ECONOMETRICS OF NETWORK FORMATION

ARUN G. CHANDRASEKHAR[‡]

1. Introduction

A growing empirical literature examines the role that networks play in a variety of economic phenomena. This research can be crudely partitioned into two branches. The first holds networks fixed and studies economic processes operating on networks (e.g., social learning, labor market search, peer effects in education). This is the subject of several other chapters in this handbook (e.g., "Some challenges in the empirics of the effects of networks," Boucher and Fortin; "Field experiments, social networks, and development," Breza; "Networked experiments: a review of methods and innovations," Aral). The second branch considers how and why economic networks form. To date, most work in this space has been theoretical, with very limited observational or experimental work on network formation. One major impediment to the progress of this research is the dearth of practicable empirical models of network formation.

This chapter will focus on the econometric issues that arise when an empirical researcher seeks to model network formation. A key goal of an empirical analysis of network formation is to estimate how and why relationships form. What governs the overall structure of the economy? Are there structural regularities that tell us something about the economic incentives underlying the interaction? The researcher may want to estimate parameters driving network formation in order to (i) test economic hypotheses and (ii) conduct counterfactual analysis.[1]

A major difficulty is that the researcher typically has a dataset consisting of a single network observed in a single period (e.g., social relationships among students in a single university, financial relationships between households in a single village). One reason that this is often the case is due to the prohibitive cost in collecting high-quality network data (see, e.g., Banerjee, Chandrasekhar, Duflo, and Jackson (2013)). Thus, the relevant thought experiment is the following: given a single network of n agents, are the parameters driving the joint distribution of agents' linking decisions, their characteristics and possibly the economic environment (including observables, unobservables, etc.) consistently estimable as $n \to \infty$? This estimation problem involves a number of challenges, described below, including but not limited to the

Date: June 22, 2015.

I thank Emily Breza, Gabriel Carroll, Aureo de Paula, Bryan Graham, Han Hong, Itay Fainmesser, Paul Goldsmith-Pinkham, Matthew Jackson, Cynthia Kinnan, Michael Leung, Elena Manresa, Mounu Prem, Xiao Yu Wang and Juan Pablo Xandri for helpful discussions.

[‡]Department of Economics, Stanford University; NBER.

[1]Another practical reason to develop network formation models is to be able to deal with missing data since data collection can be prohibitive.

degree of correlation between linking decisions, concerns about multiple equilibria, missing data. The intersection of such concerns arise in the network setting.

To highlight the core difficulty, that one large network may encode a tremendous amount of information or a very limited amount of information, observe the following. Suppose the researcher knows every two nodes are linked with probability p, where p is the parameter to be estimated. We can imagine two extremes. At one extreme, any two pairs of nodes are independent of each other. Then a single network effectively provides $\binom{n}{2}$ independent observations, so consistent estimation of p is trivial. At the other extreme, all pairs might be perfectly correlated, i.e., with probability p all nodes are linked, and with probability $1-p$ no two nodes are linked. Then it is clear that having a single network, no matter how large, is not enough to estimate p consistently.

Consider an economy with a set V of agents, whom we refer to as *nodes*. Assume there are $n = |V|$ nodes and each needs to decide whether or not to develop a relationship (e.g., a transaction, a friendship) with others. The resulting network of links, $g = (V, E)$, is called a *graph*, which consists of a set of nodes V and a set of edges E. If nodes share an edge, $ij \in E$. It is useful to represent this with an *adjacency matrix* $\mathbf{A} := \mathbf{A}(g)$, with

$$A_{ij} = \begin{cases} 1 & \text{if} \quad ij \in E \\ 0 & \quad\quad ij \notin E. \end{cases}$$

For simplicity, assume that if a link from i to j exists, then the link from j to i exists (networks are undirected), and a link either exists or it doesn't (the network is unweighted).[2]

We are interested in modeling the formation of g, in an economy where each agent may have some attributes and incentives to form relationships with other nodes.

Typically, the researcher has observations from a single network with many nodes (e.g., a university, a village). The researcher may also have a vector of covariates, $x = (x_1, ..., x_n)$. The goal is to use the data $(\mathbf{A}, x)$ to estimate the parameters of network formation from the observation of a single, large network. The challenge is that with a single network, there is potentially a lot of interdependency between the links. In sum, the researcher has a draw from

$$\mathrm{P}^n_\beta(\mathbf{A}|x)$$

which assigns likelihood to each of the possible $\binom{n}{2}$ graphs.

