ik formed as part of a different triangle ikm, or some combination of these or other combinations? Figure 3 provides an illustration.

This presents a challenge for estimating a parameter related to triangle formation since some of the observed triangles were *"directly generated"* in the formation process, and others were *"incidentally generated;"* and similarly, it presents a challenge to estimating a parameter for link formation since some truly generated links end up as parts of triangles.[25]

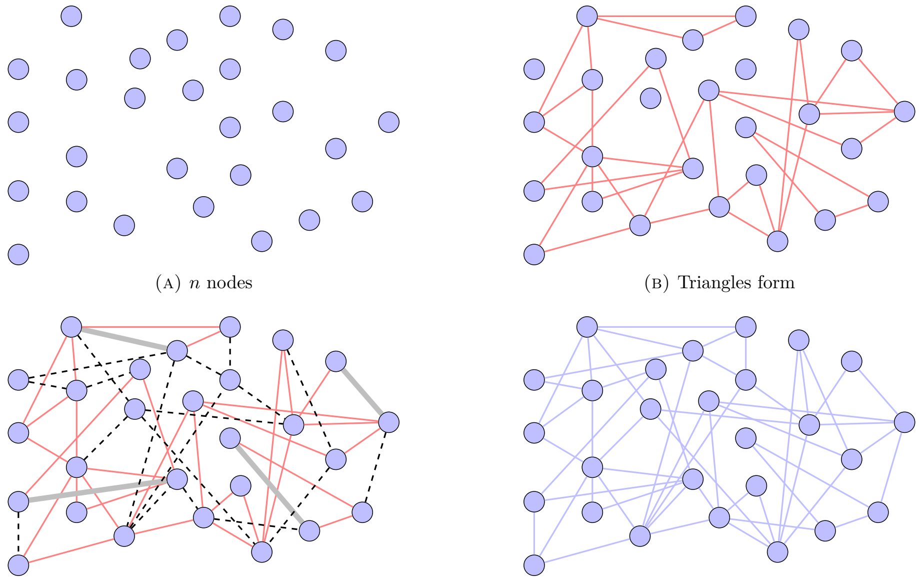
3.2. **A General Identification Result.** We first show that as the parameters of any SUGM change, so does the distribution over networks, and hence SUGMs are well-identified models.

Let P_β denote the probability distribution over a network g on n nodes under a vector of parameters β describing the probabilities of subgraph types $(G_\ell)_{\ell\in\{1,\dots,k\}}$.

Theorem **1.** *Every SUGM is identified. That is, for any finite collection of distinct types of subgraphs* $(G_\ell)_{\ell\in\{1,\dots,k\}}$ *on* n *nodes,* $\beta \neq \beta' \implies \mathrm{P}_\beta \neq \mathrm{P}_{\beta'}$.

[24] One could also have a list of subgraphs as a possible basis for the SUGM with only subset of them actually forming the true SUGM; allowing the data to tell the researcher which to include. Some of that can be done here, including the various subgraphs that might be involved and then seeing which have nontrivial parameter estimates. This marries SUGMs with model selection, a topic which could be explored further in future research.

[25] One could view this as an example of measurement error with correlation: which parts in the resulting observed graph are direct versus incidental is unobserved. The observed graph, which is a projection, and provides a count of observable subgraphs of various types, could be viewed as a mismeasurement of the list of subgraphs directly generated by the SUGM process.



(A) n nodes

(B) Triangles form

(C) Links form

(D) Resulting network

FIGURE 3. The network that is formed and eventually observed is shown in panel D. The process comes from forming triangles independently with probability β_T as in (B) in red; and also forming links, in grey, independently with probability β_L as in (C). New links are dashed while links that overlap with some link also formed in a triangle are in solid and bold. We see that there is both (i) overlap as some links coincide with links already in triangles, as well as (ii) extra triangles that were generated 'incidentally'. Given that we only observe the resulting network in panel D, we need to infer the formation of the different subgraphs carefully and not simply by directly counting observed links and triangles.

Recalling the general definition of the SUGM, this means that for any SUGM, even one comprised of subgraphs that could have nodes with varying (discrete) covariates and allowing for multiplexing, we have identification.

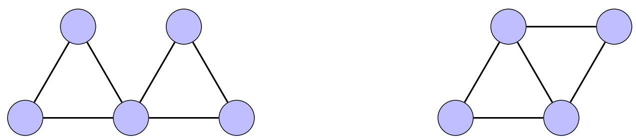
To understand why this holds, for instance in the case of links and triangles, note that as one varies (β_L, β_T), the *relative* rates of overall observed links and triangles change, as do the number of triangles that overlap with each other. One can calculate the relative rates at which incidental links and triangles are expected to be generated, and there is an invertible relationship between observed counts of links and triangles, and the underlying rates at which they were expected to be directly formed. Theorem 1 shows that this is true not only for links and triangles, but for any collection of distinct subgraphs.

We emphasize, of course, that identification does not imply that the parameters will be consistently estimated. For instance, there is always a chance that no subgraphs form, especially on a very small number of nodes. We provide results on consistency below, which require observation of a sufficiently large network and/or sufficiently many networks.

3.2.1. *Identification from Link and Triangle Counts.* Although Theorem 1 shows that SUGMs are always well-identified — i.e., distinct parameters yield distinct distributions — it is often convenient to use minimum distance based estimators based on simple moments of the network. We illustrate that this can be done with direct counts of the relative frequency of appearances of the subgraphs. In Theorem 3 below, we show that such direct counts not only identify the parameters for general subgraphs, but are also consistent and normally distributed estimators of the parameters, under suitable conditions.

In particular, here we show that a links and triangles SUGM can be identified directly from the counts of links and triangles: $S(g) = (S_L(g), S_T(g))$.

To understand the identification, consider Figure 4. Each configuration involves two triangles, but the graph in Panel B with only five links is *relatively* more easily incidentally formed than the one in Panel A. Thus, by looking at the combination of how many triangles and how likely links there are, we can sort out relative rates of the two parameters.



(A) Node adjacent triangles (B) Edge adjacent triangles

FIGURE 4. Two different configurations of two triangles - one has a count of 6 total links and the other has a count of 5 links. (A) is more relatively more likely to come directly from the formation of two triangles, and (B) is relatively more likely to come from a combination of links and triangles. The likelihoods of links and triangles can thus be deduced via careful deductions from the combination of the counts of links and triangles.

PROPOSITION **1.** *A SUGM of links and triangles is identified with moments* $S(g) = (S_L(g), S_T(g))$ *for any* $\beta = (\beta_L, \beta_T) \in (0,1)^2$. *That is, if* $(\beta'_L, \beta'_T) \neq (\beta_L, \beta_T)$ *then* $\mathrm{E}_{\beta'}[S(g)] \neq \mathrm{E}_{\beta}[S(g)]$.

Let us outline the basic ideas behind the proof, with the full proof appearing in the appendix. Let $\widetilde{q}_L$ denote the probability that a link forms conditional upon exactly one particular triangle that it could be a part of not forming. For instance, for nodes ij it is the probability that ij is formed either as a link or as part of a triangle that is not triangle hij for some other node h. In this case:

$$\mathrm{E}_{\beta_L,\beta_T}[S_L(g), S_T(g)] = \left[\beta_T + (1-\beta_T)\widetilde{q}_L, \beta_T + (1-\beta_T)(\widetilde{q}_L)^3\right]. \tag{3.1}$$

For instance, note that the term $\beta_T + (1-\beta_T)(\widetilde{q}_L)^3$ is the probability that a triangle forms, either directly (β_T), or does not form directly $(1-\beta_T)$ but then each of the links then forms on its own $(\widetilde{q}_L)^3$.[26] The term for the links is similar as it could form if some particular triangle forms, or else if that triangle does not form then it forms with probability $\widetilde{q}_L$. Although there are more direct ways to write the probability of a link forming, this particular expression is useful in the proof since it is easy to compare it to and this distinguish it from the triangle expression, as they are identical except for the exponent. This is very helpful in showing how different parameters lead to different rates of formation of links and triangles since we can isolate the difference via the $\widetilde{q}_L$ versus $(\widetilde{q}_L)^3$ expressions.

Analogs of this proposition extend to cases with covariates and multiplexing, simply with more complicated extensions of (3.1) accounting for the specific types of triangles or links needed to

[26]Conditional upon the triangle not forming directly, the links are then independent.

incidentally generate any given link or part of a triangle. Also, a general version of asymptotic identification is a by-product of Theorem 3, below.

4. Asymptotics

4.1. **Data and Asymptotic Frames.** In what follows, we assume the researcher observes $R \geq 1$ independently, and identically drawn graphs $(g_1, \ldots, g_R)$, on at least n nodes each,[27] drawn from a SUGM with a list of k subgraphs and parameters $\beta \in [0,1]^k$. Each of the k subgraphs involves no more than n nodes.

We consider two asymptotic frames, in which the size of the network and/or the number of networks become large enough for consistent estimation.

In the first frame $R \to \infty$. Here researchers have access to many networks and the empirical moments of interest converge to their expectations via observation of independent networks. This applies when a researcher is studying, for instance a number of schools, classrooms, villages, etc.

In the second frame $n \to \infty$, with the possibility (but not the requirement) that $R = 1$. That is, a researcher observes at least one network that is large. Examples include when the researcher has detailed information about a large community, friendships on social media platform, citation networks, etc. Clearly, this also extends to cover cases with large n and more than one network.

We let β depend on n and/or R as described in Section 4. We take the list of the types of subgraphs to be analyzed to be fixed.

We now provide conditions under which estimators of the parameters are consistent and describe their asymptotic distributions. Recall that we consider two frames. The first, studied in Section 4.2, holds the number of nodes, n, fixed, and allows the number of different realizations of networks R to tend to infinity. In this case estimation and inference is straightforward. There are a growing number of independent draws from the distribution and we have already in Theorem 1 proved identification. Consistency and asymptotic normality of the MLE then follow from standard arguments.

The second perspective, studied in Section 4.3, holds the number of networks observed R fixed, usually at $R = 1$, and then lets the number of nodes grow: $n \to \infty$. This is the more challenging perspective as the observations of various parts of a network are not independent. Also, the identification result from Theorem 1 does not guarantee that empirical moments converge to their expectations in a single large network. So, prove that identification is possible from simple counts of observed subgraphs (with a probability approaching one as the network grows), together with consistency and asymptotic normality. Appendix D provides simulations demonstrating consistency, that the asymptotic distribution of the parameter estimates are normal, and coverage properties.

Since in a typical SUGM the links may all be correlated, we also prove in Section 6 a new central limit theorem for correlated random variables that do not satisfy the standard mixing conditions used in time series and spatial econometrics. We use this in service of our results in Section 4.3. The central limit theorem and technique should be of interest beyond network models.

4.2. **The Many Networks Case.** We provide just a brief presentation of this first perspective since it follows standard statistical arguments (e.g., Newey and McFadden (1994)).

One has a collection of R networks, each drawn independently according to a SUGM with the same parameter β_0. We hold n fixed.

[27]For simplicity in notation, we work with each network having exactly n nodes, but one can directly extend the results by simply selecting n nodes for each network and applying all of our estimation to those subgraphs.

Theorem 2 states that a maximum likelihood estimator of the parameters is consistent and asymptotically normally distributed.

THEOREM **2.** *Consider a SUGM of k distinct types of subgraphs with $\beta_0 \in int(\mathcal{B})$, for $\mathcal{B}$ a compact subset of $(0,1)^k$. Let g_r for $r = 1, \ldots, R$ denote i.i.d. draws from this distribution. Let $\hat{\beta}$ denote the maximum likelihood estimator $\hat{\beta} = \mathrm{argmax}_{\beta \in \mathcal{B}} \frac{1}{R} \sum_r \log \mathrm{P}_\beta(g_r)$. Then $\hat{\beta} \xrightarrow{\mathrm{P}} \beta_0$. If in addition $J := \mathrm{E}[\nabla_\beta \log \mathrm{P}_{\beta_0}(g_r) \nabla_\beta \log \mathrm{P}_{\beta_0}(g_r)']$ is non-singular, then $\sqrt{R}\left(\hat{\beta} - \beta_0\right) \rightsquigarrow \mathcal{N}\left(0, J^{-1}\right)$.*

Although Theorem 2 demonstrates that a consistent and asymptotically normally distributed estimator exists, calculating the likelihood function of arbitrary networks as a function of the parameters can be computationally intensive for large networks. Thus, we also present a result on our minimum distance estimator which is computationally very straightforward since it simply involves calculating frequencies of certain subgraphs. We present it based on links and triangles as the typical case that researchers will need, but the technique can be extended as a researcher requires. As before, let $S_L(g)$ and $S_T(g)$ denote the fraction of links and triangles in the network g, with $S = (S_L, S_T)'$.

PROPOSITION **2.** *Consider a SUGM of links and triangles with $\beta_0 \in \mathcal{B}$, a compact subset of $(0,1)^2$. Let g_r for $r = 1, \ldots, R$ denote i.i.d. draws from this distribution. Let $\hat{\beta}$ denote the minimum distance estimator:*

$$\hat{\beta} := \underset{\beta \in \mathcal{B}}{\mathrm{argmin}} \left\| \frac{1}{R} \sum_r S(g_r) - \mathrm{E}_\beta \left[S\left(g\right) \right] \right\|^2 .$$

Then,

$$\hat{\beta} \xrightarrow{\mathrm{P}} \beta_0 \text{ and } \sqrt{R}\left(\hat{\beta} - \beta_0\right) \rightsquigarrow \mathcal{N}\left(0, J^{-1}\right)$$

where $H := \nabla_\beta \mathrm{E}_{\beta_0}[S(g_r)]$ and $J := HH'$.

4.3. **The Large Network Case.** Next we turn to a case where researchers have access a large network: $n \to \infty$. For the exposition, we let $R = 1$, but clearly this extends directly to having observations more than one network.

This case is considerably more challenging as it involves correlated observations generated within a network. Network data tend to be sparse, but still have local patterns such as clustering, so that people have relatively few connections compared to the potential number of links, but where one's neighbors tend to be linked to each other with much higher than an independent probability (e.g., see the background in Newman (2003); Jackson (2008)). Such clustering is the challenging aspect of the asymptotics since subgraphs are not only the directly generated subgraphs of various types but also include incidentally generated features. Thus, we have to provide new techniques for our asymptotic results.

4.3.1. *Sequences of Large Random Networks.* To describe how parameter estimates behave as a function of the number of nodes n, is useful to consider a sequence of distributions governed by parameters indexed by n and study the asymptotic behavior of estimators of parameters along the sequence. This approach is standard in the random graphs literature (e.g., see the classic book of Bollobas (2001)). Research on social networks has long observed that parameters need to adjust with the number of nodes. For example, friendship networks among a small set of agents (say 50 or 100) and large set of agents (thousands or much more) often have comparable average degrees.[28]

[28]See Chandrasekhar (2015) for examples networks of varying size ranging from village network data in sub-saharan Africa or India to university dorm friendship network data which all exhibit somewhat comparable number of links per node.

As a concrete example, consider friendships among high school students in the U.S. based on the Add Health data set (e.g., see (Currarini, Jackson, and Pin, 2009, 2010)). There are some high schools with only 30 students and others with around 3000 students. The average degree is ranges between 6 and 8 over the high schools, the link probability shrinks dramatically with n: roughly $d/30$ to roughly $d/3000$, where d is the average degree in the 6 to 8 range. Thus, irrespective of the size of their school, students have numbers of friends of the same order of magnitude; and the frequency of friendship formation must decrease with n.

So we sequence the parameters β^n. And, thus, we consider a sequence of SUGMs with subgraphs $(G_1, ..., G_k)$ that form on n nodes that are generated with probabilities $\beta^n = (\beta_1^n, ..., \beta_k^n)$. The superscript on the β^n indicates the dependence on n to allow for meeting and subgraph formation rates to vary along the sequence.

4.3.2. *Counting Subgraphs and Estimating Parameters.* Consider a SUGM and order the classes of the subgraphs, $G_1, \ldots, G_\ell, \ldots, G_k$, from 'largest' to 'smallest'. In particular, we choose the ordering of $1, \ldots, k$ so that a subgraph in G_ℓ^n cannot be a subnetwork of the subnetworks in $G_{\ell'}^n$ for $k \geq \ell' > \ell \geq 1$:

$$g_\ell \in G_\ell^n \text{ and } g_{\ell'} \in G_{\ell'}^n \text{ implies that } g_\ell \not\subset g_{\ell'}.$$

There exists at least one such ordering - for instance, any ordering in which subgraphs with more links are counted before subgraphs with fewer links. In an example with links, 2-stars and triangles: triangles precede 2-stars which precede links. Note that this is a partial order: for instance, a 'three link line' ij, jk, kl is neither a subgraph nor a supergraph of a '3-star' ij, ik, il, which is also a three link subgraph on four nodes. It is irrelevant in which order subgraphs with the same number of links are counted.

So, we count subgraphs in this order, and after having removed links associated with all of the subgraphs already counted, denoted $\widetilde{S}_\ell^n$:[29]

$$\widetilde{S}_\ell^n(g) = |\{g_\ell \in G_\ell^n : g_\ell \subset g \text{ and } g_\ell \not\cap g_{\ell''} \text{ for any } g_{\ell''} \in G_{\ell''}^n \text{ such that } g_{\ell''} \subset g \text{ for some } \ell'' < \ell\}|.$$

To define the direct parameter estimates, $\widetilde{\beta}$s, from these counts, we then need to divide by the number of possible subgraphs that could exist. Again, we work without demographics on the subgraphs, but otherwise these counts can be adjusted accordingly.

Let κ_ℓ denote the (finite number) of relabelings to count different subgraphs of type ℓ on a given set of m_ℓ nodes.[30]

The direct estimator $\widetilde{\beta}^n$ is then

$$\widetilde{\beta}_\ell^n = \frac{\widetilde{S}_\ell^n(g)}{\kappa_\ell \binom{n}{m_\ell}}. \tag{4.1}$$

As we prove next, under suitable conditions, these direct estimators are consistent estimates of the true parameters, and they are asymptotically Normally distributed.

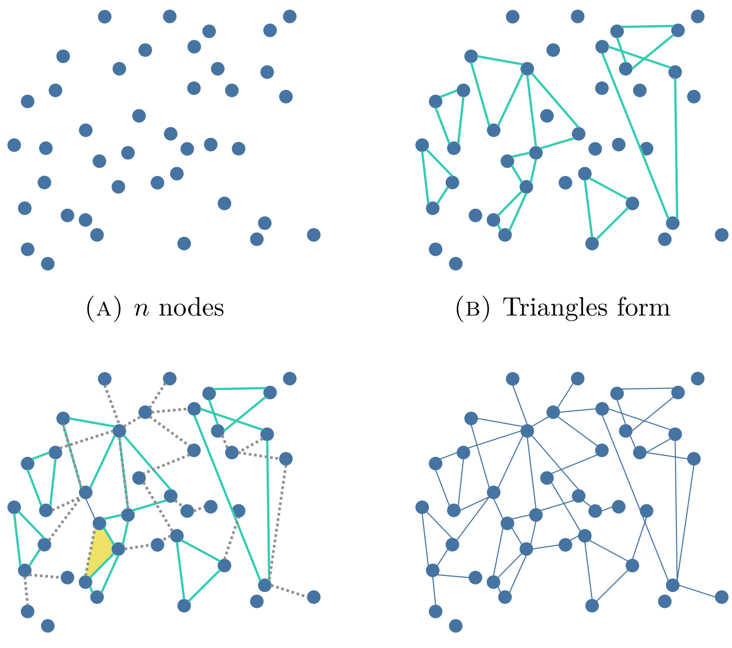
As an illustration, consider Figure 5 in which links and triangles are formed on 41 nodes. There are 9 truly generated triangles, but 10 observed overall. So, the frequency of triangles, $\widetilde{S}_T^n(g)$, is overestimated by using 10 instead of 9. The true frequency was 9/10660 but is estimated as

[29]Note in terms of the notation here, counting in order from 'largest' to 'smallest' subnetworks means that we count things from smallest to largest index ℓ: so the specification of how we ordered labels moves in the opposite direction of the size of the subgraphs.

[30]For example, note that $\kappa_\ell = 1$ for a triangle but for a K-star it is K since each star is different when a different member of the K nodes is the center.

10/10660. With respect to links, there were actually 25 truly directly generated, but one becomes part of an incidentally generated triangle and two others overlap on existing triangles, and so $\widetilde{S}^n_L(g)$ becomes 22 instead. So we estimate 22/820 while the true frequency was 25/820.

These errors are already small on a network on just 41 nodes, and as we prove next, the errors disappear completely as n grows.



(A) n nodes (B) Triangles form

(C) Links form (D) Resulting network

FIGURE 5. A network is formed on 41 nodes and is shown in panel D. The process can be thought of as first forming triangles as in (B), and links as in (C). Note that two links form on triangles, and a third link incidentally generates an extra triangle. In this network we would count $\widetilde{S}^n_T(g) = 10$, and $\widetilde{S}^n_L(g) = 22$ from (D), while the true process generated 9 triangles and 23 links directly. The estimates become $\widetilde{\beta}^n_T = \frac{10}{10660}$, and $\widetilde{\beta}^n_L = \frac{22}{820}$, while the true frequencies were $\frac{9}{10660}$ and $\frac{25}{820}$.

4.4. **Generating Classes.** To understand the rate of incidental subgraph formation, we have to track how many ways a potential subnetwork $g' \in G^n_\ell$ could be incidentally generated – some being equivalent up to relabelings.

We first provide a precise specification of what it means to be incidentally generated. We say that a subgraph $g' \in G_\ell$ for some ℓ can be *incidentally generated* by the subgraphs $\{g^j\}_{j\in J}$, indexed by J, if $g' \subset \cup_{j\in J} g^j$.

Consider any potential subgraph $g' \in G^n_\ell$ that can be incidentally generated by a set of subnetworks $\{g^j\}_{j\in J}$ with associated indices ℓ_j and also by another set $\{g^{j'}\}_{j'\in J'}$. We say that $\{g^j\}_{j\in J}$ and $\{g^{j'}\}_{j'\in J'}$ are *equivalent generators* of g' if there exists a bijection π from J to J' such that $\ell_j = \ell_{\pi(j)}$ and $|g_j \cap g'| = |g_{\pi(j)} \cap g'|$. So the equivalent generating sets have the same configurations in terms of numbers and types of subgraphs, and in terms of how many nodes each of those subgraphs intersects the given network.

So, for instance a triangle 123, could be incidentally generated by links 12, 23, and triangle 134; and an equivalent generator is links 12, 23, and triangle 135, and another is links 23, 13; and triangle 128, and so forth.

Given this equivalence relation, ignoring the specific labels of subgraphs we can define *generating classes* for any type of subgraph G_ℓ. We just keep track of the number and type of subgraphs needed, as well as how many nodes each has subgraph intersecting with the given incidentally generated subgraph.

So, each generating class $\mathcal{C}$ of some G_ℓ^n is a list $\mathcal{C} = (\ell_1, c_1, \ldots, \ell_C, c_C)$ consisting of a list of types of subgraphs used for the incidental generation and how many nodes each has intersecting with the given incidentally generated subgraph. Thus, $\mathcal{C} = (\ell_1, c_1, \ldots, \ell_C, c_C)$ is such that there $\exists g' \in G_\ell^n$ generated by some $\{g^j\}_{j\in J}$ for which $|J| = C$ and for each j: $g^j \in G_{\ell_j}^n$ and $c_j = |g^j \cap g'|$.

We order generating classes so that the indices are ordered: $\ell_j \leq \ell_{j+1}$, and lexicographically $c_j \leq c_{j+1}$ whenever $\ell_j = \ell_{j+1}$. This ensures that we avoid counting the same class twice.[31]

We only need to work with a small set of generating classes, so we restrict attention to the following:

- generating classes that are minimal: in the above J there cannot be j' such that $g' \subset \cup_{j\in J, j\neq j'} g^j$, and
- generating classes that only involve smaller subgraphs: $\ell_j \geq \ell$ for all $j \in J$.

The second condition states that we can ignore many generating classes because of our counting convention: when counting any given subgraph type, we only have to worry about incidental generation by the remaining (weakly smaller) subgraphs.

So, for a links and triangles example, where $G^n = (G_T^n, G_L^n)$ are triangles and links, there are four generating classes of a triangle: a triangle could be incidentally generated by three other triangles, two triangles and one link, two links and one triangle, or three links.[32] Under the last condition above, there are no generating classes for links to worry about, since they cannot be incidentally generated by themselves and we only count them after removing all triangles.

4.4.1. *Identification, Consistency and Asymptotic Normality.* It is convenient to express the β_ℓ^ns in the form

$$\beta_\ell^n = \frac{b_\ell}{n^{h_\ell}}$$

for some $b_\ell > 0$ and $h_\ell > 0$. This allows us to directly see how the parameter varies with n, and is a general way of encoding the rates that could come from meeting, time budgets, or any other story that gives rise to sparse sequences.

We consider the case in which $m_\ell > h_\ell$, as otherwise the expected number of subgraphs in the whole network could be bounded as n grows, precluding estimation.

The researcher can make assumptions on h_ℓ – and this order is easy to observe with simple subgraph counts, and generally quite steady across settings. For instance in most friendship networks, people have somewhere between two and a thousand connections, depending on whether one is counting close friends or acquaintances, or somewhere in between. This, as a function of the size of

[31]However, a generating class of two links and a triangle is a different generating class than one link and two triangles - this numbering just avoids the double counting of two links and a triangle separately from a triangle and two links.

[32]Here, then we would represent a generating class of two triangles and a link as $(T, 2; T, 2; L, 2)$, where this indicates that two triangles were involved and each intersected the subgraph in question in two nodes and then $L, 2$ indicates that a link was involved intersecting the subgraph in two nodes.

the network yields a direct estimate of the order of h_ℓ, which then is at least 1 or larger for links, 2 or larger for triangles, etc. – so that people have roughly some number of connections that is the same as a function of the network size. In fact, in most statistical or econometric or economic models of network formation, these are implicitly made whether knowingly or unknowingly.

We show that even without knowing the b_ℓ or h_ℓ, the parameters β_ℓ^n can be well-estimated, provided the true parameters satisfy some basic conditions (that will be satisfied by network models that are not so sparse that subgraphs are never observed, nor so dense so that they scale linearly in n).

The following conditions ensure that the direct estimation parameters are arbitrarily accurate for large enough networks.

First, for each ℓ let

$$h_\ell > m_\ell - 2. \tag{4.2}$$

This condition ensures that the overall degree of any node grows more slowly than the size of the graph. The average degree can still grow with n, but sublinearly. In particular, this condition ensures that the chance that any given link is part of multiple subgraphs is vanishing.

Next, for each ℓ consider any (minimal)[33] generating class with index J of subgraphs no larger than ℓ. Have[34]

$$h_\ell < \sum_{j\in J} h_j + c_j - m_j \tag{4.3}$$

and

$$h_{j'} + m_\ell - m_{j'} < \sum_{j\in J} h_j + c_j - m_j \tag{4.4}$$

for each $j' \in J$.

(4.3) is the requirement that a given subgraph is more likely to form directly than indirectly. (4.4) is the requirement that a given subgraph disappear into incidentally forming larger subgraphs more slowly than that subgraph forms directly.

Under these conditions, we prove identification in addition to consistency and asymptotic normality on a single large network.

It is useful to define the variance-covariance matrix

$$V_n = \operatorname{diag}\left\{ n^{2h_\ell} \frac{\beta_{0,\ell}^n (1-\beta_{0,\ell}^n)}{\kappa_\ell \binom{n}{m_\ell}} \right\}.$$

THEOREM **3.** *Consider a sequence of SUGMs of k distinct types of subgraphs with associated true parameters $0 < b_{0,\ell}$ for which $\beta_{0,\ell}^n = \frac{b_{0,\ell}}{n^{h_\ell}}$. If (4.2)-(4.4) hold for all ℓ and associated (minimal) generating classes, then $|\widetilde{b} - b_0| \xrightarrow{\mathrm{P}} 0$ and $V_n^{-1/2}\left(\widetilde{b} - b_0\right) \rightsquigarrow \mathcal{N}(0, I)$.*[35]

Although the conditions may appear hard to understand, they are actually fairly straightforward, and it is easy to see sufficient conditions that ensure them.

[33]If the condition is satisfied by minimal classes, it is automatically satisfied by larger classes.

[34]Here we simplify notation and substitute j for ℓ_j, when clear.

[35]The same conditions apply to $\widetilde{\beta} \xrightarrow{\mathrm{P}} \beta_0$ with asymptotic normality, and thus it is not necessary that the researcher have any prior knowledge of the h_ℓs.

For example, suppose that each $h_\ell = m_\ell - x$ for some same $x \in (0,2)$, so that each node has the same order probability of being a part of different sorts of subgraphs. This is the natural case, as otherwise some subgraphs become infinitely more likely than others.

In that case, all three conditions are automatically satisfied whenever the subgraphs are all cyclic subgraphs (cliques, or other subgraphs in which all nodes are parts of cycles). If some of the subgraphs are not cyclic (e.g., are lines or stars), then all three conditions hold if $x \in (0,1)$.

Corollary **1.** *Consider a sequence of SUGMs of k distinct types of subgraphs with associated true parameters $0 < b_{0,\ell}$ for which $\beta^n_{0,\ell} = \frac{b_{0,\ell}}{n^{h_\ell}}$. If $m_\ell - h_\ell = x$ for each ℓ and some $x \in (0,2)$ and either all subgraphs are cyclic or else $x < 1$, then $|\widetilde{b} - b_0| \xrightarrow{\mathrm{P}} 0$ and $V_n^{-1/2}\left(\widetilde{b} - b_0\right) \rightsquigarrow \mathcal{N}\left(0, I\right)$.*

In both results, although we state them in terms of bs, it is also the case that the ratio of $\hat{\beta}_\ell$ to $\beta_{0.\ell}$, tends to one. Furthermore, as we show in the proof, if we normalize the difference between the estimated probability and the truth by the standard deviation, this is asymptotically normally distributed. This is an equivalent representation of the above result, but is helpful to note as it does not require knowledge of h_ℓs but rather just the assumptions that they satisfy the relevant bounds from the conditions, which will generally be true of most data sets of human networks.

4.4.2. *Links and Triangles.* Theorem 3 works with parameter values for which incidental generation eventually becomes small as a function of the overall counts of the subgraphs, and works for arbitrary subgraph varieties. For SUGMs with specific subgraph types, we can explicitly calculate all the incidental rates and account for them, and develop an estimator that not only works well asymptotically, but also will be more accurate in smaller (finite) samples, where there can be nontrivial incidental generation. In particular, in this section we present a result for the links and triangles SUGM based on an estimator that fully accounts for the incidental generation (see the details in Appendix C). This involves checking identification for this sequence as well as demonstrating consistency and asymptotic normality.

In particular, we show that

$$\frac{S^n_L(g) - \mathrm{E}_{\beta^n_0}[S_L(g)]}{\sigma^n_L} \rightsquigarrow \mathcal{N}(0,1), \text{ and } \frac{S^n_T(g) - \mathrm{E}_{\beta^n_0}[S_T(g)]}{\sigma^n_T} \rightsquigarrow \mathcal{N}(0,1),$$

and jointly as well, where $(\sigma^n_L)^2 := \mathrm{var}\left(S^n_L(g)\right)$ and $(\sigma^n_T)^2 := \mathrm{var}\left(S^n_T(g)\right)$. Since

$$S^n_L(g) = \frac{\sum_{i<j} g_{ij}}{\binom{n}{2}} \text{ and } S^n_T(g) = \frac{\sum_{i<j<k} g_{ij}g_{ik}g_{jk}}{\binom{n}{3}}$$

and g^n_{ij} and g^n_{ik} are correlated for any k, S^n_L involves correlated random variables, and since any two triples in S^n_T that involve a common link are correlated, we need to carefully check that we can still prove a central limit theorem and that the correlation will not cause problems in the limit.

It is useful to define the variance-covariance matrix of the moments

$$V_n = \begin{pmatrix} \mathrm{var}(n^{h_L}S_L) & \mathrm{cov}(n^{h_L}S_L, n^{h_T}S_T) \\ \mathrm{cov}(n^{h_L}S_L, n^{h_T}S_T) & \mathrm{var}(n^{h_T}S_T) \end{pmatrix}.$$

With this defined we can state our result.

Proposition **3.** *Consider a links and triangles SUGM with associated parameters* $\beta^n_{0,L}, \beta^n_{0,T} = \left(\frac{b_{0,L}}{n^{h_L}}, \frac{b_{0,T}}{n^{h_T}}\right)$ *with* $0 \leq \underline{D} < b_{0,L}, b_{0,T} < \overline{D}$ *such that* $h_L \in (2/3, 2)$ *and* $h_T \in [h_L + 1, 3h_L]$, *with* $h_T < 3$. *Then* $|\widehat{b}^n - b_0| \xrightarrow{\text{P}} 0$ *and* [36] $V_n^{-1/2}\left(\widehat{b}^n - b_0\right) \rightsquigarrow \mathcal{N}(0, I)$.

This covers a wide range of link and triangle densities, ranging from average degree on the order $n^{1/3-\delta}$ to $n^{-1+\delta}$ for any $\delta > 0$. This covers the order constant and logarithmic growth rates of average degree studied in the literature (Newman et al., 2001; Bollobas, 2001; Jackson, 2008; Graham, 2017).

The restrictions are easily interpretable. $h_T \geq h_L + 1$ ensures that triangles are not so numerous that almost all of the links in the network lie in triangles: that n^{3-h_T} does not dwarf n^{2-h_T}. $h_T \leq 3h_L$ ensures the opposite: that triangles are not almost always incidentally formed by links, and never forming directly: $n^{3(1-h_T)}$ is not dwarfed by n^{3-3h_L}. $h_L > 2/3$ ensures that links and triangles are disentangled by imposing a density cap. Finally, $h_L < 2, h_T < 3$ ensure that there is information in the network—enough links and triangles to estimate their formation.

Again we note that although the results are stated in terms of b, these are equivalent statements to saying that ratio of the estimated ($\hat{\beta}$) and true (β_0) frequencies tend to one. And, that, when self-normalized by the standard deviations, the empirical frequencies estimated are asymptotically normally distrubted. The advantage of observing this is that it requires no knowledge of h_ℓs other than that they satisfy the relevant bounds.

5. Applications

We now apply our model to study the four examples from Section 2 to illustrate the kinds of questions SUGMs can be used to addressed.

5.1. **Data.** We use the Banerjee, Chandrasekhar, Duflo, and Jackson (2013, 2014) data consisting of a variety of social and economic networks from 75 Indian villages as well as detailed demographic background.[37] Having 75 villages worth of data allows us to show not only how the model scales with the number of nodes, but also with the number of networks observed.

The networks have households as nodes. There are an average of 220 households per village. We surveyed adults, asking them about a variety of their daily interactions, as well as their demographics (caste, education, profession, religion, family size, wealth variables, voting and ration cards, self-help group participation, savings behavior, etc.). We have network data from 89.14 percent of the 16,476 households based on interviews with 65 percent of all adults between the ages of 18 and 55. As we study the undirected, unweighted network described below, this means that we observe 98.8% of the potential links between pairs.[38] We have data concerning twelve types of interactions: (1) whose houses he or she visits, (2) who visits his or her house, (3) his or her relatives in the village, (4) non-relatives who socialize with him or her, (5) who gives him or her medical help, (6) from whom he or she borrows money, (7) to whom he or she lends money, (8) from whom he or she borrows material goods (e.g., kerosene, rice), (9) to whom he or she lends material goods, (10) from whom he or she gets important advice, (11) to whom he or she gives advice, (12) with whom he or she goes to pray (e.g., at a temple, church or mosque).

[36] The expression for V_n is different when $h_T = h_L + 1$, and is given in the proof of the proposition.

[37] See Banerjee, Chandrasekhar, Duflo, and Jackson (2013) for more information about the data.

[38] This is a new wave of data relative to our original microfinance study that includes more surveys. Note that $1 - (1 - 0.8914)^2 = 0.988$.

The answers are aggregated to the household level, but one can also work with the individual-level networks to get very similar results as those presented below. How a link is defined varies based on the application. We use undirected,[39] unweighted networks that may allow for multiplexing.

For much of what follows, we work with the borrowing and lending of material goods (questions 8 and 9, with any positive answer indicating a link being present) that we call "favor" links, and the exchange of advice (questions 10 and 11, with any positive answer indicating a link being present) that we call "info" links.

5.2. **Example 1 (cont.): Matching Features of Empirical Network Data.** Revisiting the example from Section 2.4.1, we compare a simple SUGM based on links and triangles to a number of alternative models that allow a rich set of covariates, unobserved heterogeneity through node fixed effect, and even unobserved latent space positions to model correlation.

Our goal is to demonstrate that when a SUGM is fit to the data, as compared to each of the other models, the resulting distribution over *economically relevant* network features resembles the data far better under the SUGM as compared to the other four alternatives. These measures were not used to fit the model. They the size of the giant component, average path length, and various spectral properties of the adjacency matrix (e.g., the largest eigenvalue and an eigenvalue measure of homophily). SUGM outperforms the alternatives despite the fact that the other alternatives have many variables such as numerous covariates, n fixed effects, or even n latent space variables.

Specifically, our alternatives include (a) a standard model of conditional edge independence (or a stochastic block model) that includes flexible controls for continuous covariates; (b) an extension of this model that includes n parameters to capture node fixed effects (e.g., Graham (2017)); (c) a latent space model (Hoff, Raftery, and Handcock, 2002) where nodes have unobserved arbitrary locations in $\mathbb{R}^3$ to be estimated and the probability of linking declines in their latent positions; and (d) an exponential random graph model with links, triangles, and the rich set of covariates.

Before we proceed, let us review why we look at these features of the graph structure. We look at the first eigenvalue of the adjacency matrix, which is a measure of diffusiveness of a network under a percolation process (e.g., Bollobás, Borgs, Chayes, and Riordan (2010); Jackson (2008)). This is intimately related to the expansiveness of the network – namely, for any subset of nodes the number of links leaving the subset relative to the number of links within the subset. We are also interested in the second eigenvalue of the stochasticized adjacency matrix.[40] This is a quantity that is key in local average learning processes and modulates the time to consensus (DeMarzo, Vayanos, and Zwiebel (2003); Golub and Jackson (2012)), but is also closely related to homophily (Golub and Jackson (2012)) and is labeled as such in the table below. Additionally, we look at the fraction of nodes that belong to the giant component of the network, as well as the number of isolates, as empirical networks are often not completely connected. Finally, we also consider average path length (in the largest component).

Again, we present the results for favor and info networks. These networks are reasonably connected (with more than ninety percent of the nodes being in a giant component) and yet also reasonably sparse for small networks.

[39]Some links are not reciprocated, but that is true at similar rates for the questions regarding relatives as compared to the other questions, and so much of the failure of reciprocation may simply be measurement error rather than true one-way relationships. For our purposes here, which are purely to illustrate the ability of the models to work with data, this distinction is inconsequential.

[40]The stochasticized adjacency matrix T is defined as $T_{ij} = \frac{g_{ij}}{\sum_k g_{ik}}$, where either $g_{ii} = 1$, or $g_{ik} > 0$ for some $k \neq i$, as this captures the set of people to whom i listens.

Our procedure is as follows. For every village, we estimate six network formation models. One network formation model is a link-based model (stochastic block model) in which the probabilities can depend on geographic distance, caste, the number of rooms households have, number of beds, quality of electricity provision, quality of latrines, household ownership status, and squared differences in non-binary variables. The probabilities are estimated using logistic regression and the model has 12 parameters. The next is the model of Graham (2017). This is the same formulation of the preceding model, but adds unobserved heterogeneity in the form of node-fixed effects:

$$\mathrm{P}(g_{ij} = 1 | X_{ij}) = \Lambda\left(\alpha_i + \alpha_j + \gamma' X_{ij}\right)$$

where $\Lambda(\cdot)$ is the logit link function and X_{ij} is the aforementioned vector of demographic characteristics and polynomials therein. This model has n+12 parameters per network.[41]

The third model is a latent space model

$$\mathrm{P}(g_{ij} = 1 | z_{ij}) = \Lambda\left(\alpha_i + \alpha_j - \eta \cdot \mathrm{dist}(z_i, z_j) + \gamma' X_{ij}\right)$$

where now z_i are unobserved positions in $\mathbb{R}^3$. This has $2n + 12$ parameters.

The fourth model is a links and triangles ERGM with covariates. Specifically,

$$\mathrm{P}(g) \propto \exp(\theta_L \cdot S_L(g) + \theta_T \cdot S_T(g) + \gamma' X).$$

Turning to SUGMs, in contrast, we consider only low-dimensional models. One is a the basic SUGM with links and triangles. Pairs of household are categorized as either being "close" or "far," where "close" refers to pairs of nodes that are of the same caste and "far" to those that differ in caste. Similarly, we categorize triangles as being "close" if all nodes are of the same caste and "far" otherwise. Thus, we allow for four parameters, close and far link parameters and close and far triangle parameters. The other model is a slightly richer SUGM in which we allow some nodes to be isolates, meaning there are five parameters. Neither includes any other demographic covariates nor unobserved heterogeneity.

To make the strongest point, we compare these very stark SUGMs that use only caste variables to account for homophily, to very rich covariate dependent (block) models that can incorporate a large set of covariates – including much richer demographics that are usually available to a researcher as well as node-level fixed effects in the unobserved heterogeneity model and node-level latent locations in the latent space model. We show that even though we have considerably more information on the nodes, such as geographic distance and demographic characteristics, and allow for such unobserved heterogeneities—and we do not make use of this information for the SUGMs—they recreate networks much more accurately than a link-based model that does takes advantage of a rich set of node characteristics. Adding over 12 parameters to the block model to flexibly control for demographic attributes, *or even* n*+12 parameters with unobserved heterogeneity or* $2n + 12$ *with latent locations*, does not come close to doing as well as the simple SUGMs. Moreover, since the specification developed here makes use of considerably richer data than those used in the two candidate SUGM models, it suggests that by decomposing a network into a tapestry of random structures (triangles, links, and even isolates), considerable value is added in modeling higher order features of networks in a parsimonious way.

We estimate parameters for the village network for each model and then generate random network from each model based on the estimated parameters. We do 100 such simulations for each of the

[41]Consistency of all α_i in addition to β has been proven for a dense sequence of graphs (e.g., Chatterjee et al. (2010); Graham (2017)).

Table 1. Network Properties

	Truth	*Links/Triangles SUGM*	*Links/Triangles/Isolates SUGM*	*Covariates (Block Model)*	*Covariates + Unobserved Heterogeneity (Latent Block Model)*	*Latent Space Model (with Covariates)*	*ERGM (Links/Triangles with Covariates)*
Panel A: Information	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Degree	8.096	8.076	8.042	8.815	9.621	13.500	13.526
	(0.261)	(0.263)	(0.255)	(0.311)	(0.354)	(0.139)	(0.112)
Clustering	0.220	0.159	0.147	0.051	0.075	0.077	0.134
	(0.006)	(0.003)	(0.003)	(0.003)	(0.005)	(0.001)	(0.001)
Isolates	10.972	3.503	13.787	0.499	0.873	10.269	11.987
	(0.841)	(0.408)	(0.998)	(0.092)	(0.157)	(0.138)	(0.138)
% in Giant	0.950	0.984	0.938	0.998	0.996	0.946	0.931
	(0.003)	(0.002)	(0.003)	(0.000)	(0.001)	(0.001)	(0.001)
Maximal Eigenvalue	11.914	10.453	10.816	10.374	12.583	16.463	18.174
	(0.374)	(0.301)	(0.300)	(0.321)	(0.430)	(0.144)	(0.108)
Homophily	0.887	0.815	0.804	0.686	0.680	0.876	0.772
	(0.007)	(0.009)	(0.009)	(0.010)	(0.010)	(0.001)	(0.002)
Average Path Length	3.027	2.957	2.871	2.758	2.641	3.097	2.863
	(0.048)	(0.043)	(0.040)	(0.040)	(0.037)	(0.010)	(0.012)
Panel B:Favors							
Degree	7.058	7.037	7.087	7.756	8.339	13.011	15.015
	(0.261)	(0.261)	(0.267)	(0.324)	(0.366)	(0.156)	(0.133)
Clustering	0.289	0.190	0.176	0.047	0.064	0.071	0.145
	(0.005)	(0.003)	(0.003)	(0.003)	(0.004)	(0.001)	(0.001)
Isolates	10.070	7.558	16.094	1.063	4.254	19.041	17.769
	(0.767)	(0.748)	(1.177)	(0.142)	(2.132)	(0.268)	(0.213)
% in Giant	0.951	0.962	0.925	0.995	0.979	0.873	0.901
	(0.003)	(0.003)	(0.005)	(0.001)	(0.011)	(0.002)	(0.001)
Maximal Eigenvalue	10.065	9.618	9.991	9.479	11.198	15.821	20.082
	(0.334)	(0.311)	(0.323)	(0.339)	(0.408)	(0.161)	(0.120)
Homophily	0.941	0.873	0.863	0.733	0.723	0.904	0.790
	(0.004)	(0.008)	(0.009)	(0.011)	(0.011)	(0.001)	(0.002)
Average Path Length	3.516	3.166	3.080	2.915	2.806	3.812	2.929
	(0.066)	(0.047)	(0.044)	(0.044)	(0.054)	(0.021)	(0.015)

Notes: Average value of various network statistics for the information and favor networks across 75 villages are shown in Column 1. Columns 2-7 present the average values across the 75 villages with 100 simulations per village generated from the estimated parameter value for each model specified. Column 2 presents a links and triangles in a four-parameter SUGM with covariates where links and triangle probability can vary by binary classification: whether the pairs or triples are "close" or "far", as described in the text. Column 3 adds isolates to the previous SUGM to constitute a five-parameter SUGM. Column 4 consists of a twelve-parameter conditional edge independent model that includes flexible controls for continuous covariates, as described in the text. Column 5 adds to this unobserved heterogeneity, by including fixed effects for each of node i and node j, when determining whether the pair ij are linked, which adds n-more parameters per network of size n. Column 6 adds to this latent space coordinates, a latent location for each node in 3-dimensional Euclidean space, thereby adding n-more parameters per network of size n. Column 7 studies an ERGM with links and triangles as the sufficient statistics, as well as all of the aforementioned controls. Standard errors for the means in parentheses.

75 village and for each of the models. We then compare the true network characteristics with those from the simulations.

Table 1 presents the results. Both of the SUGMs match the various features of the networks substantially better than the conditional edge independent models (with and without node fixed effects). Including isolates in the SUGM further improves the fits not only for isolates, but also for fraction in the giant component and the maximum eigenvalue. This suggests that there are more isolated households in a village for a reason outside of randomness in network formation.

The most obvious thing to note is that the link-based and also latent space models do extremely poorly when it comes to matching clustering while the SUGM does much better, and here adding unobserved dimensions to generate unconditional link correlations (e.g., clustering) does worse than a SUGM that allows correlated link formation directly. The ERGM performs better on clustering but generates excess density, diffusiveness, the spectral cut (homophily), connectedness, and average path length.

Conditioning on the triangles in the SUGM is enough to deliver better matches on all dimensions, and the difference on homophily is perhaps most interesting, since one would imagine that the block models or even latent space models could get that right given that they include many covariates. This tells us that triangles and correlation between links play a subtle but important role in homophily – something that is better picked up by a SUGM than an independent link model even when that model includes rich demographics and unobserved heterogeneity.

That SUGMs do a much better job at recreating a multitude of features of observed network structures that standard link-based models, especially with rich demographic information, models with unobserved heterogeneity, latent space models, and ERGMs is important. It suggests that there is substantial value added of modeling the formation of triangles and isolates. Knowing that our model is better able to capture the realistic correlation of links within observed networks should make us more confident in trusting the results of some other empirical applications. For example, when we look at links across social boundaries, we can be comfortable that to a first order, thinking about a SUGM with links and triangles across and within caste groups can do a good job of matching patterns in the data, and thus tracing them back to model parameters.