How much information is contained in such data for the econometrician? The crucial question is the extent to which the an event, e.g., i is linked to j but k is not linked to l, influences the likelihood of some other u and v being linked. It is useful to think of a vector of outcomes, the linking decisions for all $\binom{n}{2}$ pairs of nodes:

$$\mathrm{vec}(\mathbf{A}) = \left(\underbrace{1}_{\mathrm{A}_{1,2}}, \underbrace{0}_{\mathrm{A}_{1,3}}, ..., \underbrace{[\ ?\ ]}_{\mathrm{A}_{\mathrm{ij}}}, ..., \underbrace{[\ ?\ ]}_{\mathrm{A}_{\mathrm{n-1,n}}} \right).$$

[2]Most of what follows applies for directed and weighted graphs as well.

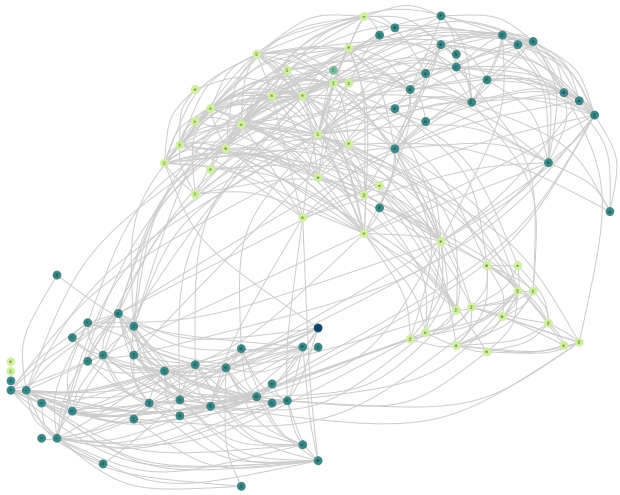


FIGURE 1. This figure, taken from Chandrasekhar, Kinnan, and Larreguy (2014) displays a village network from Karnataka, India. Nodes are households and are colored by caste. Edges denote whether members of households socialize with each other or engage in financial transactions.

Here we see the two extremes. If links are all independent, there are effectively $\binom{n}{2}$ observations in a given network.[3] However, if all links are completely correlated, the researcher actually has only one effective draw from the probability distribution, no matter the n.

In practice, the distributions we are trying to model may live somewhere in between. There is likely to be considerable interdependency among links, both for strategic reasons and due to (observable and unobservable) characteristics, but there may also be enough independence such that consistent estimation of the model's parameters is possible. This is very much like in the analysis of time-series or spatial data under standard assumptions: draws from adjacent periods or locations are highly correlated but draws that are far enough apart temporally are near-independent. A crucial distinction between the network case and the time-series or spatial case is that time-series or spatial contexts have natural embeddings into some metric space whereas in the network case there may not be such an embedding. Time has past, present, and future, and spatial data inherits, by construction, a natural geometry. For a network, there is usually no natural embedding of nodes into some space.

Our goal is to develop econometric models of network formation and have estimators that provide consistent estimates of the joint distribution, that are tractable, and that allow the researcher to conduct inference. Econometrically modeling network formation presents several challenges, some though not all unique to a networks setting. First, the econometric model must be tractable and have good asymptotic properties.

[3]Estimation would be straightforward assuming away complications caused by unobserved heterogeneity, which we discuss below.

Parameter estimates should be consistent, for instance, and estimating them should be feasible. The researcher faces a large network, which means that there are $2^{\binom{n}{2}}$ potential (undirected) networks. This can make it very difficult to conduct feasible estimation. Various techniques explored in the literature involve drawing a sample of networks over this immense, discrete space, or exploring the space of networks to determine bounds on parameter estimates. The enormity of the space can make it difficult to deal with.[4]

Second, the econometric model should reflect the structure of the network relevant to the economic problem. It is essential that the models used by the researcher reflect meaningful patterns in the data, which likely correspond to underlying economic forces. It is crucial, for instance, that network models used in applied work are consistent with stylized facts. A theoretical challenge here, as the reader will see, is that many natural network formation models are inconsistent with basic facts corresponding to empirical network data. In this chapter, we focus on the fact that economic networks (i) are sparse and (ii) exhibit correlation in the location of links.