5.3. **Example 2 (cont.): Do incentives for risk sharing drive network formation?** Continuing Example 1 from Section 2.4.2, we test whether supported relationships are significantly more likely to appear in favor exchange than informational links. The (joint) hypothesis that we are testing is that exchanging material goods is more costly and/or happens less frequently for agents, and so requires more incentives and supporting enforcement than exchanging information which is less costly and/or more frequent.

To keep the illustration in this first example clear, we abstract from covariates. We illustrate the incorporation of covariates in the examples below.

Thus, from Section 2.4.2, we know that $\beta_{T,favor} = F(\gamma_{T,favor})^3$ and $\beta_{L,favor} = F(\gamma_{L,favor})^2$, and similarly for information. Thus, the test of whether $\frac{\gamma_{T,favor}}{\gamma_{L,favor}} > \frac{\gamma_{T,info}}{\gamma_{L,info}}$ corresponds to $\frac{\beta_{T,favor}/\beta_{L,favor}^{3/2}}{\beta_{T,info}/\beta_{L,info}^{3/2}} > 1$. This test takes into account that there are more consents for a group than a pair (the 3/2), and is also robust to information links simply being more or less valuable, as it adjusts by relative link prevalence.

TABLE 2. Parameter estimates by network type

	$\widehat{\beta}_L$	$\widehat{\beta}_T$
Information	0.0131	0.0001
	(0.0002)	(0.0000)
Favors	0.0150	0.0002
	(0.0002)	(0.0000)

Notes: Standard errors computed using the results of Proposition 3.

We estimate the four parameters in question under the many independent network (n fixed, $R \to \infty$) framework. Table 2 presents the parameter estimates and standard errors. We reject the hypothesis that there is no difference in the support of favor relationships compared to information relationships ($p < 0.01$).[42] We conclude that the data are consistent with the theory that incentives for favor exchange matters in network formation in these data.

Next, we can also push this further by estimating $\widehat{\beta}_{L,v}$ and $\widehat{\beta}_{T,v}$ separately for each village v, with the large single network ($n \to \infty$, $R = 1$) paradigm for each village. This allows for heterogeneity in the parameters across villages by assuming they are drawn from entirely different distributions. We see the results in Figure 6, though standard errors are omitted for visual clarity. We see that for most villages, the favor over info ratios are higher for triangles compared to links.

[42]Specifically, the p-value is computed for a test of the null hypothesis $\frac{\beta_{T,favor}}{\beta_{T,info}} = \frac{\beta_{L,favor}^{3/2}}{\beta_{L,info}^{3/2}}$, where the parameters are held to be common across all villages in the sample.

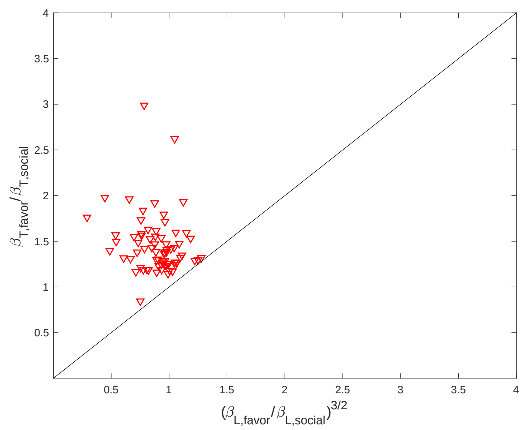


FIGURE 6. Plots of estimates of $\frac{\beta_{T,favor}}{\beta_{T,info}}$ against $\frac{\beta^{3/2}_{L,favor}}{\beta^{3/2}_{L,info}}$ by village.

5.4. **Example 3 (cont.): Links across Social Boundaries.** We link two households if members of either engaged in favor exchange with each other: that is, they borrowed or lent goods such as kerosene, rice or oil in times of need.

TABLE 3. Parameter estimates by network type

	$\widehat{\beta}_{L,same}$	$\widehat{\beta}_{T,same}$	$\widehat{\beta}_{L,diff}$	$\widehat{\beta}_{T,diff}$
Information	0.0169	0.0004	0.0065	0.00003
	(0.0002)	(0.0000)	(0.0001)	(0.0000)
Favors	0.0127	0.0004	0.0043	0.00002
	(0.0002)	(0.0000)	(0.0001)	(0.0000)

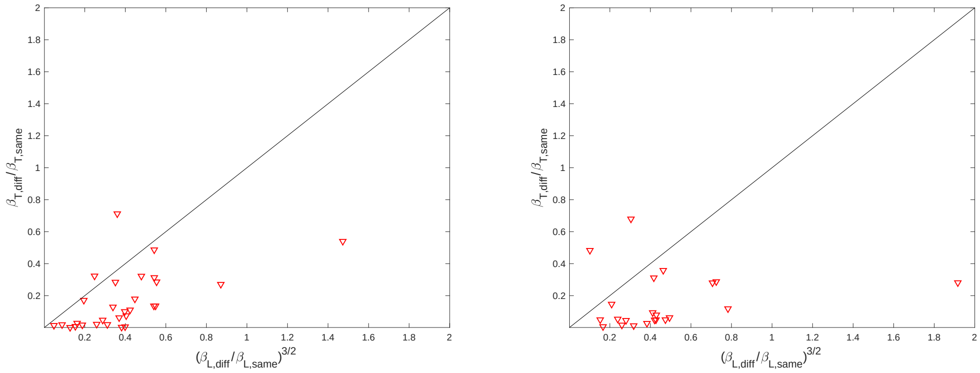
Notes: Standard errors computed using the results of Proposition 3.

Table 3 presents the parameter estimates, again where we assume that all 75 networks are independent draws from the same distribution, and a formal test rejects the null $\frac{p_T(diff)}{p_T(same)} = \left(\frac{p_L(diff)}{p_L(same)}\right)^{3/2}$ with $p < 0.001$ for each network type.

Finally, Figure 7 shows the results when we allow the parameter estimates to vary by village. For the bulk of villages, cross-caste relationships relative to within-caste relationships are more frequent as isolated links compared to being embedded in triangles, for both information and favor networks.

5.5. **Example 4 (cont.): Network Formation and the Impact of Microfinance.** Finally, we return to Example 4: how the introduction of microfinance changes network structure. We begin by estimating the SUGM parameters outlined in Lemma 3. Recall that the modeling assumptions are based on Banerjee et al. (2021) and as noted above the lending/borrowing frequencies, the return on informal lending and borrowing, and estimates of node types via a random forest classification are taken from that paper. Our goal here is to showcase the usefulness of the SUGM.

We estimate efforts using Lemma 3, and then can identify other parameters of interest as follows:



(A) Information

(B) Favors

FIGURE 7. Plots of estimates of $\frac{\beta_{T,diff}}{\beta_{T,same}}$ against $\frac{\beta^{3/2}_{L,diff}}{\beta^{3/2}_{L,same}}$ by village, for information and favors networks.

(1) $1 = \left(\sqrt{\frac{\beta_{\theta\theta}}{v^2_{\theta\theta}}}\right)^{1/2} \cdot \left(\frac{\phi(2v_{\theta\theta}+s)^3}{\beta_{\theta\theta}}\right)^{1/3}$ for every θ without microfinance and H with microfinance.

(2) $\zeta = \frac{\beta^{MF}_{HHH}}{\beta^{MF}_{HH}} \cdot \frac{\left(v^{MF}_{HH}\right)^2}{\left(v^{MF}_{HHH}\right)^3} \cdot \phi^{-1}$.

Table 4 presents the results. We see that effort for N declines (but H stays the same) in the microfinance villages. Turning to support, in order to interpret this, it is useful to scale the support factor relative to the base values v^{MF}_{NN} and v^{MF}_{HH}. Specifically, we find that the support value is 34.5% of two bilateral relationships (presumably the value of a triad with no extra payoff for closure), consistent with Jackson et al. (2012). Next, turning to the meeting technology, consistent with Feigenberg et al. (2010), we see a considerable increase in meetings among Hs with microfinance. Given the weekly meetings at microcredit centers among participants, this is unsurprising.

TABLE 4. Microfinance impacts on network formation

	$e^{\star}_H$	$e^{\star}_N$	s	ζ	ϕ
No MF	0.191	0.110	0.008	N/A	0.0065
	(0.008)	(0.0170)	(0.001)	N/A	(0.0005)
MF	0.184	0.050	0.008	20.8	0.0065
	(0.009)	(0.0013)	(0.001)	(0.002)	(0.0005)

6. A Central Limit Theorem for Correlated Random Variables

We now state a new central limit theorem that applies for a variety of settings in which all variables may be correlated (well-beyond network settings), but in which the total amount of covariance is bounded.

Many existing central limit theorems that allow for correlated random variables do not apply to our setting as they require a spatial/ ordered lattice/ geometric structure (e.g., Bolthausen (1982).

In the typical logic of central limit theorems based on strong mixing arguments in the mathematics, spatial, and time series literature, random variables are embedded in some space where there are "close" and "far" random variables and the further they are, the less correlated they are. Some researchers working on network formation (e.g., Boucher and Mourifié (2012); Leung (2014)) exploit these spatial techniques by embedding nodes in some space so that only "nearby" nodes can link and "distant" nodes cannot link (e.g., following the logic in Penrose (2003); Hoff et al. (2002)) in order to satisfy mixing conditions and apply a central limit theorem like Bolthausen (1982). As $n \to \infty$ most nodes get further and further apart and therefore essentially never link.

The reason that this is unsatisfying for our purposes is that such a strategy imposes a specific structure on the adjacency matrix. To see this, consider the simple case where nodes live on a line. Then in the adjacency matrix, only nodes within some limited distance to the left or right of any given node tend to be linked. While this may be fine for certain contexts, it is not an adequate description of a village network where there is no natural space on which some households in a village should be considered, ex ante, to be infinitely far apart (or students in a dorm or university should be considered, ex ante, to be infinitely unlikely to link to each other).

Our proof technique builds on a foundation developed by Stein (1986). That result (and precursor work in Stein (1972)) have been used before to derive central limit theorems in two literatures: time-series/spatial statistics and dependency graphs. For instance the oft-used Bolthausen (1982) central limit theorem, crucial in time-series and spatial econometrics, uses a lemma from Stein (1972) to show asymptotic normality. The basic idea is to arrange the data in some manner to identify "close" and "far" random variables and establish conditions on mixing as a function of these distances.[43]

The literature on dependency graphs has not been explored as exhaustively in econometrics. In such previous approaches, collections of random variables are represented on a graph (the graph here is a manner of representing correlation between random variables whose indices are the the nodes, not the target SUGM that we are modeling) where a link between two indices mean that they are correlated and no link means they are independent.

Our insight is generalization of the dependency graph literature: what we show is that if the overall covariances satisfy some bounds, then one can still prove a CLT no matter how that dependency is arranged (even with a complete dependency graph). This differs from much the previous literature which assumes that many variables have zero correlation. There the normalized sum is then shown to be asymptotically normally distributed provided the dependency graph is sufficiently sparse (Baldi and Rinott (1989); Goldstein and Rinott (1996); Chen and Shao (2004)). Those results overly-restrictive for our purpose. For instance, we want models where in principle *all* random variables which represent links in the SUGM can be ex ante correlated, and in overlapping ways. Even the few previous results allowing for high- and low-correlation dependency sets are too stringent to apply to our setting (Ross, 2011; Goldstein and Rinott, 1996; Chen and Shao, 2004). We work with weaker conditions that allow us to work with SUGMs, and are stated based on bounds on sums of covariances, differently from conditions in the previous literature.

We require some notation.

Consider a triangular array of (real-valued) random variables X_α^N that each have finite variance (possibly varying with N), with a set of labels $\alpha \in \Lambda^N$ such that $|\Lambda^N| = N$. For instance, in our SUGM settings the X_α^N may be an indicator of the appearance of some particular subgraph, such as a link or triangle, and α would track the pairs of nodes involved in a potential link (ij) or

[43]In this setting, a non-exhaustive but illustrative list of econometrics papers include Conley (1999), Jenish and Prucha (2009), Bester, Conley, and Hansen (2011), among numerous others.

triples of nodes in a triangle (ijk). N captures the $\binom{n}{2}$ possible links or $\binom{n}{3}$ possible triangles. So when considering link counts α would track pairs of nodes involved in links and when considering triangles α would track triples.

Let us normalize the variables by their means:

$$Z_\alpha^N = X_\alpha^N - \mathrm{E}\left[X_\alpha^N\right]$$

and let S^N be the sum

$$S^N := \sum_{\alpha\in\Lambda^N} Z_\alpha^N.$$

We provide conditions under which a normalized statistic

$$\overline{S}^N := \frac{S^N}{a_N^{1/2}} \rightsquigarrow \mathcal{N}(0,1),$$

where the normalizer, a_N, is a measure of the variance of S^N, defined below.

6.1. **Dependency neighborhoods.** For each α, N, we partition the index set Λ^N into two pieces. In particular, we define a set, called a *dependency neighborhood*, for each α, N:

$$\Delta(\alpha, N) \subset \Lambda^N \text{ such that } \alpha \in \Delta(\alpha, N).$$

The conditions for $\eta \in \Delta(\alpha, N)$ are defined below. For each α, we need Λ^N to be partitioned into a set $\Delta(\alpha, N)$ and its complement in a specific manner to satisfy a few sufficient conditions.

$\Delta(\alpha, N)$ includes indices η where the corresponding X_η's have relatively "high" correlation with X_α, and its complement includes the indices η where the corresponding X_η's that have relatively "low" correlation with X_α. There is substantial freedom in defining these sets, but an easy rule to applying them to (non-sparse) SUGMs is to set the $\Delta(\alpha, N)$ sets to include the other subgraphs with which the subgraph α shares some edges and could have potentially been incidentally generated.

We show that under conditions on the relative correlations inside and outside of the dependency neighborhoods, a central limit theorem applies.

6.2. **The Central Limit Theorem.** Let

$$a_N := \sum_{\alpha;\eta\in\Delta(\alpha,N)} \operatorname{cov}(Z_\alpha, Z_\eta),$$

be the total sum of variance-covariances across all the pairs of variables in each other's dependency neighborhoods, and recall the normalized statistic

$$\overline{S}^N = \frac{S^N}{a_N^{1/2}}.$$

In what follows, we maintain the assumption that $a_N \to \infty$, as otherwise there is insufficient variation to obtain a central limit theorem.

The following are the key conditions for the theorem:

$$\sum_{\alpha;\eta,\gamma\in\Delta(\alpha,N)} \mathrm{E}\left[|Z_\alpha| Z_\eta Z_\gamma\right] = o\left(a_N^{3/2}\right), \tag{6.1}$$

$$\sum_{\alpha,\alpha',\eta\in\Delta(\alpha,N),\eta'\in\Delta(\alpha',N)} \operatorname{cov}(Z_\alpha Z_\eta, Z_{\alpha'} Z_{\eta'}) = o\left(\left(a^N\right)^2\right), \tag{6.2}$$

$$\sum_{\alpha;\eta\notin\Delta(\alpha,N)} \mathrm{E}\left(Z_\alpha Z_\eta \cdot \mathrm{sign}\left(\mathrm{E}[Z_\alpha Z_\eta | Z_\eta]\right)\right) = o\left(a_N\right). \tag{6.3}$$

All conditions are intuitive. Loosely put, Condition (6.1) captures the idea that most of the covariance between random variables in a dependency neighborhood for α is coming from covariances between the reference random variable X_α and a member of the neighborhood X_η, rather than from dependency between two other members X_η and X_γ. Some of them can have high dependency, but in total they cannot. So in constructing our normalizer a_N we need only consider the covariance terms between reference variables and their neighbors.[44]

Condition (6.2) is similar but it looks at the dependency between two members (η,η') of different dependency neighborhoods of two distinct reference nodes (α,α'). It says, again, that the *total* amount of covariance across members of different dependency neighborhoods, when considering any two pairs of reference nodes, is small relative to the total sum of variances.

Condition (6.3) states that covariances between reference nodes and members outside of its dependency neighborhood are relatively small. This is intuitive and motivates the strategy in defining dependency neighborhoods in the first place. Note that if, for instance, $\mathrm{E}[Z_\alpha Z_\eta|Z_\eta] \geq 0$, then Condition (6.3) is simply that $\sum_{\alpha;\eta\notin\Delta(\alpha,N)} \mathrm{cov}\left(Z_\alpha Z_\eta\right) = o\left(a_N\right)$.

Theorem **4.** *If (6.1)-(6.3) are satisfied, then* $\overline{S}^N \rightsquigarrow \mathcal{N}(0,1)$.

It is useful to consider the special case in which $\Delta(\alpha,N) = \{\alpha\}$, which extends but nests many standard central limit theorems. This is useful when we get to the case of sparse networks, where incidental networks are unlikely and the correlation between different subgraphs becomes small.

Corollary **2.** *If* $\mathrm{E}[Z_\alpha Z_\eta|Z_\eta] \geq 0$ *for every* $\eta \neq \alpha$, *and*[45]

(i) $\sum_{\alpha,\eta} \mathrm{cov}(Z_\alpha^2, Z_\eta^2) = o\left(\left(a^N\right)^2\right)$, *and*

(ii) $\sum_{\alpha\neq\eta} \mathrm{cov}(Z_\alpha, Z_\eta) = o\left(a^N\right)$,

then $\overline{S}^N \rightsquigarrow \mathcal{N}(0,1)$.

Moreover, if the X_α*s are Bernoulli random variables with* $\mathrm{E}[X_\alpha] \to 0$ *(uniformly), then (ii) implies (i).*

Note that (i) is often satisfied whenever (ii) is, so this is an easy corollary based on one intuitive condition: the overall sum of covariances between different variables cannot be too large relative to the sum of their variances.

The proof of this Central Limit Theorem applies some key techniques pioneered by Stein (1972, 1986) (see also Bolthausen (1982); Baldi and Rinott (1989); Ross (2011)) to a more general structure than have been analyzed before. Rather than relying on mixing conditions typically applied in the spatial and network settings, which do not apply to SUGMs, we extend Stein's approximation argument to develop bounds that are looser and more permissive, which turns out to be essential.

7. Concluding Remarks

[44] Note that an easy sufficient condition for (6.1) is that $\sum_{\alpha;\eta,\gamma\in\Delta(\alpha,N)} \mathrm{E}\left[X_\alpha X_\eta X_\gamma\right] = o\left(a_N^{3/2}\right)$, since $|Z_\alpha| \leq X_\alpha$.

[45] If $\mathrm{E}[Z_\alpha Z_\eta|Z_\eta] \geq 0$ does not hold, then (ii) can just be substituted by (6.3).

We have developed a new class of models—SUGMs—in which networks are formed via a basis set of subgraphs. The parameters are always identified and we study conditions when the parameters have estimators that are consistent and asymptotically normally distributed. En route, we develop a new central limit theorem for dependent random variables which extends the dependency graph literature and also does not require a geometric (lattice-like) ordering of covariances of the kind used in the time series and spatial literatures. We believe this is of independent interest.

Our model is useful for empirical work. We show that it models economically relevant features of real-world network data better than the standard alternatives: stochastic block models, unobserved heterogeneity models, latent space models, and ERGMs. Further, we have illustrated that it is easy to microfound and use to test important hypotheses such as whether a network provides incentives to sustain informal contracts, whether people are willing to interact across caste publicly, and even how network formation changes in response to the introduction of credit.

Future research could explore, among other things, richer inclusion of covariates in subgraphs, a data-driven approach to select subgraphs for inclusion in the model, statistical properties of other specific empirically-relevant SUGMs not studied here, and systematic bootstrap techniques for inference for use in complex implementations of these models.

References

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015): "Systemic Risk and Stability in Financial Networks," *American Economic Review*, 105(2), 564–608. 1

Andrews, D. (1990): "Generic Uniform Convergence," Tech. rep., Cowles Foundation for Research in Economics, Yale University. C.1, C.2

Badev, A. (2013): "Discrete Games in Endogenous Networks: Theory and Policy," *mimeo: University of Pennsylvania*. 1.1

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2016): "Social Networks and Housing Markets," *unpublished manuscript*. 3.1

Baldi, P. and Y. Rinott (1989): "On normal approximations of distributions in terms of dependency graphs," *The Annals of Probability*, 1646–1650. 1.2, 6, 6.2

Banerjee, A., E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan (2021): "Changes in social network structure in response to exposure to formal credit markets," . 2.4.4, 20, 2.4.4, 5.5

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013): "Diffusion of Microfinance," *Science*, 341, DOI: 10.1126/science.1236498, July 26 2013. 1, 3.1, 5.1, 37

Banerjee, A. V., A. Chandrasekhar, E. Duflo, and M. O. Jackson (2014): "Gossip: Identifying Central Individuals in a Social Network," *Available at SSRN 2425379*. 1.2, 5.1

Barabasi, A. and R. Albert (1999): "Emergence of scaling in random networks," *Science*, 286, 509. 1.1, 2.4.1

Bester, C. A., T. G. Conley, and C. B. Hansen (2011): "Inference with dependent data using cluster covariance estimators," *Journal of Econometrics*, 165, 137–151. 11, 43

Bhamidi, S., G. Bresler, and A. Sly (2008): "Mixing time of exponential random graphs," *Arxiv preprint arXiv:0812.2265*. 1.1

Bhattacharyya, S., P. J. Bickel, et al. (2015): "Subsampling bootstrap of count features of networks," *Annals of Statistics*, 43, 2384–2411. 10

Bickel, P., A. Chen, and E. Levina (2011): "The method of moments and degree distributions for network models," *Annals of Statistics*, 39, 2280–2301. 2

Biscio, C. A. N., A. Poinas, and R. Waagepetersen (2018): "A note on gaps in proofs of central limit theorems," *Statistics & Probability Letters*, 135, 7–10. C.3.2, C.3.2

Blitzstein, J. and P. Diaconis (2011): "A sequential importance sampling algorithm for generating random graphs with prescribed degrees," *Internet mathematics*, 6, 489–522. 3

Blumenstock, J., N. Eagle, and M. Fafchamps (2011): "Charity and Reciprocity in Mobile Phone-Based Giving: Evidence from Rwanda," *mimeo: University of California, Berkeley*. 3.1

Bollobas, B. (2001): *Random Graphs*, Cambridge University Press. 2.4.1, 4.3.1, 4.4.2

Bollobás, B., C. Borgs, J. Chayes, and O. Riordan (2010): "Percolation on dense graph sequences," *The Annals of Probability*, 38, 150–183. 5.2

Bolthausen, E. (1982): "On the central limit theorem for stationary mixing random fields," *Annals of Probability*, 10, 1047–1050. 1.2, 11, 6, 6.2

Boucher, V. and I. Mourifié (2012): "My friend far far away: Asymptotic properties of pairwise stable networks," *Available at SSRN*, 2170803. 1.1, 6

Bramoullé, Y., S. Currarini, M. Jackson, P. Pin, and B. Rogers (2012): "Homophily and Long-Run Integration in Social Networks," *Journal of Economic Theory*, 147, 1754–1786. 1.1, 18

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2017): "Using aggregated relational data to feasibly identify network structure without network data," Tech. rep., National Bureau of Economic Research. 7, 2.4

Cabrales, A., A. Calvó-Armengol, and Y. Zenou (2011): "Social interactions and spillovers," *Games and Economic Behavior*, 72, 339–360. 23

Cai, J., A. deJanvry, and E. Sadoulet (2015): "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 7:2, 81–108. 1

Calvo-Armengol, A. and M. Jackson (2007): "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of Economic Theory*, 132, 27–46. 2.4.1

Calvo-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer Effects and Social Networks in Education," *The Review of Economic Studies*, 76, 1239–1267. 1

Canen, N., M. O. Jackson, and F. Trebbi (2020): "Endogenous Networks and Legislative Activity," *SSRN http://ssrn.com/abstract=2823338*. 23

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013): "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81, 855–882. 1

Chandrasekhar, A. (2015): "Econometrics of Network Formation," *mimeo: Stanford*. 28

Chandrasekhar, A. and M. Jackson (2012): "Tractable and Consistent Random Graph Models," *SSRN Working Paper: http://ssrn.com/abstract=2150428*. 1.1

Chandrasekhar, A. and R. Lewis (2013): "Econometrics of sampled networks," Stanford working paper. 2.4

Chaney, T. (2016): "Networks in international trade," *The Oxford Handbook on the Eco- nomics of Networks, edited by Bramoullé, Yann, Andrea Galeotti and Brian Rogers, Oxford: Oxford University Press*. 1

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

Chatterjee, S., P. Diaconis, and A. Sly (2010): "Random graphs with a given degree sequence," *Arxiv preprint arXiv:1005.1136*. 1.1, 3, 41

Chen, L. H. and Q.-M. Shao (2004): "Normal approximation under local dependence," *The Annals of Probability*, 32, 1985–2028. 1.2, 6

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman (2010): "An Empirical Model for Strategic Network Formation," *NBER Working Paper*. 1.1

Coleman, J. (1988): "Social Capital in the Creation of Human Capital," *American Journal of Sociology*, 94, S95–S120. 2.4

Conley, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of econometrics*, 92, 1–45. 11, 43

Currarini, S., M. Jackson, and P. Pin (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77, 1003–1045. 1.1, 2.4.1, 17, 18, 3.1, 4.3.1

——— (2010): "Identifying the roles of race-based choice and chance in high school friendship network formation," *Proceedings of the National Academy of Sciences*, 107, 4857–4861. 1.1, 2.4.1, 17, 18, 4.3.1

de Paula, A. (2015): "Econometrics of Network Models," *mimeo, to appear in "Advances in Economics and Econometrics, Theory and Applications: Eleventh World Congress of the Econometric Society" Cambridge University Press*. 9

De Paula, Á., S. Richards-Shubik, and E. Tamer (2018): "Identifying preferences in networks with bounded degree," *Econometrica*, 86, 263–288. 1.1

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions*," *Quarterly journal of economics*, 118, 909–968. 5.2

Elliott, M., B. Golub, and M. O. Jackson (2014): "Financial Networks and Contagion," *American Economic Review*, 104(10), 3115–3153. 1

Feigenberg, B., E. Field, and R. Pande (2010): "Building Social Capital through Microfinance," *NBER Working Papers*. 2.4.4, 5.5

Frank, O. and D. Strauss (1986): "Markov graphs," *Journal of the American Statistical Association*, 832–842. 4

Gai, P. and S. Kapadia (2010): "Contagion in financial networks," *Proceedings of the Royal Society A*, 466, 2401–2423. 1

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996): "Crime and Social Interactions," *The Quarterly Journal of Economics*, 111(2), 507–548. 1

Goldstein, L. and Y. Rinott (1996): "Multivariate normal approximations by Stein's method and size bias couplings," *Journal of Applied Probability*, 1–17. 1.2, 6

Golub, B. and M. Jackson (2012): "How Homophily Affects the Speed of Learning and Best-Response Dynamics," *Quarterly Journal of Economics*, 127, 1287–1338. 2.4.1, 5.2

Graham, B. S. (2017): "An econometric model of network formation with degree heterogeneity," *Econometrica*, 85, 1033–1063. 1.1, 3, 4.4.2, 5.2, 41

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002): "Latent space approaches to social network analysis," *Journal of the american Statistical association*, 97, 1090–1098. 1.1, 5.2, 6

Holland, P. W. and S. Leinhardt (1981): "An exponential family of probability distributions for directed graphs: Rejoinder," *Journal of the American Statistical Association*, 76, 62–65. 3

Jackson, M. (2008): *Social and economic networks*, Princeton: Princeton University Press. 1.1, 4.3, 4.4.2, 5.2

Jackson, M. and B. Rogers (2007): "Meeting strangers and friends of friends: How random are social networks?" *The American economic review*, 97, 890–915. 1.1, 6, 2.4.1

Jackson, M. and A. Watts (2001): "The Existence of Pairwise Stable Networks," *Seoul Journal of Economics*, 14(3), 299–321. 1.1

Jackson, M. and A. Wolinsky (1996): “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory*, 71, 44–74. 2.4.2

Jackson, M. O. (2005): “A Survey of Models of Network Formation: Stability and Efficiency,” *Group Formation in Economics; Networks, Clubs and Coalitions, ed. G Demange, M Wooders. Cambridge, UK: Cambridge Univ. Press*. 1.1

——— (2019): *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*, Pantheon. 1

Jackson, M. O. and S. Nei (2015): “Networks of Military Alliances, Wars, and International Trade,” *Proceedings of the National Academy of Sciences*, 112(50), 15277–15284. 1

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (2012): “Social Capital and Social Quilts: Network Patterns of Favor Exchange,” *American Economic Review*, 102, 1857–1897. 1.2, 2.4, 2.4.2, 2.4.4, 5.5

Jackson, M. O., B. W. Rogers, and Y. Zenou (2016): “The Economic Consequences of Social Network Structure,” *Journal of Economic Literature (forthcoming)*. 1

Jenish, N. and I. R. Prucha (2009): “Central limit theorems and uniform laws of large numbers for arrays of random fields,” *Journal of Econometrics*, 150, 86–98. 11, 43, C.2

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009): “Trust and Social Collateral*,” *Quarterly Journal of Economics*, 124, 1307–1361. 2.4.1

Koenig, M., D. Rohner, M. Thoenig, and F. Zilibotti (2015): “Networks in Conflict: Theory and Evidence from the Great War of Africa,” *SSRN: http://ssrn.com/abstract=2704419*. 1

Krackhardt, D. (1988): “Predicting with Networks: Nonparameteric Multiple Regression Analysis of Dyadic Data,” *Social Networks*, 10, 359–381. 4

Leung, M. (2014): “A Random-Field Approach to Inference in Large Models of Network Formation,” *Stanford Working Paper*. 1.1, 6

Lubold, S., A. G. Chandrasekhar, and T. H. McCormick (2020): “Identifying the latent space geometry of network models through analysis of curvature,” Tech. rep., National Bureau of Economic Research. 1.1

McCormick, T. H. and T. Zheng (2015): “Latent surface models for networks using Aggregated Relational Data,” *Journal of the American Statistical Association*, 110, 1684–1695. 7

McFadden, D. (1973): “Conditional logit analysis of qualitative choice behavior,” *Institute of Urban and Regional Development, University of California*. 2.4.2

McPherson, M., L. Smith-Lovin, and J. Cook (2001): “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–444. 2.4.1

Mele, A. (2017a): “A structural model of Dense Network Formation,” *Econometrica*, 85, 825–850. 5, 1.1

——— (2017b): “A structural model of homophily and clustering in social networks,” . 5

Mele, A. and L. Zhu (2017): “Approximate variational estimation for a model of network formation,” *arXiv preprint arXiv:1702.00308*. 5

Munshi, K. and M. Rosenzweig (2006): “Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy,” *The American Economic Review*, 96, 1225–1252. 2.4.3

Newey, W. and D. McFadden (1994): “Large sample estimation and hypothesis testing,” *Handbook of econometrics*, 4, 2111–2245. 4.2, A, C.3.3

Newman, M., S. Strogatz, and D. Watts (2001): “Random graphs with arbitrary degree distributions and their applications,” *Physical Review E*, 64, 26118. 4.4.2

Newman, M. E. (2003): "The Structure and Function of Complex Networks," *SIAM Review*, 45:2, 167–256. 4.3

Park, J. and M. E. Newman (2004): "Statistical mechanics of networks," *Physical Review E*, 70, 066117. 3

Patacchini, E. and Y. Zenou (2008): "The strength of weak ties in crime," *European Economic Review*, 52, 209–236. 1

Penrose, M. (2003): *Random Geometric Graphs*, Oxford University Press. 1.1, 6

Pötscher, B. and I. Prucha (1997): *Dynamic Nonlinear Econometric Models*, Springer-Verlag, New York. C.1, C.2

Pötscher, B. M. and I. R. Prucha (1991): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models," *Econometric Reviews*, 10, 253–325. C.1

Ross, N. (2011): "Fundamentals of Stein's method," *Probab. Surv*, 8, 210–293. 6, 6.2, B.1, B.2

Shalizi, C. and A. Rinaldo (2012): "Consistency under Sampling of Exponential Random Graph Models," *ArXiv 1111.3054v3*. 1.1

Sheng, S. (2013): "Identification and Estimation of Network Formation Games," *mimeo*. 1.1

Stein, C. (1972): "A bound for the error in the normal approximation of a sum of dependent random variables," *In: Proceedings of Berkeley Symposium M.S.P.*, 2, 583–603. 11, 6, 6.2

——— (1986): "Approximate computation of expectations," *Lecture Notes-Monograph Series*, 7, i–164. 1.2, 6, 6.2, B.1, B.1

Wasserman, S. and P. Pattison (1996): "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp," *Psychometrika*, 61, 401–425. 4

Watts, D. and S. Strogatz (1998): "Collective dynamics of small-world networks," *Nature*, 393, 440–442. 2.4.1

Appendix A. Proofs

Proof of Lemma 1. Note that for $z \in \{favor, info\}$, $\frac{\gamma_{T,z}}{\gamma_{L,z}} = \frac{\beta_{T,z}^{1/3}}{\beta_{L,z}^{1/2}}$ and so the condition becomes $\frac{\beta_{T,favor}^{1/3}}{\beta_{L,favor}^{1/2}} > \frac{\beta_{T,info}^{1/3}}{\beta_{L,info}^{1/2}}$ from which the result directly follows. ∎

Proof of Lemma 2. Having two randomly picked nodes bump into each other within a community, there is a $f^2+(1-f)^2$ probability of the nodes being of the same type, and a $1-(f^2+(1-f)^2)$ probability of them being of different types.[46] Thus, the relative meeting frequency of different type links compared same type links is

$$\frac{\pi_L(diff)}{\pi_L(same)} = \frac{1-(f^2+(1-f)^2)}{f^2+(1-f)^2}.$$

For triangles, picking three individuals out of the community at any point in time would lead to a $f^3+(1-f)^3$ probability that all three are of the same type, and $1-(f^2+(1-f)^2)$ of them being of mixed types, and so

$$\frac{\pi_T(diff)}{\pi_T(same)} = \frac{1-(f^3+(1-f)^3)}{f^3+(1-f)^3}.$$

It follows directly that for $f \in (0,1)$:

$$\frac{\pi_T(same)}{\pi_T(diff)} < \frac{\pi_L(same)}{\pi_L(diff)}. \tag{A.1}$$

So different type triangles are more likely to have opportunities to form under this random mixing model than different type links. In particular, note that $\frac{p_{T,diff}}{p_{T,same}} < \frac{p_{L,diff}}{p_{L,same}}$ if and only if $\left(\frac{\beta_{T,diff}}{\beta_{T,same}}\frac{\pi_T(diff)}{\pi_T(same)}\right)^{1/3} < \left(\frac{\beta_{L,diff}}{\beta_{L,same}}\frac{\pi_L(same)}{\pi_L(diff)}\right)^{1/2}$. In summary, given (A.1), sufficient condition for $\frac{p_{T,diff}}{p_{T,same}} < \frac{p_{L,diff}}{p_{L,same}}$ is that $\frac{\beta_{T,diff}}{\beta_{T,same}} < \left(\frac{\beta_{L,diff}}{\beta_{L,same}}\right)^{3/2}$. ∎

Proof of Theorem 1. Order subgraph types so that the number of links a subgraph of type ℓ is nondecreasing in ℓ. Let ℓ^* be the smallest ℓ for which $\beta_\ell \neq \beta'_\ell$.

Consider a particular (labeled) subgraph $g' \in G_{\ell^*}$. Let $p_\beta(g')$ denote the probability that the subgraph g' (without any extra links) forms from some collection of subgraphs in G_ℓ for $\ell < \ell^*$. We can then write the probability of forming the subgraph g' as

$$p_\beta(g') + (1-p_\beta(g'))\beta_{\ell^*},$$

where recall that β_{ℓ^*} is the probability that g' forms directly. Let $no_\beta(g')$ denote the probability that all $g'' \in G_\ell$ for $\ell < \ell^*$ such that $g'' \subset g'$ *do not* form. Then the probability that none of the links in g' are present as parts of subgraphs that do not extend beyond g' is then

$$no_\beta(g')(1-\beta_{\ell^*}).$$

Let $\emptyset$ denote the empty network. It then follows that

$$\frac{\mathrm{P}_\beta(g')}{\mathrm{P}_\beta(\emptyset)} = \frac{p_\beta(g') + (1-p_\beta(g'))\beta_{\ell^*}}{no_\beta(g')(1-\beta_{\ell^*})}.$$

[46]To keep things simple, we consider equal-sized groups, but the argument extends with some adjustments to asymmetric sizes.

So the probability that the realized network is exactly g' compared to the probability that it is the empty network, depends only on the probability that g' forms directly or incidentally from subgraphs of it, over the probability that no subgraph of g' (including itself) forms.

Note that this expression is strictly increasing in β_{ℓ^*} since $p_\beta(g') < 1$ and $no_\beta(g') > 0$. By the definition of ℓ^*: $p_\beta(g') = p_{\beta'}(g') < 1$ and $no_\beta(g') = no_\beta(g')$. It then follows that

$$\frac{\mathrm{P}_\beta(g')}{\mathrm{P}_\beta(\emptyset)} \neq \frac{\mathrm{P}_{\beta'}(g')}{\mathrm{P}_{\beta'}(\emptyset)},$$

which establishes the claim. ■

Proof of Proposition 1. First, note that $1-(1-\beta_T^n)^x$ is the probability that some link is formed as part of at least one triangle out of x possible triangles that could have it as an edge (independently of whether it also forms directly).

Next, note that the probability that a link forms conditional on some particular triangle that it could be a part of *not forming* is[47]

$$\widetilde{q}_L = \beta_L + (1-\beta_L)\left(1-(1-\beta_T)^{n-3}\right). \tag{A.2}$$

Given this, note that the probability that a link forms can be written as

$$q_L = \beta_T + (1-\beta_T)\widetilde{q}_L, \tag{A.3}$$

noting that a link could form as part of a triangle that it is part of, or else form conditional upon that triangle not forming.

We can write the probability of some triangle forming as

$$q_T = \beta_T + (1-\beta_T)(\widetilde{q}_L)^3, \tag{A.4}$$

where the first expression β_T is the probability that the triangle is directly generated, and then the second expression $(1-\beta_T)(\widetilde{q}_L)^3$ is the probability that it was not generated directly, but instead all three of the edges formed on their own (which happen independently, conditional on the triangle not forming, which has probability $(\widetilde{q}_L)^3$). The result follows from Lemma A.1, with $x_1 = \beta_L$, $x_2 = \beta_T$, $q_L = a_1(x)$, $q_T = a_2(x)$ and $\widetilde{q}_L = f(x)$. ■

Lemma **A.1.** *Let $x = (x_1, x_2) \in (0,1)^2$ and $a(x) = (a_1(x), a_2(x))$ be two real-valued functions*

$$\begin{aligned} a_1(x) &= x_2 + (1-x_2) f(x) \\ a_2(x) &= x_2 + (1-x_2) f(x)^3, \end{aligned}$$

with

$$f(x) = x_1 + (1-x_1)\left[1-(1-x_2)^N\right] = 1-(1-x_1)(1-x_2)^N$$

for some integer $N \geq 0$. Then $x \neq x' \implies a(x) \neq a(x')$.

Proof. Suppose the contrary. Then

$$x_2' + (1-x_2') f(x') = x_2 + (1-x_2) f(x) \text{ and } x_2' + (1-x_2') f(x')^3 = x_2 + (1-x_2) f(x)^3.$$

First, note that if $x_2' = x_2$, then since these are both less than one, the first equation above implies that $f(x') = f(x)$. However, that is not possible since f is increasing in x_1 and $x_1' \neq x_1$ -

[47]That is, consider a given pair of nodes i, j and a third node k. Consider the probability that link ij is formed conditional on triangle ijk not forming directly as a triangle.

recalling that $x' \neq x$ and $x_2' = x_2$. Thus, $x_2' \neq x_2$, and so without loss of generality consider the case in which $x_2' < x_2$. This implies that both

$$f(x') = bf(x) + c$$

and

$$f(x')^3 = bf(x)^3 + c,$$

where $b = \frac{1-x_2}{1-x_2'} \in (0,1)$ and $c = \frac{x_2 - x_2'}{1-x_2'} \in (0,1)$, and $b + c = 1$.

This implies that

$$bf(x)^3 + 1 - b = (bf(x) + 1 - b)^3.$$

This as an equation of the form

$$by^3 + 1 - b = (by + 1 - b)^3$$

where $b \in (0,1)$ and $y \in (0,1)$. Note that the left hand side is larger when $y = 0$ and the two are equal when $y = 1$, and that the derivative of the difference is

$$3by^2 - 3b(by + 1 - b)^2 = 3b\left[y^2 - (by + 1 - b)^2\right] < 0.$$

The difference is decreasing over the entire interval, and hits 0 at the end. Thus, the difference is always positive in $(0,1)$ and there is no solution, meaning our supposition was incorrect. ■

Lemma **A.2.** *Any event (in the discrete σ-algebra generated by all possible realizations of all subgraphs) associated with any SUGM has a probability that is an analytic function (and so it is in C^∞), and has derivatives and cross partials at all levels being uniformly continuous and bounded on the whole parameter space of $[0,1]^k$.*

Proof. An 'outcome' is a specification of exactly which subgraphs form and which do not - so a complete specification of what happens. Any event then corresponds to a set of outcomes, and so its probability is a sum of probabilities of the outcomes. Each outcome's probability is of the form

$$\prod_\ell \beta_\ell^{z_\ell}(1 - \beta_\ell)^{m_\ell - z_\ell}$$

where z_ℓ indicates how many subgraphs of type ℓ are present in the outcome. As each of these functions is analytic (and hence in C^∞), all of the derivatives and partials, cross partials, etc., are continuous and bounded on $[0,1]^k$ and hence uniformly continuous on $[0,1]^k$. Any event is then a finite sum of analytic functions and so the result follows directly. ■

Proof of Theorem 2. We verify the conditions of Theorem 2.5 of Newey and McFadden (1994) for consistency. Assumption (i) holds by Theorem 1 and we assume compactness of the parameter space (Assumption (ii)). Continuity of $\log \mathrm{P}_\beta(g)$ at each β with probability one is mechanical in our model since subgraph probabilities are analytic functions of the parameters (Lemma A.2). Finally, the uniform bound of assumption (iv) holds since n is fixed, there are only a finite number of graphs in consideration, each with assigned probabilities in a compact set of parameters, and there is a positive probability of seeing any graph in $\mathcal{G}^n$. Therefore, the supremum must be finite.

We verify the conditions of Theorem 3.3 of Newey and McFadden (1994) for asymptotic normality. We have assumed (i), interiority of the parameter, and our model by construction places positive mass on all of $\mathcal{G}^n$. We have assumed (iv). Lemma A.2 implies Assumptions (ii), (iii), and (v). Because all events have probabilities that are analytic functions of parameters, with all derivatives and cross-partials being uniformly continuous and bounded in the parameter space, the norms of the maximal derivative ($\|\nabla_\beta \mathrm{P}_\beta(g)\|$) and second derivatives ($\|\nabla_{\beta\beta} \mathrm{P}_\beta(g)\|$) of the probability functions, as well as the log likelihood ($\|\nabla_\beta \log \mathrm{P}_\beta(g)\|$), have uniform and finite upper bounds. ■

Proof of Proposition 2. First we check consistency by the conditions of Theorem 2.6 of Newey and McFadden (1994). Here each observation is an independently drawn network. For Assumption (i) let $\widehat{W}$ be the identity matrix and then apply Proposition 1. For (ii), we have assumed that the parameter space $\mathcal{B}$ is compact. (iii) follows from the fact that $\mathrm{E}_\beta\left[S\left(g_r\right)\right]$ is continuous at each β with probability one since it composes continuous functions of parameter entries. Finally (iv) follows from the fact that since both S_ℓ are shares, they are strictly less than 1.

Next we check asymptotic normality by the conditions of Theorem 3.4 of Newey and McFadden (1994). Since we have β_0 in the interior of the compact parameter space, so (i) is met. We see (iii) holds since by definition the subgraph counts are fractions between 0 and 1. Both (ii) (that the empirical moment function is continuously differentiable in a neighborhood of the true parameter) and (iv) (that the gradient of the moment function is continuous at the true parameter and that it satisfies a ULLN) follow from Lemma A.2. Analytic functions are C^∞, so there are arbitrarily many derivatives. Finally, for (v), that HH' is non-singular follows from the linear independence of rows of H. ■

Proof of Theorem 3. When obvious, we omit superscript n's to simplify notation, but they are implicit. It follows that,

$$\widetilde{\beta}_\ell = \left(\frac{S_\ell^{true}}{\kappa_\ell \binom{n}{m_\ell}} + \frac{\widetilde{S}_\ell^{true} - S_\ell^{true}}{\kappa_\ell \binom{n}{m_\ell}} + \frac{\widetilde{S}_\ell(g) - \widetilde{S}_\ell^{true}}{\kappa_\ell \binom{n}{m_\ell}} \right) \tag{A.5}$$

where S_ℓ^{true} is the number of truly generated such subgraphs (unobserved) on the whole network, and $\widetilde{S}_\ell^{true}$ is the number of truly generated such subgraphs (unobserved) on the networks that the after removing the links in $D_\ell(g) = \{ij : \ ij \in g', g' \subset g, g' \in G_{\ell'}, \ell' < \ell\}$, and $\binom{n}{m_\ell}$ counts the number of ways to pick m_ℓ nodes out of n.

We show below that $|\widetilde{S}_\ell^{true} - S_\ell^{true}| = o_p(S_\ell^{true})$ and $|\widetilde{S}_\ell(g) - \widetilde{S}_\ell^{true}| = o_p(\widetilde{S}_\ell^{true})$; which then also implies that $\widetilde{S}_\ell(g) - \widetilde{S}_\ell^{true} = o_p(S_\ell^{true})$. Together with (A.5), these tell us that

$$\widetilde{\beta}_\ell = \left(\frac{S_\ell^{true}}{\kappa_\ell \binom{n}{m_\ell}} \right) (1 + o_p(1)). \tag{A.6}$$

Note that $S_\ell^{true}(g)$ has a binomial distribution with parameter $\beta_{0,\ell}^n$. From this and (A.6), it then follows that

$$\widetilde{\beta} - \beta_0^n \rightsquigarrow \mathcal{N}(0, \sigma_\ell)$$

where $\sigma_\ell = \left(\frac{\beta_{0,\ell}^n (1 - \beta_{0,\ell}^n)}{\kappa_\ell \binom{n}{m_\ell}} \right)^{1/2}$.

Next, note that the $S_\ell^{true}(g)$ are independent across ℓ. From (A.6) it then follows that

$$\sum_\ell \alpha_\ell \widetilde{\beta}_\ell = \sum_\ell \alpha_\ell \frac{S_\ell^{true}}{\kappa_\ell \binom{n}{m_\ell}} (1 + o_p(1))$$

for any $\alpha \in [0,1]^k$, with $\sum_\ell \alpha_\ell = 1$. Given the independence of S_ℓ^{true} across ℓ, it then follows that the random variable on the right hand side converges to being normal.[48] Then, by the Cramér-Wold

[48]Note that under the assumption that $m_\ell > h_\ell$ there are a growing number of observations of each subgraph.

Theorem, this implies that the $\widetilde{\beta}_\ell$ are jointly normally distributed in the limit, and so

$$\Sigma^{-1/2}(\widetilde{\beta} - \beta_0^n) \rightsquigarrow \mathcal{N}(0, I)$$

where $\Sigma_{\ell,\ell} = \frac{\beta_{0,\ell}^n(1-\beta_{0,\ell}^n)}{\kappa_\ell \binom{n}{m_\ell}}$ and the off-diagonals are all 0.

Thus, to complete the proof we show that $|\widetilde{S}_\ell^{true} - S_\ell^{true}| = o_p(S_\ell^{true})$ and $|\widetilde{S}_\ell(g) - \widetilde{S}_\ell^{true}| = o_p(\widetilde{S}_\ell^{true})$.

To establish these claims, we establish two facts. One is that the probability that some observed subgraph of type ℓ was incidentally generated (by subgraphs that are no larger than it in the ordering) is $o_p(1)$. This establishes that $|\widetilde{S}_\ell(g) - \widetilde{S}_\ell^{true}| = o_p(\widetilde{S}_\ell^{true})$. The other is that a truly formed subgraph of type ℓ becomes part of an incidentally generated subgraph of type $\ell' < \ell$ is $o_p(1)$. This establishes that $|\widetilde{S}_\ell^{true} - S_\ell^{true}| = o_p(S_\ell^{true})$.