Third, since the decisions to link can be thought of as a discrete choice, endogeneity and the presence of unobserved heterogeneity present identification issues. For instance, network models, by their very nature, put outcome variables – whether or not certain links are built – on both the left and right-hand side of equations. The event that $ij \in E$ can depend on the event that $kl \in E$ and vice versa. The structural equations may have multiple solutions, and therefore, issues surrounding multiple equilibria may need to be confronted (Bresnahan and Reiss, 1991; Tamer, 2003).

In this chapter, we will attempt to survey the literature and take stock of the approaches developed to deal with each of these issues. The models used in the literature often typically map to some preferences for agents over networks. A fairly general way to describe these models is

$$u_i(g, x, \epsilon; \beta),$$

where x is a vector of (observable and unobservable) agent attributes and ϵ is a vector of unobservable shocks. Let $g + ij$ be shorthand for the graph $(V, E \cup \{ij\})$ and similarly $g - ij$ be the shorthand for the graph $(V, E \setminus \{ij\})$. The basic decision individuals are making can be thought of as evaluating their marginal utility from maintaining or dropping a candidate link ij,

$$U_i(ij, g, x, \epsilon; \beta) := u_i(g + ij, x, \epsilon; \beta) - u_i(g - ij, x, \epsilon; \beta).$$

Therefore, the researcher is tasked with deciding how interdependent her model will be by making choices over specifications of preferences and disturbances. For instance, if

$$u_i(g, x, \epsilon; \beta) = \sum_j (\beta - \epsilon_{ij}) A_{ij}$$

then it is easy to see that the decisions to form links ij and ik are independent, if shocks were independent.

[4]Insights from empirical data can help here. While there are many possible networks and certain models are unable to be estimated in a computationally feasible model, conditioning on classes of models that replicate certain empirical features such as sparsity may reduce the problem significantly.

On the other hand, preferences may be more complex. For instance, consider

$$u_i(g, x, \epsilon; \beta) = \sum_{j \neq i} (\beta_L - \epsilon_{ij}) L_{ij} + \sum_{j<k \neq i} (\beta_T - \epsilon_{ijk}) T_{ijk}$$

where L_{ij} denotes a link ij that is not part of a triangle ($A_{ij} = 1$, $A_{ik}A_{jk} = 0$ for all k) and T_{ijk} denotes a triangle (i, j, k mutually linked). Here, the agent receives different payoffs for being in different types of relationships. Having a direct friend may be less beneficial than having a friend who is also a mutual friend of another friend. This is related to a simple model of favor exchange, in which being part of a group of three generates a payoff, whereas having a link or two friends who themselves are unlinked may generate no payoff (Jackson, Barraquer, and Tan, 2012).

Of course, there is nothing special in writing a model in which agents receive payoffs from links and from triangles. One could imagine writing a model based on utility from being in various subgraphs $g_\ell \subset g$ and payoff functions $v(g_\ell, x_\ell, \epsilon_\ell; \beta_\ell)$, which can depend on the covariates of those nodes involved in the subgraph $x_\ell = (x_{1,\ell}, ..., x_{n,\ell})$:

$$u_i(g, x, \epsilon; \beta) = \sum_{i \in g_\ell} v(g_\ell, x_\ell, \epsilon_\ell; \beta_\ell).$$

How these preferences are specified, how an equilibrium is selected, and the nature of the errors all will influence to a large degree the ability of the model to both reflect realistic features of empirical data while also delivering estimators that are feasible and enable us to conduct inference.

This chapter makes a modest contribution to a growing set of review articles on network formation models across many disciplines (e.g., Albert and Barabasi (2002); Goldenberg, Zheng, Fienberg, and Airoldi (2010); Blume, Brock, Durlauf, and Ioannides (2010); Graham (2014b)). Most recently Graham (2014b) surveys methods of identification in social networks, paying particular attention to distinguishing between structural transitivity (correlation in links due to payoff incentives) versus homophily (correlation in links due to unobserved individual characteristics). My goal here is to acquaint readers, in a self-contained manner, with a number of network formation models used in the graph theory, statistics, sociology, and econometrics literatures, with a view to both how well they map to real-world data as well as the sorts of microfoundations they implicitly assume. I hope this serves as a toolbox and allows them to survey a wide set of modeling choices to see what might be useful for their specific aims.

The remainder of the chapter is organized as follows. In Section 2, I review some common network-based datasets used in the literature and document key stylized facts of empirical networks that are useful to bear in mind when thinking about modeling decisions.