Let z_ℓ^n denote the probability that any given $g' \in G_\ell^n$ is incidentally generated. We now show that $z_\ell^n/\beta_{0,\ell}^n = o(1)$, which establishes the first claim. Consider $g_\ell \in G_\ell^n$ and a (minimal, ordered) generating subclass $\mathcal{C} = (\ell_j, c_j)_{j\in J}$, and for which $\ell_j \geq \ell$ fr all j.

We show that the probability z_ℓ^n that it is generated by this subclass goes to zero relative to $\beta_{0,\ell}^n$, and since there are at most $M_\ell \leq k^{m_\ell}$ such generating classes, this implies that $z_\ell^n/\beta_{0,\ell}^n \to 0$.

Consider a subnetwork in $G_{\ell_j}^n$. The probability of getting at least one such network that has the c_j nodes out of the m_ℓ in g_ℓ is no more than

$$\kappa_{\ell_j}\binom{n}{m_{\ell_j} - c_j}\beta_{0,\ell_j}^n \leq \kappa_{\ell_j} n^{m_{\ell_j} - c_j}\beta_{0,\ell_j}^n.$$

Then, we can bound the desired ratio by

$$\frac{z_\ell^n}{\beta_{0,\ell}^n} \leq \frac{\prod_{j\in J} n^{m_{\ell_j}-c_j}\kappa_{\ell_j}\beta_{0,\ell_j}^n}{\beta_{0,\ell}^n} \leq \frac{\prod_{j\in J} n^{m_{\ell_j}}\kappa_{\ell_j}\beta_{0,\ell_j}^n}{n^{\sum_j c_j}\beta_{0,\ell}^n} = \frac{n^{\sum_{j\in J} m_{\ell_j} - h_{\ell_j} - c_j}\kappa_{\ell_j} b_{0,\ell_j}}{n^{-h_\ell} b_{0,\ell}} \to 0,$$

where the last convergence is guaranteed by (4.3).

The second claim follows from a similar calculation. It is sufficient to show that the probability that some subgraph of type $\ell_{j'}$ becomes part of a subgraph of type $\ell < \ell_{j'}$ (where $j' \in J$ is part of a generating class of some $\ell < \ell_{j'}$), compared to the likelihood of the formation of a subgraph of type $\ell_{j'}$, is of vanishing order. Again, as there are a finite number of larger subgraphs, and a finite number of generating classes, it is sufficient to show this for a generic $\ell < \ell_{j'}$ and generic generating class. In the following, the numerator is on the order of the expected number of incidentally formed subgraphs of type ℓ from this type of generating class, while the denominator is the expected number of the subgraphs of type ℓ.

$$\frac{\kappa_\ell\binom{n}{m_\ell}\prod_{j\in J} n^{m_{\ell_j}-c_j}\kappa_{\ell_j}\beta_{0,\ell_j}^n}{\kappa_{\ell_{j'}}\binom{n}{m_{j'}}\beta_{0,\ell_{j'}}^n} = \Theta\left(\frac{n^{m_{\ell_j}-c_j-h_j}}{n^{h_{\ell_{j'}}}}\right) \to 0,$$

where the convergence to 0 follows from (4.4).

Finally, by multiplying and dividing by n^{h_ℓ} and collecting terms, it follows that $|\widetilde{b} - b_0| \xrightarrow{\text{P}} 0$ and $V_n^{-1/2}\left(\widetilde{b} - b_0\right) \rightsquigarrow \mathcal{N}(0, I)$. To see this, observe that $\Sigma^{-1/2}(\widetilde{\beta} - \beta_0^n) = V_n^{-1/2}\left(\widetilde{b} - b_0\right)$. ■

Proof of Corollary 1. Note that $\sum_j c_j \geq m_\ell + (|C|-1)z$ for some $z \geq 1$, where $z \geq 2$ if subgraphs are acyclic (each subgraph in the incidental set overlaps the others with at least one node, and at least two if the subgraphs are acyclic). The conditions then simplify directly. ■

Appendix B. Proof of Central Limit Theorem 4 and Corollary 2

B.1. **Stein's Lemma.** Our proof uses a lemma from Stein (1986). We review it here, both to be self-contained and also to explain why this approach to proving asymptotic normality is useful and distinct from other approaches in the networks literature. The key observation of Stein (1986) is that if a random variable satisfies

$$\mathrm{E}[f'(Y) - Yf(Y)] = 0$$

for every $f(\cdot)$ that is continuously differentiable, then it must have a standard normal distribution.

This observation leads to a useful lemma, that allows one to characterize the Kolmogorov distance between a random variable Y and a standard normally distributed Z, denoted $d_K(Y, Z)$. We can bound this from above by (a constant times) the Wasserstein distance, $d_W(Y, Z)$, which itself is bounded by the below expression.

Lemma **B.1** (Stein (1986); Ross (2011))**.** *If Y is a random variable and Z has the standard normal distribution, then*

$$d_W(Y, Z) \leq \sup_{\{f:||f||,||f''||\leq 2,||f'||\leq\sqrt{2/\pi}\}} \left|\mathrm{E}[f'(Y) - Yf(Y)]\right|.$$

Further $d_K(Y, Z) \leq (2/\pi)^{1/4}(d_W(Y, Z))^{1/2}$.

By this lemma, if we show that a normalized sum of random variables satisfies

$$\sup_{\{f:||f||,||f''||\leq 2,||f'||\leq\sqrt{2/\pi}\}} \left|\mathrm{E}[f'(\overline{S}^N) - \overline{S}^N f(\overline{S}^N)]\right| \to 0,$$

then $d_W(\overline{S}^N, Z) \to 0$, and so it must be asymptotically normally distributed.

B.2. **Proofs of Theorem 4 and Corollary 2.** The following lemmas are useful in the proof.

Lemma **B.2.** *A solution to* $\max_h \mathrm{E}[Zh(Y)]$ *s.t.* $|h| \leq 1$ *(where h is measurable) is* $h(Y) = \mathrm{sign}(\mathrm{E}[Z|Y])$, *where we break ties, setting* $\mathrm{sign}(\mathrm{E}[Z|Y]) = 1$ *when* $\mathrm{E}[Z|Y] = 0$.

Proof. This can be seen from direct calculation:

$$\mathrm{E}[Zh(Y)] = \int_Y \mathrm{E}[Z|Y]h(Y)d\mathrm{P}(Y)$$

Maximizing $\mathrm{E}[Z|Y]h(Y)$ pointwise when $|h| \leq 1$ is achieved by setting $h(Y) = \mathrm{sign}(\mathrm{E}[Z|Y])$, and we break ties by setting $\mathrm{sign}(\mathrm{E}[Z|Y]) = 1$ when $\mathrm{E}[Z|Y] = 0$, as that makes no difference in the integral. ■

Lemma **B.3.** $\mathrm{E}[XYh(Y)]$ *when $h(\cdot)$ is measurable and bounded by $\sqrt{\frac{2}{\pi}}$ satisfies*

$$\mathrm{E}[XYh(Y)] \leq \sqrt{\frac{2}{\pi}}\mathrm{E}\left[XY \cdot \mathrm{sign}(\mathrm{E}[X|Y]Y)\right].$$

Proof. This follows from Lemma B.2, setting $Z = XY$. ■

Proof of Theorem 4. By Lemma B.1, it is sufficient to show that the appropriate sequence of random variables $\overline{S}^N$ satisfies

$$\sup_{\{f:||f||,||f''||\leq 2,||f'||\leq\sqrt{2/\pi}\}} \left|\mathrm{E}[f'(\overline{S}^N) - \overline{S}^N f(\overline{S}^N)]\right| \to 0.$$

Recall

$$a_N = \sum_{\alpha,\eta\in\Delta(\alpha,N)} \operatorname{cov}\left(Z_\alpha, Z_\eta\right),$$

and

$$\overline{S}^N = S^N / a_N^{1/2}.$$

For ease of notation, we omit the superscript Ns below. Let

$$S_\alpha := \sum_{\eta\notin\Delta(\alpha,N)} Z_\eta \text{ and } \overline{S}_\alpha := S_\alpha / a^{1/2}.$$

Observe that

$$\mathrm{E}\left[\overline{S} f\left(\overline{S}\right)\right] = \mathrm{E}\left[\frac{1}{a^{1/2}}\sum_\alpha Z_\alpha \cdot f\left(\overline{S}\right)\right] = \mathrm{E}\left[\frac{1}{a^{1/2}}\sum_\alpha Z_\alpha \left(f\left(\overline{S}\right) - f\left(\overline{S}_\alpha\right)\right)\right] + \mathrm{E}\left[\frac{1}{a^{1/2}}\sum_\alpha Z_\alpha \cdot f\left(\overline{S}_\alpha\right)\right].$$

The first step is to show that

$$\left|\mathrm{E}\left[\frac{1}{a^{1/2}}\sum_\alpha Z_\alpha \cdot f\left(\overline{S}_\alpha\right)\right]\right| = o(1),$$

by employing condition (6.3).

In order to do this, we can expand the term to

$$\begin{aligned}
\left|\mathrm{E}\left[\frac{1}{a_N^{1/2}}\sum_{\alpha\in\Lambda} Z_\alpha \cdot f\left(\overline{S}_\alpha\right)\right]\right| &= \left|\mathrm{E}\left[\frac{1}{a_N^{1/2}}\sum_{\alpha\in\Lambda} Z_\alpha \cdot f\left(\frac{1}{a_N^{1/2}}\sum_{\eta\notin\Delta(\alpha,N)} Z_\eta\right)\right]\right| \\
&\leq \underbrace{\left|\mathrm{E}\left[\frac{1}{a_N^{1/2}}\sum_{\alpha\in\Lambda} Z_\alpha \cdot f\left(0\right)\right]\right|}_{=0 \text{ since } \mathrm{E}[Z_\alpha]=0.} \\
&\quad + \left|\mathrm{E}\left[\frac{1}{a_N^{1/2}}\sum_{\alpha\in\Lambda} Z_\alpha \cdot \left(\frac{1}{a_N^{1/2}}\sum_{\eta\notin\Delta(\alpha,N)} Z_\eta\right) \cdot f'\left(\widehat{\overline{S}_\alpha}\right)\right]\right|
\end{aligned}$$

where $\widehat{\overline{S}_\alpha}$ is an intermediate value between $\overline{S}_\alpha$ and 0.

To bound the second term, we apply Lemma B.3 to conclude that

$$\left|\frac{\mathrm{E}\left[\sum_{\alpha\in\Lambda;\eta\notin\Delta(\alpha,N)} Z_\alpha Z_\eta f'\left(\widehat{\overline{S}_\alpha}\right)\right]}{a_N}\right| \leq \sqrt{\frac{2}{\pi}}\left|\frac{\mathrm{E}\left[\sum_{\alpha\in\Lambda;\eta\notin\Delta(\alpha,N)} Z_\alpha Z_\eta \cdot \operatorname{sign}\left(\mathrm{E}\left[Z_\alpha Z_\eta \middle| \; Z_\eta\right]\right)\right]}{a_N}\right|.$$

Thus, it is sufficient that

$$\mathrm{E}\left[\sum_{\alpha\in\Lambda;\eta\notin\Delta(\alpha,N)} Z_\alpha Z_\eta \cdot \operatorname{sign}\left(\mathrm{E}\left[Z_\alpha Z_\eta | Z_\eta\right]\right)\right] = o(a_N) \tag{B.1}$$

to ensure that

$$\left|\frac{\mathrm{E}\left[\sum_{\alpha\in\Lambda;\eta\notin\Delta(\alpha,N)} Z_\alpha \cdot Z_\eta \cdot f'\left(\widehat{\overline{S}_\alpha}\right)\right]}{a_N}\right| = o(1),$$

which is ensured by (6.3).

Next, the second step of the proof is to apply a similar reasoning as in Ross (2011) with an $o(1)$ adjustment (from the first step above), to write

$$\left|\mathrm{E}\left[f'(\overline{S})-\overline{S}f(\overline{S})\right]\right| \leq \left|\mathrm{E}\left[\frac{1}{a^{1/2}}\sum_{\alpha} Z_\alpha(f(\overline{S})-f(\overline{S}_\alpha)-(\overline{S}-\overline{S}_\alpha)f'(\overline{S})\right]\right| + \left|\mathrm{E}\left[f'(\overline{S})\left(1-\frac{1}{a^{1/2}}\sum_{\alpha} Z_\alpha(\overline{S}-\overline{S}_\alpha)\right)\right]\right| + o(1),$$

and then to show that the right hand side of this expression goes to 0.

By a Taylor series approximation and given the bound on the derivatives of f, it follows that

$$\left|\mathrm{E}\left[f'(\overline{S})-\overline{S}f(\overline{S})\right]\right| \leq \frac{||f''||}{2a^{1/2}}\sum_{\alpha}\mathrm{E}\left[|Z_\alpha|\left(\overline{S}-\overline{S}_\alpha\right)^2\right] + \left|\mathrm{E}\left[f'(\overline{S})\left(1-\frac{1}{a^{1/2}}\sum_{\alpha} Z_\alpha(\overline{S}-\overline{S}_\alpha)\right)\right]\right| + o(1).$$

Let us denote the first two terms on the right hand side as A_1 and A_2 respectively. We bound each, and show that each is $o(1)$, which then completes the proof.

$$A_1 = \frac{||f''||}{2a^{3/2}}\sum_{\alpha}\mathrm{E}\left[|Z_\alpha|\left(\sum_{\eta\in\Delta(\alpha,N)} Z_\eta\right)^2\right] = \frac{||f''||}{2a^{3/2}}\sum_{\alpha;\eta\in\Delta(\alpha,N),\gamma\in\Delta(\alpha,N)}\mathrm{E}\left[|Z_\alpha|Z_\eta Z_\gamma\right] = o(1),$$

where the last equality follows from (6.1).

Next,

$$\begin{aligned}
A_2 &= \left|\mathrm{E}\left[f'(\overline{S})\left(1-\frac{1}{a^{1/2}}\sum_{\alpha} Z_\alpha(\overline{S}-\overline{S}_\alpha)\right)\right]\right| = \frac{1}{a}\left|\mathrm{E}\left[f'(\overline{S})\left(a-\sum_{\alpha,\eta\in\Delta(\alpha,N)} Z_\alpha Z_\eta\right)\right]\right| \\
&\leq \frac{||f'||}{a}\mathrm{E}\left|\left(a-\sum_{\alpha,\eta\in\Delta(\alpha,N)} Z_\alpha Z_\eta\right)\right| = \frac{||f'||}{a}\mathrm{E}\left|\left(\sum_{\alpha,\eta\in\Delta(\alpha,N)} Z_\alpha Z_\eta - \mathrm{E}\left[Z_\alpha Z_\eta\right]\right)\right| \\
&\leq \frac{\sqrt{2}}{a\sqrt{\pi}}\left(\mathrm{var}\left[\sum_{\alpha,\eta\in\Delta(\alpha,N)} Z_\alpha Z_\eta\right]\right)^{1/2} = \frac{\sqrt{2}}{a\sqrt{\pi}}\left(\sum_{\alpha,\alpha',\eta\in\Delta(\alpha,N),\eta'\in\Delta(\alpha',N)}\mathrm{cov}\left(Z_\alpha Z_\eta, Z_{\alpha'}Z_{\eta'}\right)\right)^{1/2},
\end{aligned}$$

where the last inequality follows by Cauchy-Schwarz. The final expression is $o(1)$ by (6.2). ■

Proof of Corollary 2. We apply Theorem 4 to the case in which $\Delta(\alpha,N)=\{\alpha\}$. (6.1) becomes

$$\sum_{\alpha}\mathrm{E}\left[|Z_\alpha|^3\right] = o\left(\left(\sum_{\alpha}\mathrm{var}\left(Z_\alpha\right)\right)^{3/2}\right)$$

which becomes

$$\sum_{\alpha}\mathrm{var}\left(Z_\alpha\right)^{3/2} = o\left(\left(\sum_{\alpha}\mathrm{var}\left(Z_\alpha\right)\right)^{3/2}\right),$$

which is satisfied directly, given that $\sum_\alpha \mathrm{var}\left(Z_\alpha\right)$ is growing without bound.

(i) and (ii) correspond directly to (6.2) and (6.3) respectively (noting that the sign is always positive).

We now show that for Bernoulli random variables with uniformly vanishing means, (i) holds whenever (ii) holds. Observe that

$$\mathrm{cov}\left(Z_\alpha^2, Z_\eta^2\right) = \mathrm{cov}\left(\left(X_\alpha-\mu_\alpha\right)^2, \left(X_\eta-\mu_\eta\right)^2\right)$$

$$= \text{cov}\left(X_\alpha^2 - 2X_\alpha\mu_\alpha + \mu_\alpha^2, X_\eta^2 - 2X_\eta\mu_\eta + \mu_\eta^2\right)$$
$$= \text{cov}\left(X_\alpha^2, X_\eta^2\right) - 2\mu_\alpha \text{cov}\left(X_\alpha, X_\eta^2\right) - 2\mu_\eta \text{cov}\left(X_\alpha^2, X_\eta\right) + 4\mu_\alpha\mu_\eta \text{cov}\left(X_\alpha, X_\eta\right).$$

Because they are Bernoulli, $\text{cov}\left(X_\alpha^k, X_\eta^{k'}\right) = \text{cov}\left(X_\alpha, X_\eta\right)$ for any $k, k' > 0$. Since the means tend to zero, this means

$$\text{cov}\left(Z_\alpha^2, Z_\eta^2\right) = \text{cov}\left(X_\alpha, X_\eta\right)\left(1 + o\left(1\right)\right).$$

Therefore satisfying (ii) implies (i) (noting also that $a_N \geq 1$ so $a_N^2 \geq a_N$). ■

For Online Publication Only

Appendix C. Proof of Proposition 3

Let the moment (normalized) be

$$\widehat{M}(\beta) = R_n S(g) - \mathrm{E}_\beta \left[R_n S(g)\right],$$

where $R_n = \operatorname{diag}\left\{n^{h_L}, n^{h_T}\right\}$ properly normalizes the moments. So, for example, for links we have

$$\widehat{M}_L^n(\beta) = \frac{n^{h_L}}{\binom{n}{2}} \sum_{i<j} \left\{g_{ij} - \mathrm{E}_\beta g_{ij}\right\} = \frac{n^{h_L}}{\binom{n}{2}} \sum_{i<j} g_{ij} - q_L(\beta)\, n^{h_L}.$$

The objective function is

$$\widehat{Q}_n(g, \beta) := \widehat{M}^n(\beta)' \, \widehat{M}^n(\beta).$$

And we need

$$\overline{Q}_n(\beta) = \mathrm{E}\left[\widehat{M}^n(\beta)\right]' \mathrm{E}\left[\widehat{M}^n(\beta)\right],$$

which is the non-stochastic analogue.

C.1. **Identification.** We prove identification for sequences of parameters, in the sense of identifiable uniqueness in the sense of Lemma 3.1 of Pötscher and Prucha (1997). See also "Assumption ID" in Andrews (1990) and Pötscher and R. Prucha (1991). The parameters β_0^n are identifiably unique in the sense that for any $\varepsilon > 0$

$$\operatorname{liminf}_{n\to 0}\left[\inf_{\beta\in\mathcal{B}:\ \delta\left(\beta,\beta_0^n\right)>\varepsilon} \left|\overline{Q}^n(\beta) - \overline{Q}^n(\beta_0^n)\right|\right] > 0.$$

We take the usual Euclidean metric $\|\hat{b} - b_0\|$ to calculate distance between two vectors $\hat{b}, b_0$. Note that, since in our setting while b_0 is uniformly bounded from above and below in n, β_0^n—the subgraph probability vector—has entries that tend to zero at hypothesized rates.

It is useful to note that in our setting, not only will we show that $\|\hat{b} - b_0\| \xrightarrow{\mathrm{P}} 0$ but in fact for a metric $\delta(\cdot,\cdot)$, we have $\delta(\hat{\beta}, \beta_0) \xrightarrow{\mathrm{P}} 0$. In fact, the former follows from the latter mechanically.

To see why this is useful, first consider the degenerate estimator $\hat{\beta}^n = 0$ and observe $\|0 - \beta_0^n\| \xrightarrow{\mathrm{P}} 0$. That is, for a sequence of models in which the probability of any given subgraph tends to zero—mechanically true in any sparse random graph model—by definition the zero vector is a consistent estimator for the probability parameters, though this is informative.

The right metric for this sequence is to set[49]

$$\delta(x, y) := \max_\ell \left[\frac{|x_\ell - y_\ell|}{\max(|x_\ell|, |y_\ell|)}\right], \tag{C.1}$$

then the requirement becomes

$$\delta(\hat{\beta}^n, \beta_0^n) = \max_\ell \frac{\left|\widehat{\beta}_\ell^n - \beta_{0,\ell}^n\right|}{\max(\left|\widehat{\beta}_\ell^n\right|, \left|\beta_{0,\ell}^n\right|)} \xrightarrow{\mathrm{P}} 0.$$

This requires that $\widehat{\beta}_\ell^n$ and $\beta_{0,\ell}^n$ be proportional to each other far enough along the sequence. Thus, if β_0^n approaches 0, saying that $\widehat{\beta}_\ell^n$ is a good estimate of it under this metric also requires that $\widehat{\beta}_\ell^n$ approach 0 at the same rate, which is a much stronger conclusion than just requiring that the two parameters converge in the usual Euclidean metric.

[49] We take $0/0 = 0$.

Returning to our degenerate estimator $\widehat{\beta}_\ell^n = 0$, note

$$\delta(0, \beta_0^n) = \max_\ell \frac{n_\ell^h \cdot \left|0 - \beta_{0,\ell}^n\right|}{b_{0,\ell}} = \max_\ell \frac{|0 - b_{0,\ell}|}{b_{0,\ell}} = 1$$

which does not tend to zero; the δ metric tells us that this an inconsistent estimator.

Finally, if $\delta(\widehat{\beta}, \beta^n) \xrightarrow{\text{P}} 0$ then $|\widehat{b} - b^n| \xrightarrow{\text{P}} 0$ so any proof of consistency in β-space implies and so the results in the paper follow as corollaries to the results below. To see this observe

$$\begin{aligned}\delta\left(\widehat{\beta}, \beta^n\right) &= \max_\ell \frac{\left|\widehat{\beta}_\ell - \beta_\ell^n\right|}{\max\left(\left|\widehat{\beta}_\ell\right|, |\beta_\ell^n|\right)} = \max_\ell \frac{n^{h_\ell}\left|\widehat{\beta}_\ell - \beta_\ell^n\right|}{n^{h_\ell}\max\left(\left|\widehat{\beta}_\ell\right|, |\beta_\ell^n|\right)} \\ &= \max_\ell \frac{\left|\widehat{b}_\ell - b_\ell^n\right|}{\max\left(\left|\widehat{b}_\ell\right|, |b_\ell^n|\right)} \geq \max_\ell \frac{\left|\widehat{b}_\ell - b_\ell^n\right|}{\overline{D}}.\end{aligned}$$

since by assumption b_ℓ lives in a compact set with maximum $\overline{D}$. Since $\delta\left(\widehat{\beta}, \beta^n\right) \xrightarrow{\text{P}} 0$ then so must $\max_\ell \frac{\left|\widehat{b}_\ell - b_\ell^n\right|}{\overline{D}}$, proving the result.

Proposition C.1. *Consider a links and triangles SUGM with associated parameters $\beta_{0,L}^n, \beta_{0,T}^n = \left(\frac{b_{0,L}}{n^{h_L}}, \frac{b_{0,T}}{n^{h_T}}\right)$ with $h_L \in \left(\frac{1}{2}, 2\right)$ and $h_T \in [h_L + 1, 3h_L]$, with $h_T < 3$. Then $\beta_{0,L}^n, \beta_{0,T}^n$ are identifiably unique.*

Proof of Proposition C.1.

Write[50]

$$\beta^n = \left(\frac{b_L^n}{n^{h_L}}, \frac{b_T^n}{n^{h_T}}\right) \qquad \beta_0^n = \left(\frac{b_{0,L}^n}{n^{h_L}}, \frac{b_{0,T}^n}{n^{h_T}}\right),$$

where $b_L^n, b_T^n, b_{0,L}^n, b_{0,T}^n$ lie in $[\underline{D}, \overline{D}]$

Let $r_L^n = 1/n^{h_L}$ and $r_T^n = 1/n^{h_T}$.

First, note that $1 - (1 - \beta_T^n)^x$ is the probability that some link is formed as part of at least one triangle out of x possible triangles that could have it as an edge (independently of whether it also forms directly).

Next, note that the probability that a link forms conditional on some particular triangle that it could be a part of *not forming* is[51]

$$\widetilde{q}_L^n = \beta_L^n + (1 - \beta_L^n)\left(1 - (1 - \beta_T^n)^{n-3}\right). \tag{C.2}$$

So, we can write the probability of some triangle forming as

$$q_T^n := \mathrm{E}_{\beta_L^n, \beta_T^n}\left[S_T(g)\right] = \beta_T^n + (1 - \beta_T^n)(\widetilde{q}_L^n)^3, \tag{C.3}$$

where the first expression β_T^n is the probability that the triangle is directly generated, and then the second expression $(1 - \beta_T^n)(\widetilde{q}_L^n)^3$ is the probability that it was not generated directly, but instead all three of the edges formed on their own (which happen independently, conditional on the triangle not forming, which has probability $(\widetilde{q}_L^n)^3$).

[50]We allow the constants to depend on n to capture that some applications have both rates and constants that adjust with scale, and we may want to fit across data of networks of varying sizes. But this is largely semantic, as estimating any particular network has only one b, and one can ignore the superscripts on the bs if one likes.

[51]That is, consider a given pair of nodes i, j and a third node k. Consider the probability that link ij is formed conditional on triangle ijk not forming directly as a triangle.

It is useful to note that since $\beta_L^n = o(1)$, $(1-\beta_L^n) \to 1$ and since $h_T > 1$, $|(1-\beta_T^n)^{n-3} - (1 - \frac{b_{0,T}^n}{n^{h_T-1}})| \to 0$. Thus,[52]

$$\widetilde{q}_L^n = \Theta\left(\frac{1}{n^{h_L}} + \frac{1}{n^{h_T-1}}\right) = \Theta\left(\frac{1}{n^{h_L}}\right)$$

where the second equality follows since $h_T \geq h_L + 1$.

Next, note that the probability that a link forms is

$$q_L^n := \mathrm{E}_{\beta_L^n,\beta_T^n}\left[S_L(g)\right] = \beta_L^n + (1-\beta_L^n)\left(1-(1-\beta_T^n)^{n-2}\right), \tag{C.4}$$

where the first expression β_L^n is the probability that the link is directly generated, and then the second expression $(1-\beta_L^n)\left(1-(1-\beta_T^n)^{n-2}\right)$ is the probability that it was not generated directly, but instead appeared as an edge in some triangle (and there are $n-2$ such possible triangles).

It is also useful to write this in a very different way:

$$q_L^n := \mathrm{E}_{\beta_L^n,\beta_T^n}\left[S_L(g)\right] = \beta_T^n + (1-\beta_T^n)\widetilde{q}_L^n, \tag{C.5}$$

noting that a link could form as part of a triangle that it is part of, or else form conditional upon that triangle not forming.

The following derivative expressions are useful:

$$\frac{\partial \widetilde{q}_L^n}{\partial \beta_L^n} = (1-\beta_T^n)^{n-3} \qquad \frac{\partial \widetilde{q}_L^n}{\partial \beta_T^n} = (n-3)(1-\beta_L^n)(1-\beta_T^n)^{n-2}. \tag{C.6}$$

$$\frac{\partial q_L^n}{\partial \beta_L^n} = (1-\beta_T^n)^{n-2}.$$

$$\frac{\partial q_T^n}{\partial \beta_L^n} = 3(1-\beta_T^n)(\widetilde{q}_L^n)^2\frac{\partial \widetilde{q}_L^n}{\partial \beta_L^n} = 3(\widetilde{q}_L^n)^2(1-\beta_T^n)^{n-2}.$$

$$\frac{\partial q_L^n}{\partial \beta_T^n} = 1-\widetilde{q}_L^n + (1-\beta_T^n)\frac{\partial \widetilde{q}_L^n}{\partial \beta_T^n} = 1-\widetilde{q}_L^n + (n-3)(1-\beta_L^n)(1-\beta_T^n)^{n-1}.$$

$$\frac{\partial q_T^n}{\partial \beta_T^n} = 1-(\widetilde{q}_L^n)^3 + 3(1-\beta_T^n)(\widetilde{q}_L^n)^2\frac{\partial \widetilde{q}_L^n}{\partial \beta_T^n} = 1-(\widetilde{q}_L^n)^3 + 3(\widetilde{q}_L^n)^2(n-3)(1-\beta_L^n)(1-\beta_T^n)^{n-1}.$$

Given that $\beta_L^n = o(1)$ (since $h_L > 0$), $\beta_T^n = o(1/n)$ (since $h_T > 1$), and $\widetilde{q}_L^n = \Theta\left(\frac{1}{n^{h_L}}\right)$ the above expressions imply that:

$$\frac{\partial q_L^n}{\partial \beta_L^n} = 1 - o(1), \tag{C.7}$$

$$\frac{\partial q_T^n}{\partial \beta_L^n} = \Theta\left(\frac{1}{n^{2h_L}}\right), \tag{C.8}$$

$$\frac{\partial q_L^n}{\partial \beta_T^n} = n-2-o(1), \tag{C.9}$$

$$\frac{\partial q_T^n}{\partial \beta_T^n} = \Theta\left(\max[1, n^{1-2h_L}]\right). \tag{C.10}$$

[52]We use Bachmann-Landau notation so $f(n) = \Theta(g(n))$ means that f is bounded above and below asymptotically by g. That is, $\exists k_1 > 0, \exists k_2 > 0, \exists n_0$ such that $\forall n > n_0$, $k_1 g(n) \leq f(n) \leq k_2 g(n)$.

Note that (C.7)-(C.10) hold for any parameters $h_L > 0$ and $3h_L > h_T \geq h_L + 1$ - and thus uniformly for any β^n in a compact set of such h_L, h_T, and thus as long as we restrict attention to β^n in that compact set, we have the same order derivatives and so then we approximate:

$$\text{(C.11)} \qquad \frac{\mathrm{E}_{\beta^n}[S_L(g)] - \mathrm{E}_{\beta_0^n}[S_L(g)]}{r_L^n} \approx n^{h_L}\left[\frac{b_L^n - b_{0,L}^n}{n^{h_L}} + (n-2)\frac{b_T^n - b_{0,T}^n}{n^{h_T}}\right]$$
$$\approx b_L^n - b_{0,L}^n + (b_T^n - b_{0,T}^n)\Theta(n^{h_L+1-h_T}),$$

and

$$\text{(C.12)} \qquad \frac{\mathrm{E}_{\beta^n}[S_T(g)] - \mathrm{E}_{\beta_0^n}[S_T(g)]}{r_T^n} \approx n^{h_T}\left[\frac{b_L^n - b_{0,L}^n}{n^{h_L}}\Theta(1/n^{2h_L}) + \frac{b_T^n - b_{0,T}^n}{n^{h_T}}\Theta\left(\max[1, n^{1-2h_L}]\right)\right]$$
$$\approx (b_L^n - b_{0,L}^n)\Theta(n^{h_T-3h_L}) + (b_T^n - b_{0,T}^n)\Theta\left(\max[1, n^{1-2h_L}]\right).$$

To establish identifiable uniqueness (given the additive separability of $\overline{Q}^n(\beta)$ across L, T) it is sufficient to argue that for any $\varepsilon > 0$ there exists $\phi > 0$ such that for large enough n, if $\delta((\beta_L^n, \beta_T^n), (\beta_{0,L}^n, \beta_{0,T}^n)) > \varepsilon$, then at least one of the following inequalities holds:

$$\text{(C.13)} \qquad \left|\frac{\mathrm{E}_{\beta^n}[S_L(g)] - \mathrm{E}_{\beta_0^n}[S_L(g)]}{r_L^n}\right| > \phi$$

or

$$\text{(C.14)} \qquad \left|\frac{\mathrm{E}_{\beta^n}[S_T(g)] - \mathrm{E}_{\beta_0^n}[S_T(g)]}{r_T^n}\right| > \phi.$$

Note that $\delta((\beta_L^n, \beta_T^n), (\beta_{0,L}^n, \beta_{0,T}^n)) > \varepsilon$ translates into $|b_L^n - b_{0,L}^n| > c\varepsilon$ and/or $|b_T^n - b_{0,T}^n| > c\varepsilon$ for some $c > 0$. If the second inequality holds, then by (C.12) it follows that (C.14) holds. If (C.14) does not hold for any ϕ, then by (C.12) it must be that $|b_L^n - b_{0,L}^n| > c\varepsilon$ while $|b_T^n - b_{0,T}^n| < \delta^n$ for a sequence $\delta^n \to 0$. In that case, noting that since $h_T \geq h_L - 1$ (and so the second term of (C.11) is of order at most 1 times δ^n while the first term is at least $c\varepsilon$ in magnitude), then by (C.11) it follows that (C.13) holds. ■

C.2. **Consistency.**

Proposition **C.2.** *Consider a links and triangles SUGM with associated parameters* $\beta_{0,L}^n, \beta_{0,T}^n = \left(\frac{b_{0,L}}{n^{h_L}}, \frac{b_{0,T}}{n^{h_T}}\right)$ *with* $h_L \in \left(\frac{1}{2}, 2\right)$ *and* $h_T \in [h_L + 1, 3h_L]$, *with* $h_T < 3$. *Then* $\delta(\widehat{\beta}, \beta_0^n) \xrightarrow{\mathrm{P}} 0$ *and therefore* $\|\hat{b}^n - b\| \xrightarrow{\mathrm{P}} 0$.

Proof of Proposition C.2. The proof follows from checking the conditions of Lemma 3.1 of Pötscher and Prucha (1997) (see also Jenish and Prucha (2009)) or equivalently Andrews (1990), Lemma 6. Clearly $\mathcal{B}$ is compact, the weighting function is the identity matrix so it is positive semi-definite, and the moment function is continuous in β. Identifiable uniqueness was demonstrated in Proposition C.1. Uniform convergence remains.

Observe is that this just requires showing

$$\sup_\beta \left|\widehat{M}^n(\beta) - E\widehat{M}^n(\beta)\right| = o_p(1)$$

as

$$\sup_\beta \left|\widehat{Q}_n(g, \beta) - \overline{Q}_n(\beta)\right| \leq \sup_\beta \left|\widehat{M}^n(\beta)' \widehat{M}^n(\beta) - \mathrm{E}\left[\widehat{M}^n(\beta)\right]' E\left[\widehat{M}^n(\beta)\right]\right|$$

$$\leq \sup_{\beta}\left|\left\{\widehat{M}^{n}(\beta)-\mathrm{E}\left[\widehat{M}^{n}(\beta)\right]\right\}'\widehat{M}^{n}(\beta)\right|+\sup_{\beta}\left|\mathrm{E}\left[\widehat{M}^{n}(\beta)\right]'\left\{\widehat{M}^{n}(\beta)-\mathrm{E}\left[\widehat{M}^{n}(\beta)\right]\right\}\right|$$
$$\leq 2K\cdot\sup_{\beta}\left|\widehat{M}^{n}(\beta)-\mathrm{E}\left[\widehat{M}^{n}(\beta)\right]\right|$$

for a constant K, recalling we have assumed $\underline{D}_L < b_L < \overline{D}_L$ and $\underline{D}_T < b_T < \overline{D}_T$.

So, we show that $\sup_{\beta}\left|\widehat{M}^{n}(\beta)-\mathrm{E}\widehat{M}^{n}(\beta)\right| = o_p(1)$. It is enough to show pointwise convergence, which is clear by inspection, and stochastic equicontinuity.

Stochastic equicontinuity requires that for any $\epsilon>0$, there exists $\eta>0$ such that

$$\limsup_{n}\mathrm{P}\left\{\sup_{\beta}\sup_{\beta'\in\delta(\beta,\beta')<\eta}\left|\widehat{M}^{n}(\beta)-\widehat{M}^{n}(\beta')\right|>\epsilon\right\}<\epsilon$$

as in Andrews (1990) and others. A sufficient condition is a Lipschitz condition: for every β,β',

$$\left|\widehat{M}(\beta)-\widehat{M}(\beta')\right| = O_p(1)\cdot\delta(\beta,\beta').$$

Recall that $\Delta = h_T - h_L$. It is also useful to note (see the proof of Proposition C.1) that

$$\left|q_L^n(\beta)-q_L^n(\beta')\right|\leq(1+o(1))\left|\beta_L-\beta_L'\right|+\Theta(n)\left|\beta_T-\beta_T'\right|$$

and

$$\left|q_T^n(\beta)-q_T^n(\beta')\right|\leq\Theta\left(n^{-2h_L}\right)\left|\beta_L-\beta_L'\right|+\Theta(1)\left|\beta_T-\beta_T'\right|.$$

Returning to the moments computation:

$$\begin{aligned}
\left|\widehat{M}(\beta)-\widehat{M}(\beta')\right| &\leq n^{h_L}\left|q_L(\beta)-q_L(\beta')\right|+n^{h_T}\left|q_T(\beta)-q_T(\beta')\right|\\
&\leq(1+o(1))\left|\beta_L-\beta_L'\right|n^{h_L}+\Theta(n)\left|\beta_T-\beta_T'\right|n^{h_L}\\
&+\Theta\left(n^{-2h_L}\right)\left|\beta_L-\beta_L'\right|n^{h_T}+\Theta(1)\left|\beta_T-\beta_T'\right|n^{h_T}\\
&\leq\Theta(1)\,\delta_L(\beta_L,\beta_L')+\Theta\left(n^{1-\Delta}\right)\delta_T(\beta_T,\beta_T')\\
&+\Theta\left(n^{h_T-3h_L}\right)\delta_L(\beta_L,\beta_L')+\Theta(1)\,\delta_T(\beta_T,\beta_T')\\
&\leq\Theta(1)\,\delta(\beta,\beta')
\end{aligned}$$

since $\Delta\geq 1$ and $h_T\leq 3h_L$, the result follows. ■

C.3. **Asymptotic Normality.** In what follows, begin with the restrictions required for identification and consistency which are $h_L\in\left(\frac{1}{2},2\right)$ and $h_T\in[h_L+1,3h_L]$, with $h_T<3$. Asymptotic normality will require further tightening of the restriction, though a more general treatment of the arguments will show more permissive rates to get normality of the shares, but would violate identification and consistency requirements for the minimum distance estimator.

C.3.1. *Asymptotic Normality of Link and Triangle Shares.* Let $Y_L := \sum_{i<j} g_{ij}$ denote the sum of links which takes the place of S in our general CLT (since we have used S_L to denote share above).

LEMMA **C.1.** *Assume the rate requirements for identifiable uniqueness and consistency. Then*

$$a_L^{-1/2}\left(Y_L-\mathrm{E}\left[Y_L\right]\right)\rightsquigarrow\mathcal{N}(0,1)$$

if $h_L\in\left(\frac{2}{3},2\right)$ *and* $h_T\in[h_L+1,3h_L]$, *with* $h_T<3$.

Proof. We apply the main theorem where α indexes a link ij. We define the dependency neighborhood $\Delta(\alpha, N) := \{\eta : \eta \cap \alpha \neq \emptyset\}$, so $\Delta\left(ij, \binom{n}{2}\right) = \{ij\} \cup \{ik : k \neq i, j\} \cup \{k : k \neq i, j\}$. Therefore dependency neighborhoods include all node-adjacent links.

Condition (6.3) is obvious from the definition of $\Delta\left(ij, \binom{n}{2}\right)$, because if ij and kl do not share nodes, no triangle nor link can generate both. Thus they are independent and the left-hand side term is 0.

Next, we verify Condition (6.1) as follows, using the sufficient condition from footnote 41.

$$\mathrm{E}\left|X_\alpha X_\eta X_\gamma\right| = \mathrm{P}\left(X_\alpha X_\eta X_\gamma\right) \quad \text{s.t.} \quad \eta, \gamma \in \Delta(\alpha, N).$$

We have three cases where all indices are distinct and two cases where at least two indices are identical. Enumerating them, we have

(1) ij, jk, il (a line) - there are $O\left(n^4\right)$ of these.
(2) ij, ik, il (a star) - there are $O\left(n^4\right)$ of these.
(3) ij, jk, ik (a triangle) - there are $O\left(n^3\right)$ of these.
(4) ij, ij, ik or ij, jk, jk (two repeat) - there are $O\left(n^3\right)$ of these.
(5) ij, ij, ij (all repeat) - there are $O\left(n^2\right)$ of these.

From the proof of identification, recall q_L is the probability of a link forming in the graph, which can be due to a link forming directly or as a part of a triangle. Also recall that $\widetilde{q}_L$ is the probability of a link forming if a particular triangle that it could be a part of does not form. Finally, let $\widetilde{q}_L'$ denote the probability that a link forms conditional on two triangles, that it could be part a part of, not forming. Note that we have, and will continue to, suppress the dependence on n unless explicitly needed.

We can construct loose upper bounds on the probabilities of the various structures:

(1) Line: $\beta_{0,T}^2 + 2\left(1-\beta_{0,T}\right)\beta_{0,T}\widetilde{q}_L + \left(1-\beta_{0,T}\right)^2 \widetilde{q}_L^2 \widetilde{q}_L' \leq \beta_{0,T}^2 + 2\beta_{0,T} q_L + q_L^3$.
(2) Star:

$$\beta_{0,T}^3 + 3\left(1-\beta_{0,T}\right)\beta_{0,T}^2 + 3\left(1-\beta_{0,T}\right)^2 \beta_{0,T}\widetilde{q}_L' + \left(1-\beta_{0,T}\right)^3 \left(\widetilde{q}_L'\right)^3 \leq 4\beta_{0,T}^2 + 3\beta_{0,T} q_L + q_L^3.$$

(3) Triangle: $\beta_{0,T} + \left(1-\beta_{0,T}\right)\left(\widetilde{q}_L'\right) \leq \beta_{0,T} + q_L^3$.
(4) Two repeat: $\beta_{0,T} + \left(1-\beta_{0,T}\right)\left(\widetilde{q}_L'\right)^2 \leq \beta_{0,T} + q_L^2$.
(5) All repeat: $q_{L.}$

Then it follows that

$$\mathrm{E}\left|X_\alpha X_\eta X_\gamma\right| \leq \Theta\left(n^4\left(\beta_{0,T}^2 + \beta_{0,T} q_L + q_L^3\right) + n^3\left(\beta_{0,T} + q_L^2\right) + n^2 q_L\right), \tag{C.15}$$

where we omit the dominated term from triangles.

It is straightforward to show that for $k \neq i$, by binomial approximation and bounds on lower order terms

$$\operatorname{cov}\left(X_{ij}, X_{jk}\right) = \beta_{0,T}\left(1-\beta_{0,T}\right)\left(1-\widetilde{q}_L\right)^2 \leq O\left(\beta_{0,T}\right).$$

Then it follows directly that

$$a_N = \Theta\left(n^2 q_L + n^3 \beta_{0,T}\right).$$

For the sufficient condition for (6.1) we need to compare this to the bound on $\mathrm{E}\left|X_\alpha X_\eta X_\gamma\right|$ from (C.15) and show

$$\mathrm{E}\left|X_\alpha X_\eta X_\gamma\right| = o\left(a_N^{3/2}\right) = o\left(n^3 q_L^{3/2} + n^{9/2}\beta_{0,T}^{3/2}\right),$$

and so we need to show

$$n^4\left(\beta_{0,T}^2 + \beta_{0,T} q_L + q_L^3\right) + n^3\left(\beta_{0,T} + q_L^2\right) + n^2 q_L = o\left(n^3 q_L^{3/2} + n^{9/2}\beta_{0,T}^{3/2}\right).$$

This imposes a number of constraints, and omitting the parts that are obviously satisfied (e.g., $n^4\beta_{0,T}^2 + n^3 q_L^2 = o\left(n^3 q_L^{3/2} + n^{9/2}\beta_{0,T}^{3/2}\right)$), this reduces to

$$n^4\left(\beta_{0,T} q_L + q_L^3\right) + n^3\left(\beta_{0,T}\right) + n^2 q_L = o\left(n^3 q_L^{3/2} + n^{9/2}\beta_{0,T}^{3/2}\right). \tag{C.16}$$

Recall in addition $h_L > 1/2$ and $h_T \in [h_L + 1, 3h_L]$ from the identification and consistency requirements. Noting that, as in the proof of identification, (working there with $\widetilde{q}_L$ which is of the same order)

$$q_L^n = \Theta\left(n^{-h_L} + n^{-(h_T-1)}\right) = \Theta\left(n^{-h_L}\right)$$

since $h_T - 1 \geq h_L$. The condition (C.16) is then satisfied if we can show that

$$n^{4-h_T-h_L} + n^{4-3h_L} + n^{3-h_T} + n^{2-h_L} < n^{3-(3/2)h_L} + n^{9/2-(3/2)h_T}. \tag{C.17}$$

Since $h_T \geq h_L + 1$,

$$n^{4-h_T-h_L} + n^{4-3h_L} + n^{3-h_T} + n^{2-h_L} \leq n^{3-3h_L} + n^{4-3h_L} + n^{2-h_L} + n^{2-h_L}$$

and thus to show (C.17) it is enough to show that

$$\max\left\{4 - 3h_L, 2 - h_L\right\} < 3 - (3/2)\, h_L,$$

which holds since $2/3 < h_L < 2$ (which exactly correspond to the crossing points).

Next we turn to Condition (6.2). We will show that this is implied by the above restrictions. To do this, we compute terms of the form

$$\operatorname{cov}\left(\left(g_{ij} - q_L\right)\left(g_{ik} - q_L\right), \left(g_{rs} - q_L\right)\left(g_{st} - q_L\right)\right)$$

since $\eta \in \Delta(\alpha, N)$ and $\eta' \in \Delta(\alpha', N)$; here we allow for the cases that $k = i$ and $r = t$. Iterating on expectations, one can show that

$$\operatorname{cov}\left(\left(g_{ij} - q_L\right)\left(g_{ik} - q_L\right), \left(g_{rs} - q_L\right)\left(g_{st} - q_L\right)\right) \leq \mathrm{E}\left[g_{ij} g_{jk} g_{rs} g_{st}\right].$$

It is easy to see that if $\{i, j, k\} \cap \{r, s, t\} = \emptyset$ then the covariance is zero since the events are independent. Thus, we are summing over the cases in which the intersection is non-empty. The cases with intersection of two or more nodes are handled as we already did above, noting that the condition here is less restrictive ($a_N > 1$ so $a_N^2 > a_N^{3/2}$).

So we restrict attention to the case where there is only one node of intersection. In this case the intersection could come from:

(1) $s = j$, so two-stars joined at the center,
(2) $r = i$, so a line,
(3) $s = i$, so the center of one star is attached to the leaf of the other.

These exhaust all configurations up to a relabeling.

Consider the event $g_{ij} g_{jk} g_{rs} g_{st} = 1$. Assume we are in case 1. This has the highest probability relative to the other two cases, so we can construct a crude bound using this to finish the result. The probability is of order no more than

$$\beta_{0,T}^3 + \beta_{0,T}^2 + \beta_{0,T}^2 q_L + \beta_{0,T} q_L^2 + q_L^4.$$

Therefore we check

$$n^5\left(\beta_{0,T}^3 + \beta_{0,T}^2 + \beta_{0,T}^2 q_L + \beta_{0,T} q_L^2 + q_L^4\right) = o\left(n^4 q_L^2 + n^6 \beta_{0,T}^2\right).$$

The relevant rate on the right-hand side is n^{4-2h_L}, since $h_T \geq h_L + 1$. Dividing both sides by n^4, the inequality boils down to four conditions:

(1) $1 - 3h_T < -2h_L$ or $1 + 2h_L < 3h_T$, which is implied by $h_T \geq h_L + 1$;
(2) $1 - 2h_T < -2h_L$ or $h_L + \frac{1}{2} < h_T$ which is also implied by $h_T \geq h_L + 1$;
(3) $1 - 2h_T - h_L < -2h_L$ or $1 + h_L < 2h_T$, which is also implied by $h_T \geq h_L + 1$;
(4) $1 - 4h_L < -2h_L$ or $h_L > 1/2$, which is implied by $h_L > 2/3$.

This concludes the proof. ■

Let $Y_T := \sum_{i<j<k} g_{ijk}$ denote the sum of links which takes the place of S in our general CLT (since we have used S_T to denote share above).

Lemma **C.2.** *Assume the rate requirements for identifiable uniqueness and consistency. Then*

$$a_T^{-1/2}\left(Y_T - \mathrm{E}\left[Y_T\right]\right) \rightsquigarrow \mathcal{N}\left(0,1\right)$$

if $h_L \in \left(\frac{2}{3}, 2\right)$ *and* $h_T \in \left[h_L + 1, 3h_L\right]$, *with* $h_T < 3$.

Proof. For this proof we appeal to Corollary 2. Here we set $\Delta\left(ijk, \binom{n}{3}\right) = \{ijk\}$.

We begin with (ii) since it implies (i) for Bernoullis, which is equivalent to showing

$$\sum_{\alpha \neq \eta} \mathrm{cov}\left(X_\alpha, X_\eta\right) = o\left(N \cdot \mathrm{var}\left(X_\alpha\right)\right).$$

Applying the calculation in the proof of Proposition C.1,

$$\mathrm{var}\left(X_\alpha\right) = q_T\left(1 - q_T\right) = \Theta\left(\beta_{0,T}\right).$$

We calculate the covariances for the various cases of α, η and check when they are of lesser order. We have two relevant cases: where the two indices intersect on one node and when they intersect on two nodes. By independence if they do not intersect at all, the covariance is zero.

(1) $|\alpha \cap \eta| = 1$: $\mathrm{cov}\left(X_\alpha, X_\eta\right) = \Theta\left(\beta_{0,T} q_L^4\right)$ and there are $O\left(n^5\right)$ of these.

The triangles are node-adjacent. Since at least one link needs to form together, this can only happen if the joint node is part of a triangle and neither of the triangles formed directly. This gives $\beta_{0,T}^4 \tilde{q}_L^4 = O\left(\beta_{0,T} q_L^4\right)$.

(2) $|\alpha \cap \eta| = 2$: $\mathrm{cov}\left(X_\alpha, X_\eta\right) = \Theta\left(\beta_{0,T} q_L^2 + q_L^5\right)$ and there are $O\left(n^4\right)$ of these.

The triangles are edge-adjacent. This is because we need the common link from each triangle to have formed together and not have already formed independently in both cases, which can happen only if exactly one of the triangles formed directly and the other did not, or else neither triangle formed and all of the links have to form. So, this is of order $\beta_{0,T} \tilde{q}_{L:}^2 + \left(\tilde{q}_L'\right)^5 \leq \beta_{0,T} q_L^2 + q_L^5$.

Each of these must be of order $o\left(n^3 \beta_{0,T}\right)$ for the result to hold. The conditions therefore are

(1) $5 - h_T - 4h_L < 3 - h_T$ or $1/2 < h_L$, which is satisfied since $h_L > 2/3$.
(2) $4 - h_T - 2h_L < 3 - h_T$ or $1/2 < h_L$ which is the same as above.
(3) $4 - 5h_L < 3 - h_T$ or $h_T < 5h_L - 1$. But notice that $3h_L < 5h_L - 1$ so long as $h_L > 1/2$, and so this is implied by $h_T < 3h_L$

This completes the proof. ■

C.3.2. *Joint Asymptotic Normality.* Let $Y = (Y_L, Y_T)'$, the vector of the sums of links and triangles.

Lemma **C.3.** *If* $h_L \in \left(\frac{2}{3}, 2\right)$ *and* $h_T \in \left[h_L + 1, 3h_L\right]$, *with* $h_T < 3.$, *then*

$$a^{-1/2}\left(Y - \mathrm{E}\left[Y\right]\right) \rightsquigarrow \mathcal{N}\left(0, I\right)$$

where a *is the variance-covariance matrix defined below.*

Proof. We will apply the Cramér-Wold device to show joint normality through showing all weighted (normalized) sums are normally distributed. Specifically, Lemma 2.1 of Biscio et al. (2018) contains a useful generalization which we use.

Let $\Lambda = \{ijk: \ i,j,k \in [1:n]\} \cup \{ij: \ i,j \in [1:n]\}$. This consists of a set of $\binom{n}{2} + \binom{n}{3}$ terms. Notice that the set has two types of random variables coming from links and triangles. We now alter the dependency set for the sake of links (for triangles they may remain the same). Specifically

$$\Delta(ij;N) := \{ik: \ \forall k\} \cup \{jk: \ \forall k\} \cup \{irs \text{ and } jrs: \ \forall r,s\}.$$

Let

$$a = \begin{pmatrix} a_L & a_{LT} \\ a_{LT} & a_T \end{pmatrix}.$$

where the two diagonal variance terms have been studied

(1) $a_L = \sum_{ij,rs \in \Delta(ij,N)} \operatorname{cov}(X_{ij}, X_{rs})$,
(2) $a_T = \sum_{ijk} \operatorname{var}(X_{ijk})$, and
(3) $a_{LT} = \sum_{ij.rst \in \Delta(ij,N)} \operatorname{cov}(X_{ij}, X_{rst})$.

We need to check that for every $w \in \mathbb{R}^2$,

$$(w' a_N w)^{-1/2} w'(Y - \mathrm{E}[Y]) \rightsquigarrow \mathcal{N}(0,1)$$

which is Lemma 2.1 of Biscio et al. (2018). But this reduces to checking the conditions of Theorem 4 for these now w-weighted sums.

We need to calculate growth rates for the new covariance term:

$$a_{LT} = \sum \operatorname{cov}(X_{ij}, X_{rst}) = \Theta\left(n^4 \beta_{0,T} q_L + n^4 \beta_{0,T}^2 + n^3 \beta_{0,T}\right).$$

We have for any weighted sum of a_L, a_T, and a_{LT} (w is fixed in n so does not matter) the order

$$\left(n^2 q_L + n^3 \beta_{0,T}\right) + \left(n^4 \beta_{0,T} q_L + n^4 \beta_{0,T}^2 + n^3 \beta_{0,T}\right) + \left(n^3 \beta_{0,T}\right)$$

or collecting terms and dropping the obviously dominated ones

$$n^4 \beta_{0,T} q_L + n^3 \beta_{0,T} + n^2 q_L$$

where, notice, the latter two terms were the rates of a_L and a_T and the possible new component is given by the first term.

Again Conditions (6.3) is obvious, so we check the other two.

Condition (6.1) is as follows. We examine the new terms not covered by the prior two lemmas and appeal to the sufficient condition in footnote 41. These are of the form

$$\mathrm{E}\left|X_\alpha X_\eta X_\gamma\right| = \mathrm{P}\left(X_\alpha X_\eta X_\gamma = 1\right)$$

which we now bound. The loose bounds across the two cases are

(1) one link and two triangles: This must constitute *edge adjacent* triangles. Otherwise we automatically have independence. This leaves 4 nodes, so order n^4 terms with a bound on probability $\beta_{0,T}^2$ which is notably loose.
(2) one link, one triangle, and a second link: This has a loose upper bound of $\beta_{0,T}^2 + \beta_{0,T} q_L^2$. This leaves 4 nodes again so order n^4 of these.

This exhausts the list.

So we compare

(1) $n^4 \beta_{0,T}^2 < n^6 \beta_{0,T}^{3/2} q_L^{3/2} + n^{9/2} \beta_{0,T}^{3/2} + n^3 q_L^{3/2}$

(a) $4 - 2h_T < 6 - \frac{3}{2}h_T - \frac{3}{2}h_L$ so $h_L < \frac{4}{3} + h_T$ which holds for every $h_T \geq h_L + 1$.
(b) $4 - 2h_T < 9/2 - \frac{3}{2}h_T$ so this is always true.
(c) $4 - 2h_T < 3 - \frac{3}{2}h_L$ so $\frac{1}{2} + \frac{3}{4}h_L < h_T$ which already holds.

(2) $n^4 \left(\beta_{0,T}^2 + \beta_{0,T} q_L^2\right) < n^6 \beta_{0,T}^{3/2} q_L^{3/2} + n^{9/2} \beta_{0,T}^{3/2} + n^3 q_L^{3/2}$.
(a) $4 - h_T - 2h_L < 6 - \frac{3}{2}h_T - \frac{3}{2}h_L$ so $h_T < 4 + h_L$ but this holds precisely because $h_L < 2$ and $h_T \leq 3h_L$.
(b) $4 - h_T - h_L < 9/2 - \frac{3}{2}h_T$ is mechanical.
(c) $4 - h_T - h_L < 3 - \frac{3}{2}h_L$ follows from $1 + \frac{1}{2}h_L < h_T$ which is true by assumption.

As a consequence, we have

$$\sum_{\alpha;\eta,\gamma \in \Delta(\alpha,N)} \mathrm{E}\left|X_\alpha X_\eta X_\gamma\right| = o\left(a^{3/2}\right)$$

since we have controlled the within-link and within-triangle terms in the prior two Lemmas and the cross-term above.

Next we need to verify Condition (6.2). The condition will be met with the exact same rates. To see this, first observe that we only need to consider terms of the form

$$\mathrm{cov}\left(\left(g_{ij} - q_L\right) \cdot \left(g_{ik} - q_L\right), \left(g_{rs} - q_L\right) \cdot \left(X_{rvw} - q_T\right)\right) \leq \mathrm{E}\left[g_{ij} g_{ik} g_{rs} g_{rv} g_{rw}\right]$$

and

$$\mathrm{cov}\left(\left(g_{ij} - q_L\right) \cdot \left(X_{ikl} - q_T\right), \left(g_{rs} - q_L\right) \cdot \left(X_{rvw} - q_T\right)\right) \leq \mathrm{E}\left[g_{ij} g_{ik} g_{kl} g_{rs} g_{rv} g_{rw}\right].$$

That is, the reference nodes in this condition must be pairs because triples have dependency neighborhoods that are singletons. Also observe that the neighbors considered must have at least one triangle because the all-links case has already been covered in Lemma C.1.

We have two cases to consider: a 7-node case with $\{i,j,k\} \cap \{r,s,v,w\} \neq \emptyset$ and an 8-node case with $\{i,j,k,l\} \cap \{r,s,v,w\} \neq \emptyset$.

Let us begin with the 7-node case. Here we can have one or two intersections (we have already calculated cases with 4 or fewer nodes, meaning three or more intersections). Begin with a single node in common. One can check that amongst all configurations (which intersects a two-star ijk with a triangle with a leaf (ruv and rs)), an upper bound on the probability of the structure forming is of order $\beta_{0,T}^2 q_L$. There are order n^6 such potential structures. We need to compare this to the square of the weighted sum of variance-covariance terms, which by the above is

$$\left(n^4 \beta_{0,T} q_L + n^3 \beta_{0,T} + n^2 q_L\right)^2.$$

Observe that the first term is the only one to consider—the other two have been studied,

$$6 - 2h_T - h_L < 4 - 2h_L \iff 2 + h_L < 2h_T$$

follows directly from $h_T \geq h_L + 1$ so the result follows in this case.

Next we can look at two nodes in common. This involves a number of configurations of one or two triangles and a collection of leafs and/or stars. Here we have order n^5 free nodes and we can check a loose upper bound on the probability of formation is of order $\beta_{0,T} q_L^2$. As such we see

$$5 - h_T - 2h_L < 4 - 2h_L$$

which is implied by $h_T > 1$, which is satisfied in our setting since $h_T \geq h_L + 1 > 1 + 2/3$.

This covers all cases that have not previously been calculated. So then we turn to the 8-node case. We can have one or two nodes in common before we repeat calculations already covered. If we have one node in common, the loose upper bound is probability of order $\beta_{0,T}^2 q_L^2$ and there are of course n^7 such potential collections of nodes. The condition to check is then

$$7 - 2h_T - 2h_L < 4 - 2h_L \iff 3 < 2h_T.$$

But this is mechanical since $h_T \geq 1 + 2/3$ in any case.

The final case to consider is with two nodes in common in this case of 8 nodes with two in common. In this case, a loose upper bound is $\beta_{0,T}^2 q_L$ and there are n^6. But we have studied this above and the restriction is already satisfied.

Therefore, with Λ and $\Delta(\alpha, N)$ for $\alpha \in \Lambda$ defined as above, we have shown that Conditions (6.1) and (6.2) are satisfied. Since the remaining conditions were already discussed, the sum is asymptotically normally distributed. But, since the weights were arbitrary (indeed fixed in n and therefore did not contribute to any of the dependency calculations) the result held irrespective of $(w_1, w_2) \in \mathbb{R}^2$, so joint normality follows. ■

C.3.3. *Asymptotic Normality of SUGM Estimators.* It is useful to define the variance-covariance matrix of the moments and a rate matrix

$$V_n = \begin{pmatrix} \mathrm{var}(n^{h_L} S_L) & \mathrm{cov}(n^{h_L} S_L, n^{h_T} S_T) \\ \mathrm{cov}(n^{h_L} S_L, n^{h_T} S_T) & \mathrm{var}(n^{h_T} S_T) \end{pmatrix} \quad \text{and } R_n = \begin{pmatrix} n^{h_L} & 0 \\ 0 & n^{h_T} \end{pmatrix}.$$

PROPOSITION **C.3.** *Consider a links and triangles SUGM with associated parameters* $\beta_{0,L}^n, \beta_{0,T}^n = \left(\frac{b_{0,L}}{n^{h_L}}, \frac{b_{0,T}}{n^{h_T}}\right)$ *with* $0 \leq \underline{D} < b_{0,L}, b_{0,T} < \overline{D}$ *such that*

$$h_L \in (2/3, 2) \textit{ and } h_T \in [h_L + 1, 3h_L] \textit{ with } h_T < 3.$$

Consider the minimum distance estimator $\widehat{\beta}$ *using moments* $S = (S_L(g), S_T(g))$*. Then*

$$\delta\left(\widehat{\beta}^n, \beta_0^n\right) \xrightarrow{\mathrm{P}} 0$$

and[53]

$$V_n^{-1/2} R_n \left(\widehat{\beta}^n - \beta_0^n\right) \rightsquigarrow \mathcal{N}(0, I).$$

Proof of Proposition C.3.

The proof of the result follows the outline of standard results on asymptotic normality of parameter estimates (e.g., Newey and McFadden (1994)).

It is convenient to normalize things via a change of variables via the diagonal normalizing matrix $R_n = \mathrm{diag}\{n^{h_L}, n^{h_T}\}$ to a parameter vector $b := R_n\beta$, so that the magnitude of the parameter vector does not change with n. Observe that $\delta\left(\widehat{\beta}, \beta_0^n\right) \xrightarrow{\mathrm{P}} 0$ if and only if $\widehat{b} \xrightarrow{\mathrm{P}} b_0$, and consistency in the δ-metric holds by Proposition C.2.

It is also useful to then define the expected and empirical moment functions in terms of this rescaled parameter

$$\widehat{M}_L(b) = \left[\frac{n^{h_L}}{\binom{n}{2}} \sum_{i<j} g_{ij} - \overline{q}_L(b_L, b_T)\right]$$

and

$$\widehat{M}_T(b) = \left[\frac{n^{h_T}}{\binom{n}{3}} \sum_{i<j<k} g_{ij}g_{jk}g_{ik} - \overline{q}_T(b_L, b_T)\right]$$

where $\overline{q}_L(b) = n^{h_L} q_L$ and $\overline{q}_T(b) = n^{h_T} q_T$ are the normalized expectations given parameter $(b_L, b_T) = R_n\beta$.

[53]The expression for V_n is different when $h_T = h_L + 1$, and is given in the proof of the proposition.

Let $\Delta = h_T - h_L$. We treat two separate cases, $\Delta > 1$ and $\Delta = 1$. The second case allows links to generate triangles at a similar rate as triangles, and so is a more complex case to treat, and so each step of the argument involves different arguments for the two cases.

From the first order condition of GMM estimation, we take a mean value expansion around the true normalized parameter b_0 by applying the mean-value theorem, and then solve for $\widehat{b} - b_0$.[54] Note that the mean value $\bar{b}$ is evaluated component by component in the matrix $\nabla \widehat{M}\left(\bar{b}\right)$. This abuse of notation is standard (e.g., Newey and McFadden (1994)).

$$R_n\left(\widehat{\beta} - \beta_0\right) = \left(\widehat{b} - b_0\right) = -\left[\nabla \widehat{M}\left(\widehat{b}\right)' \nabla \widehat{M}\left(\bar{b}\right)\right]^{-1} \nabla \widehat{M}\left(\widehat{b}\right)' \widehat{M}\left(b_0\right).$$

Below we will show that for $\Delta > 1$

$$-\left[\nabla \widehat{M}\left(\widehat{b}\right)' \nabla \widehat{M}\left(\bar{b}\right)\right]^{-1} \nabla \widehat{M}\left(\widehat{b}\right)' \xrightarrow{\mathrm{P}} I$$

and by Lemma C.3, for

$$V_n = \begin{pmatrix} \operatorname{var}(n^{h_L} S_L) & \operatorname{cov}(n^{h_L} S_L, n^{h_T} S_T) \\ \operatorname{cov}(n^{h_L} S_L, n^{h_T} S_T) & \operatorname{var}(n^{h_T} S_T) \end{pmatrix}$$

it follows that

$$V_n^{-1/2} \widehat{M}\left(b_0\right) \rightsquigarrow \mathcal{N}\left(0, I\right).$$

Therefore by Slutzky's theorem, it follows that

$$V_n^{-1/2} R_n\left(\widehat{\beta} - \beta_0\right) \rightsquigarrow \mathcal{N}\left(0, I\right).$$

Thus, to complete the proof for the case of $\Delta > 1$, it suffices to show that

$$-\left[\nabla \widehat{M}\left(\widehat{b}\right)' \nabla \widehat{M}\left(\bar{b}\right)\right]^{-1} \nabla \widehat{M}\left(\widehat{b}\right)' \xrightarrow{\mathrm{P}} I$$

For the case of $\Delta = 1$ we will end up with a different expression for the limit of $-\left[\nabla \widehat{M}\left(\widehat{b}\right)' \nabla \widehat{M}\left(\bar{b}\right)\right]^{-1} \nabla \widehat{M}\left(\widehat{b}\right)'$ and so will have a different covariance and normalization.

To find the limit of these gradient terms, we need to compute $\nabla \bar{q}$, where we define

$$\begin{aligned}
\bar{q}_L\left(b\right) &:= n^{h_L}\left[q_L\left(\beta\right)\right] \\
&= n^{h_L}\left[\beta_L + \left(1 - \beta_L\right)\left[1 - \left(1 - \beta_T\right)^{n-2}\right]\right] \\
&= b_L + \left(n^{h_L} - b_L\right)\left(n - 2\right) \cdot \frac{b_T}{n^{h_T}} \\
&= b_L + \frac{b_T}{n^{h_T - h_L - 1}} + o\left(1\right).
\end{aligned}$$

Similarly

$$\begin{aligned}
\bar{q}_T\left(b\right) &:= n^{h_T}\left[q_T\left(\beta\right)\right] \\
&= n^{h_T}\left[\beta_T + \left(1 - \beta_T\right)\left\{\beta_L + \left(1 - \beta_L\right)\left[1 - \left(1 - \beta_T\right)^{n-2}\right]\right\}^3\right] \\
&= b_T + \left(n^{h_T} - b_T\right)\left\{\frac{b_L}{n^{h_L}} + \frac{b_T}{n^{h_T - 1}} - \frac{b_T}{n^{h_T + h_L - 1}} b_L\right\}^3
\end{aligned}$$

[54]This is valid because b_0 is assumed to lie in the interior of B, a compact set, which then implies the sequence of $\mathcal{B}^n$ under consideration.

$$\approx b_T + \left\{ \underbrace{\frac{b_L}{n^{h_L - h_T/3}} + \frac{b_T}{n^{2h_T/3-1}} - \frac{b_T}{n^{h_T+h_L-1-h_T/3}} b_L}_{=:x} \right\}^3$$

where we dropped the final term that is of lesser order.

Note that the third term will always be of lesser order, so

$$x = O\left(\left\{ \frac{b_L}{n^{h_L - h_T/3}} + \frac{b_T}{n^{2h_T/3-1}} \right\} \right).$$

Also notice

$$h_L - h_T/3 < 2h_T/3 - 1 \iff \Delta > 1.$$

Thus, if $\Delta > 1$ only the first term in x matters, while if $\Delta = 1$ then the two terms are of the same order.

Finally it will be useful to write

$$\nabla \bar{q}_L = \begin{pmatrix} 1 + o(1) \\ n^{1-\Delta} + o\left(n^{1-\Delta}\right) \end{pmatrix}$$

and

$$\nabla \bar{q}_T = \begin{pmatrix} 3x^2 \left(\frac{1}{n^{h_L - h_T/3}} (1 + o(1)) \right) \\ 1 + 3x^2 \left(\frac{1}{n^{2h_T/3-1}} (1 + o(1)) \right) \end{pmatrix}.$$

Consider the case where $\Delta > 1$. Then

$$\nabla \bar{q}_L = \begin{pmatrix} 1 + o(1) \\ n^{1-\Delta} + o\left(n^{1-\Delta}\right) \end{pmatrix} = \begin{pmatrix} 1 + o(1) \\ o(1) \end{pmatrix}$$

and

$$\nabla \bar{q}_T = \begin{pmatrix} o(1) \\ 1 + o(1) \end{pmatrix},$$

since $3h_L > h_T$ and $h_T > \frac{3}{2}$.

Now consider the case where $\Delta = 1$. Again

$$\nabla \bar{q}_T = \begin{pmatrix} o(1) \\ 1 + o(1) \end{pmatrix}$$

but in this case

$$\nabla \bar{q}_L = \begin{pmatrix} 1 + o(1) \\ n^{1-\Delta} + o\left(n^{1-\Delta}\right) \end{pmatrix} = \begin{pmatrix} 1 + o(1) \\ 1 + o(1) \end{pmatrix}.$$

Notice that $\bar{q}(b)$ is a continuously differentiable function of $b \in B$, where B is compact, and $\nabla_b \bar{q}(b)$ has a bounded derivative. This allows us to write

$$\nabla \widehat{M}(\hat{b}) = \nabla \widehat{M}(b_0) + o_p(1) = -\nabla \bar{q}(b_0) + o_p(1).$$

We explicitly compute the inverse of $\nabla \bar{q}(b)' \nabla \bar{q}(b)$ below, which exists.

If $\Delta > 1$ we can write

$$-\nabla \widehat{M}(b) = \begin{pmatrix} \frac{\partial \bar{q}_L}{\partial b_L} & \frac{\partial \bar{q}_L}{\partial b_T} \\ \frac{\partial \bar{q}_T}{\partial b_L} & \frac{\partial \bar{q}_T}{\partial b_T} \end{pmatrix} = \begin{pmatrix} 1 + o(1) & o(1) \\ o(1) & 1 + o(1) \end{pmatrix}$$

and if $\Delta = 1$ we can write

$$-\nabla \widehat{M}(b) = \begin{pmatrix} \frac{\partial \overline{q}_L}{\partial b_L} & \frac{\partial \overline{q}_L}{\partial b_T} \\ \frac{\partial \overline{q}_T}{\partial b_L} & \frac{\partial \overline{q}_T}{\partial b_T} \end{pmatrix} = \begin{pmatrix} 1 + o(1) & 1 + o(1) \\ o(1) & 1 + o(1) \end{pmatrix}.$$

We can also compute

$$\begin{aligned} \nabla \widehat{M}(b)' \nabla \widehat{M}(b) &= \begin{pmatrix} \frac{\partial \overline{q}_L}{\partial b_L} & \frac{\partial \overline{q}_T}{\partial b_L} \\ \frac{\partial \overline{q}_L}{\partial b_T} & \frac{\partial \overline{q}_T}{\partial b_T} \end{pmatrix} \begin{pmatrix} \frac{\partial \overline{q}_L}{\partial b_L} & \frac{\partial \overline{q}_L}{\partial b_T} \\ \frac{\partial \overline{q}_T}{\partial b_L} & \frac{\partial \overline{q}_T}{\partial b_T} \end{pmatrix} \\ &= \begin{pmatrix} \left(\frac{\partial \overline{q}_L}{\partial b_L}\right)^2 + \left(\frac{\partial \overline{q}_T}{\partial b_L}\right)^2 & \frac{\partial \overline{q}_L}{\partial b_L}\frac{\partial \overline{q}_L}{\partial b_T} + \frac{\partial \overline{q}_L}{\partial b_T}\frac{\partial \overline{q}_T}{\partial b_T} \\ \frac{\partial \overline{q}_L}{\partial b_L}\frac{\partial \overline{q}_L}{\partial b_T} + \frac{\partial \overline{q}_T}{\partial b_L}\frac{\partial \overline{q}_T}{\partial b_T} & \left(\frac{\partial \overline{q}_L}{\partial b_T}\right)^2 + \left(\frac{\partial \overline{q}_T}{\partial b_T}\right)^2 \end{pmatrix} \end{aligned}$$

and so the inverse is

$$\left[\nabla \widehat{M}(b)' \nabla \widehat{M}(b)\right]^{-1} = \frac{1}{\det\left[\nabla \widehat{M}(b)' \nabla \widehat{M}(b)\right]} \begin{pmatrix} \left(\frac{\partial \overline{q}_L}{\partial b_T}\right)^2 + \left(\frac{\partial \overline{q}_T}{\partial b_T}\right)^2 & -\left[\frac{\partial \overline{q}_L}{\partial b_L}\frac{\partial \overline{q}_L}{\partial b_T} + \frac{\partial \overline{q}_L}{\partial b_T}\frac{\partial \overline{q}_T}{\partial b_T}\right] \\ -\left[\frac{\partial \overline{q}_L}{\partial b_L}\frac{\partial \overline{q}_L}{\partial b_T} + \frac{\partial \overline{q}_T}{\partial b_L}\frac{\partial \overline{q}_T}{\partial b_T}\right] & \left(\frac{\partial \overline{q}_L}{\partial b_L}\right)^2 + \left(\frac{\partial \overline{q}_T}{\partial b_L}\right)^2 \end{pmatrix}.$$

The determinant is given by

$$\begin{aligned} \det\left[\nabla \widehat{M}(b)' \nabla \widehat{M}(b)\right] &= \left[\left(\frac{\partial \overline{q}_L}{\partial b_T}\right)^2 + \left(\frac{\partial \overline{q}_T}{\partial b_T}\right)^2\right]\left[\left(\frac{\partial \overline{q}_L}{\partial b_L}\right)^2 + \left(\frac{\partial \overline{q}_T}{\partial b_L}\right)^2\right] \\ &\quad - \left[\frac{\partial \overline{q}_L}{\partial b_L}\frac{\partial \overline{q}_L}{\partial b_T} + \frac{\partial \overline{q}_T}{\partial b_L}\frac{\partial \overline{q}_T}{\partial b_T}\right]^2 \\ &= \left(\frac{\partial \overline{q}_L}{\partial b_L}\frac{\partial \overline{q}_T}{\partial b_T} - \frac{\partial \overline{q}_T}{\partial b_L}\frac{\partial \overline{q}_L}{\partial b_T}\right)^2. \end{aligned}$$

If $\Delta > 1$ then the determinant is $1 + o(1)$. If $\Delta = 1$ it is the same.

So the inverse is, if $\Delta > 1$,

$$\left[\nabla \widehat{M}(b)' \nabla \widehat{M}(b)\right]^{-1} = \frac{1}{1 + o(1)} \begin{pmatrix} 1 + o(1) & o(1) \\ o(1) & 1 + o(1) \end{pmatrix}.$$

We can compute the final object in the case $\Delta > 1$ as

$$\begin{aligned} -\left[\nabla \widehat{M}(b_0)' \nabla \widehat{M}(b_0)\right]^{-1} \nabla \widehat{M}(b_0)' &= \frac{1}{1 + o(1)} \begin{pmatrix} 1 + o(1) & o(1) \\ o(1) & 1 + o(1) \end{pmatrix} \begin{pmatrix} 1 + o(1) & o(1) \\ o(1) & 1 + o(1) \end{pmatrix} \\ &= \begin{pmatrix} 1 + o(1) & o(1) \\ o(1) & 1 + o(1) \end{pmatrix} \xrightarrow{\mathrm{P}} I, \end{aligned}$$

which completes the argument for the case of $\Delta > 1$.

Meanwhile if $\Delta = 1$ then the inverse is

$$\left[\nabla \widehat{M}(b)' \nabla \widehat{M}(b)\right]^{-1} = \frac{1}{1 + o(1)} \begin{pmatrix} 2 + o(1) & -1 + o(1) \\ -1 + o(1) & 1 + o(1) \end{pmatrix}.$$

Therefore,

$$\begin{aligned} -\left[\nabla \widehat{M}(b_0)' \nabla \widehat{M}(b_0)\right]^{-1} \nabla \widehat{M}(b_0)' &= \frac{1}{1 + o(1)} \begin{pmatrix} 2 + o(1) & -1 + o(1) \\ -1 + o(1) & 1 + o(1) \end{pmatrix} \begin{pmatrix} 1 + o(1) & o(1) \\ 1 + o(1) & 1 + o(1) \end{pmatrix} \\ &= \begin{pmatrix} 1 + o(1) & -1 + o(1) \\ o(1) & 1 + o(1) \end{pmatrix} \xrightarrow{\mathrm{P}} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Now consider the case with $\Delta = 1$. In this case since $n^{2-h_L} = n^{3-h_T}$, it follows from our calculations above that

$$\sqrt{n^{2-h_L}} R_n \left(\widehat{\beta} - \beta_0\right) = \sqrt{n^{2-h_L}} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{n^{h_L}}{\binom{n}{2}} \sum g_{ij} - \overline{q}_L\left(b_0\right) \\ \frac{n^{h_T}}{\binom{n}{3}} \sum g_{ij} g_{ik} g_{jk} - \overline{q}_T\left(b_0\right) \end{pmatrix}$$
$$= \sqrt{n^{2-h_L}} \begin{pmatrix} \left\{\frac{n^{h_L}}{\binom{n}{2}} \sum g_{ij} - \overline{q}_L\left(b_0\right)\right\} - \left\{\frac{n^{h_T}}{\binom{n}{3}} \sum g_{ij} g_{ik} g_{jk} - \overline{q}_T\left(b_0\right)\right\} \\ \frac{n^{h_T}}{\binom{n}{3}} \sum g_{ij} g_{ik} g_{jk} - \overline{q}_T\left(b_0\right). \end{pmatrix},$$

which still jointly converge to a mean zero random variable, but with a different variance-covariance matrix:

$$V_n = \begin{pmatrix} \mathrm{var}(n^{h_L} S_L - n^{h_T} S_T) & \mathrm{cov}(n^{h_L} S_L - n^{h_T} S_T, n^{h_T} S_T) \\ \mathrm{cov}(n^{h_L} S_L - n^{h_T} S_T, n^{h_T} S_T) & \mathrm{var}(n^{h_T} S_T) \end{pmatrix}$$

for the $\Delta = 1$ case. ■

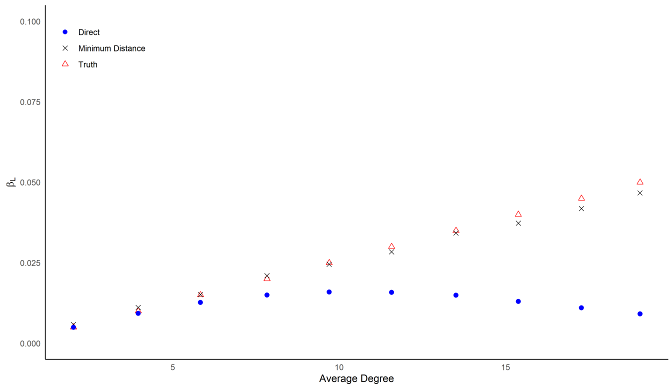
Appendix D. Simulations

We demonstrate the consistency of $\hat{\beta}$, asymptotically normal distribution of $\hat{\beta}$, and quality of estimated standard errors. We show that the minimum distance estimator performs well in the links and triangles model—estimates are centered on the true value of the parameters, the distribution indeed looks asymptotically normal, and the constructed 95% confidence interval has proper coverage.

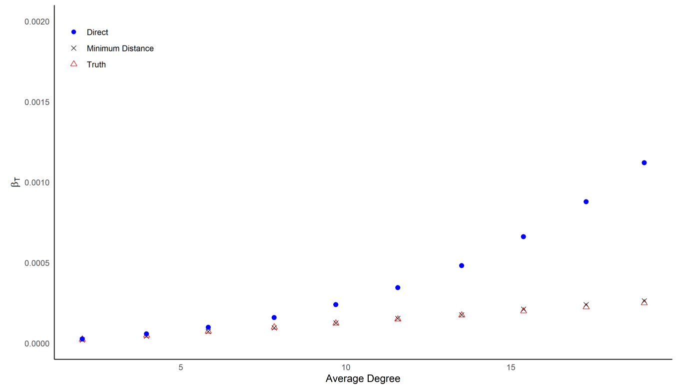
We set $h_L = 1$, $h_T = h_L + 1$, and $b_{0,L} = b_{0,T}$, which we vary in order to generate the requisite average degrees described below.

D.1. **Consistency.** First we present the point estimates of our minimum distance and direct estimators over a wide range of average degrees. Consistent with the theory, the estimators coincide at low densities and as we look at increasingly dense graphs, the direct estimator misattributes direct link formation to triangles and inherits a bias, whereas the minimum distance estimator does not. Figure D.1 presents the results.

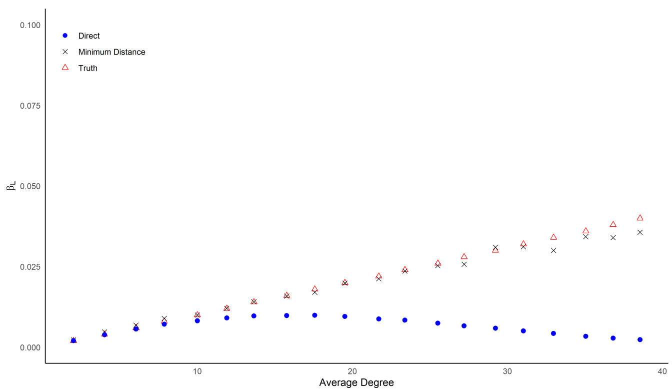
D.2. **Asymptotic Normality.** Next, we turn to the asymptotic distribution of the estimator and show that it is normal. We display the (standardized) parameter estimates $(\hat{\beta}_L, \hat{\beta}_T)$ and a standard normal distribution for lower and higher degrees in the range consistent with data and across a broad range of network sizes. The results, displayed in Figures D.2 and D.3 clearly show that the normal approximation is good.



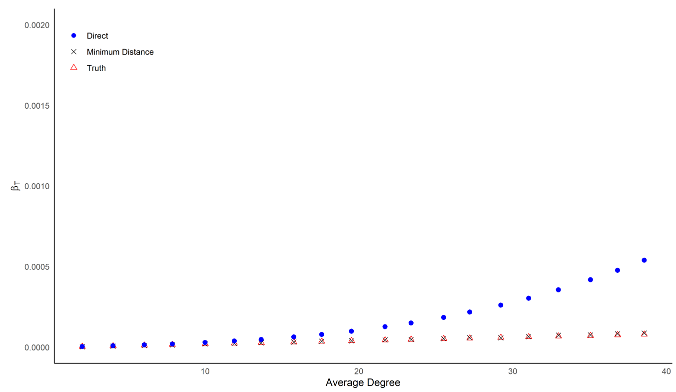
(A) $n = 200,\ \beta_L$



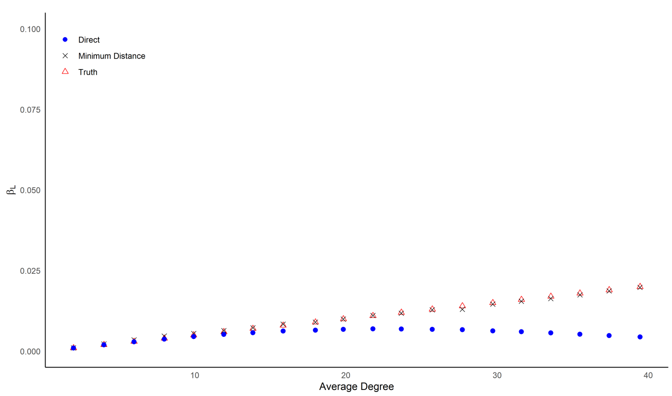
(B) $n = 200,\ \beta_T$



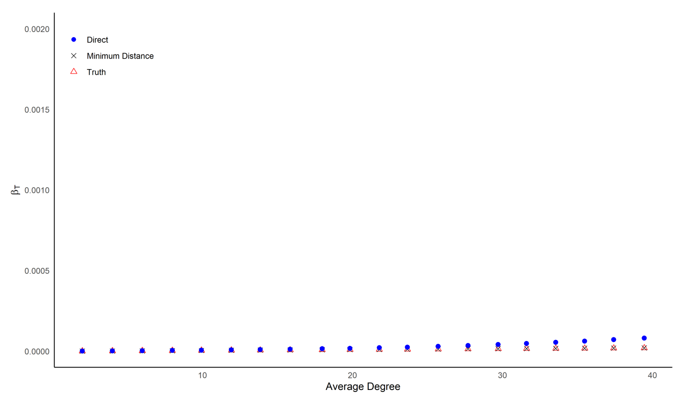
(C) $n = 500,\ \beta_L$



(D) $n = 500,\ \beta_T$

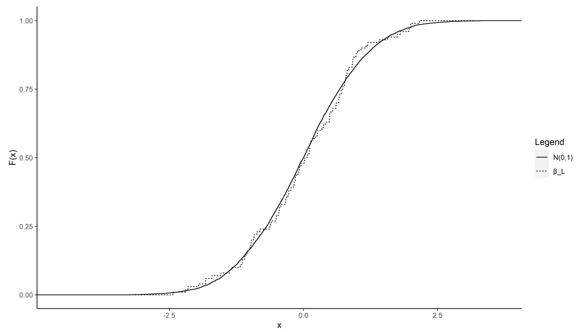


(E) $n = 1000,\ \beta_L$

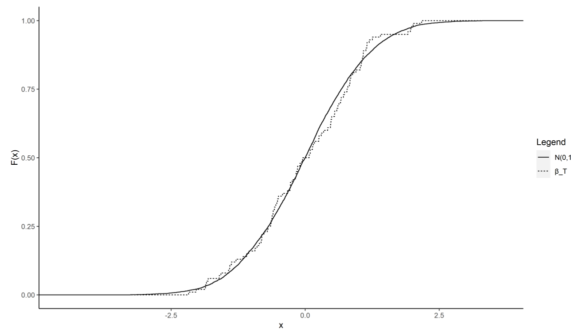


(F) $n = 1000,\ \beta_T$

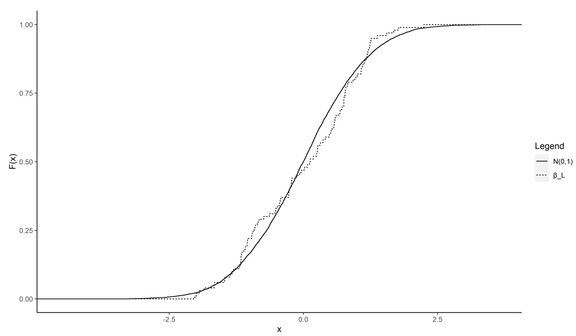
FIGURE D.1. Estimates from the mimimum distance and direct estimators for $n \in \{200, 500, 1000\}$ with 100 simulations used.



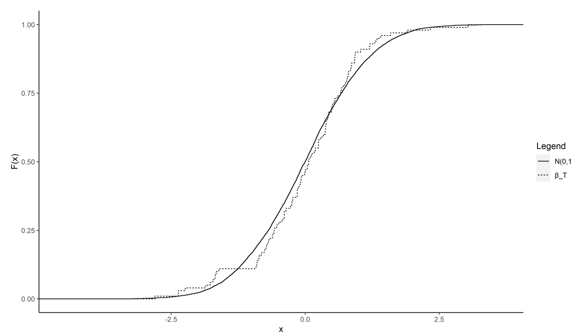
(A) $n = 200,\ \beta_L$



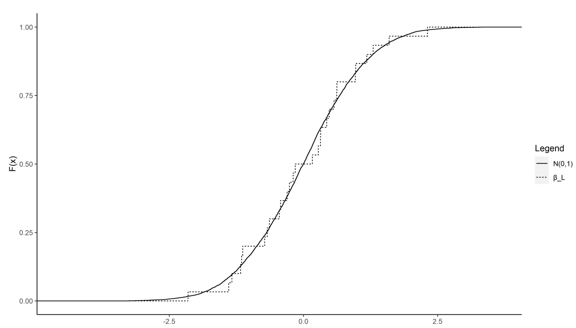
(B) $n = 200,\ \beta_T$



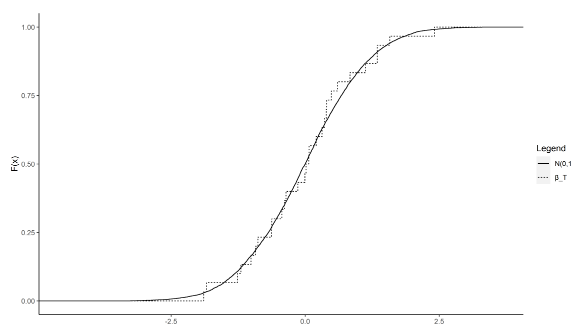
(C) $n = 500,\ \beta_L$



(D) $n = 500,\ \beta_T$

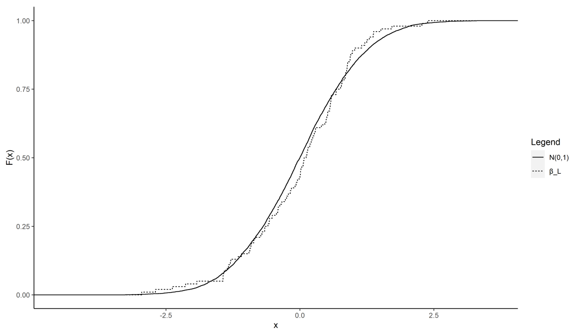


(E) $n = 1000,\ \beta_L$

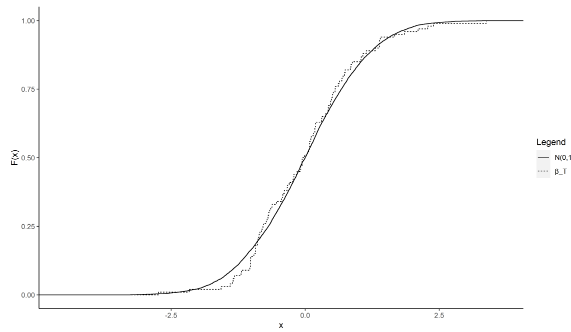


(F) $n = 1000,\ \beta_T$

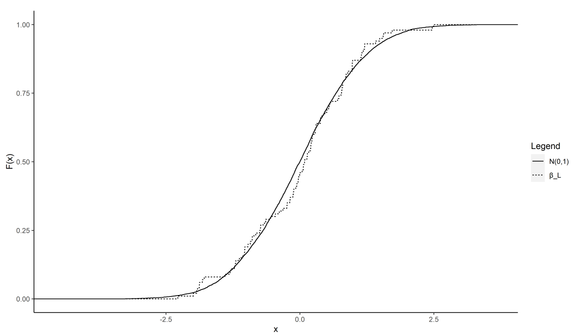
FIGURE D.2. CDFs of standardized parameter estimates $\widehat{\beta}_L, \widehat{\beta}_T$ for average degrees of 5.9 with $n \in \{200, 500, 1000\}$ with 100 simulations.



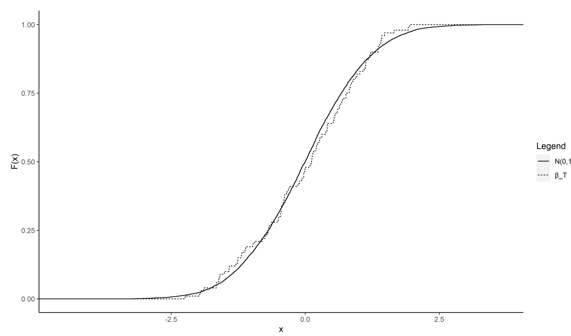
(A) $n = 200,\ \beta_L$



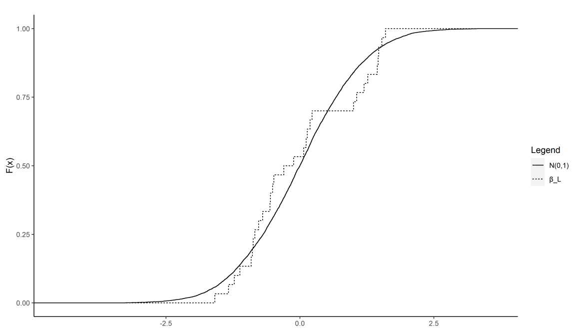
(B) $n = 200,\ \beta_T$



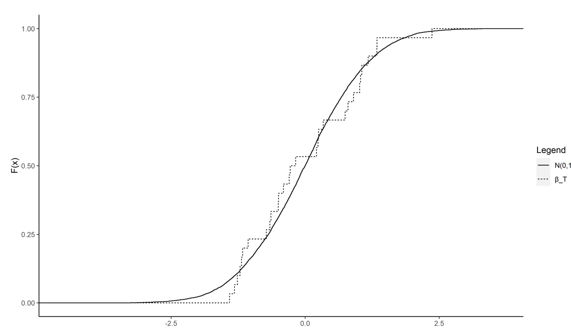
(C) $n = 500,\ \beta_L$



(D) $n = 500,\ \beta_T$



(E) $n = 1000,\ \beta_L$



(F) $n = 1000,\ \beta_T$

FIGURE D.3. CDFs of standardized parameter estimates $\widehat{\beta}_L, \widehat{\beta}_T$ for average degrees of 13.5 with $n \in \{200, 500, 1000\}$ with 100 simulations.

D.3. **Coverage.** Having seen that both the parameter estimates are accurate and appear to be asymptotically normally distributed, consistent with theory, finally we investigate the analytic, estimated standard errors. We construct these using the minimum distance estimator of Proposition 3. We omit for brevity but note that by simply simulating the model from the estimated parameters and then calculating standard errors through this process has excellent coverage properties, but developing the theory of a bootstrap is beyond the scope of the present paper and remains a topic for future work.

TABLE D.1. Coverage for β_L

Avg. Degree	100	150	200	250	300	350	400
7.80	0.960	0.950	0.960	0.960	0.935	0.925	0.940
9.78	0.965	0.985	0.975	0.975	0.975	0.960	0.870
11.66	0.935	0.950	0.985	0.96	0.955	0.970	0.94
13.54	0.930	0.930	0.955	0.965	0.97	0.985	0.965
15.44	0.890	0.935	0.940	0.920	0.955	0.970	0.965

Notes: Coverage of the 95% confidence interval for the link probability parameter implied by Proposition 3, for various average degrees for various network sizes. For each network size we and each degree we conduct 200 simulations.

TABLE D.2. Coverage for β_T

Avg. Degree	100	150	200	250	300	350	400
7.80	0.960	0.950	0.960	0.960	0.895	0.925	0.940
9.78	0.965	0.985	0.975	0.975	0.975	0.960	0.870
11.66	0.935	0.950	0.985	0.990	0.955	0.970	0.985
13.54	0.930	0.930	0.955	0.965	0.995	0.985	0.965
15.44	0.890	0.935	0.940	0.920	0.955	0.970	0.965

Notes: Coverage of the 95% confidence interval for the triangle probability parameter implied by Proposition 3, for various average degrees for various network sizes. For each network size we and each degree we conduct 200 simulations.

To take stock at an aggregate level, in this range the overall average coverage is 0.952 and 0.953 for links and triangles respectively.