Section 3 looks at models that are conditionally edge-independent. In these models the linking decisions only depend on agents' attributes and (independent) shocks, but a link elsewhere in the graph, conditional on these attributes and shocks, has no correlation with a considered link. The main concepts dealt with here pertain to heterogeneity. Agents may be members of unobserved groups and may even have unobserved

heterogeneity that systematically influences their otherwise independent linking decisions. Such concerns, even without introducing interdependence in linking decisions, can generate challenging econometric problems.

Section 4 considers models that allow for higher order dependence between links.[5] In those models, agents' payoffs can depend on links elsewhere in the graph in a direct manner. This renders some models used in the literature entirely unusable, because estimators may not be feasible or consistent for the parameters they are aiming to recover. Models that succeed, of course, either limit the degree to which preferences across agents can be entangled by maintaining that payoffs can come from subgraphs that are not too large, or generate exogenous reasons why agents that are in some sense far enough away can never link to each other and therefore generate a form of independence that way.

Section 5 concludes and provides a discussion of the tradeoffs explored throughout the chapter.

2. Stylized Facts

Let us first turn our attention to some datasets in order to establish stylized facts. I will focus on two themes:[6] (1) the density of links, and (2) the correlation in link presence.

Table 1. Properties of various network datasets

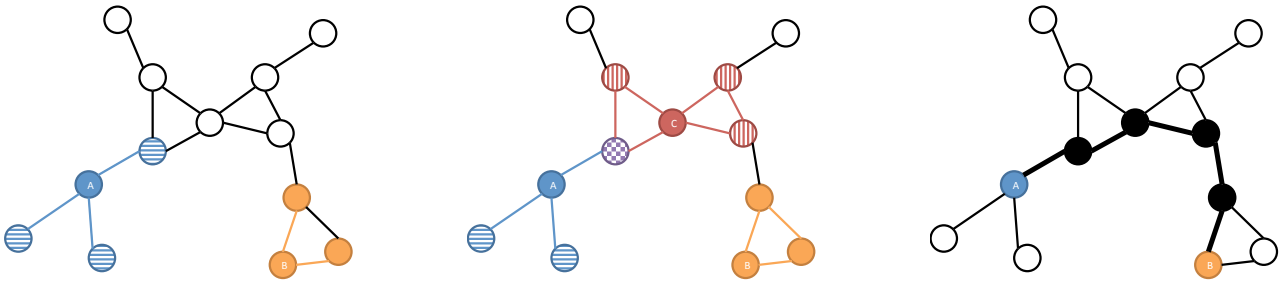
Data Source	Average Degree	Density	Clustering	Average Path Length	Number of Networks	Average Number of Nodes
AddHealth	3.361	0.042	0.411	--	14	80.36
China Villages	3.266	0.113	--	2.578	185	28.82
Harvard Dorm	10.830	0.019	0.181	3.260	1	569
Harvard Facebook-Based	7.917	0.003	0.174	4.571	1	2360
Karnataka Villages	17.378	0.089	0.303	2.337	75	213.37
Malawi Villages	6.189	0.061	0.234	3.089	21	133.48
Uganda Village	8.300	0.069	0.230	2.500	1	119

Development economics and labor economics are two subfields that have made strong contributions to empirical network economics. Therefore, I focus on datasets that make use of social network data from villages and schools. The data includes networks from the AddHealth dataset (the sample used in Badev (2013)), Harvard social network data (used in Leider et al. (2009); Mobius et al. (2010)), and several rural network datasets from the developing world (Cai et al. (2012); Banerjee et al. (2014); Ngatia (2011); De Weerdt and Dercon (2006)). A link in the village data typically represents

[5]Surely models of unobserved heterogeneity and models of higher order dependence can be merged.

[6]The literature has focused on a number of other stylized facts about empirical networks. For example, networks have short paths and exhibit both positive correlation in degree and negative correlation between a node's degree and clustering. See Jackson (2008) or Jackson and Rogers (2007a) for further discussion.

whether a member of one household has a social, financial, informational, or familial relationship with a member of another household. A link in the US university/school data represents a friendship link. The results are presented in Table 1.



(A) Degree (B) Clustering (C) Path Length

FIGURE 2. Illustrates network concepts. (A) Degrees are $d_A(g) = 3$ and $d_B(g) = 2$. (B) Clustering given by $c_A(g) = 0$, $c_B(g) = 1$, and $c_C(g) = 2/\binom{4}{2} = 1/3$. (C) Path length is $\gamma(A, B) = 5$ and $diam(g) = 6$.

2.1. **Density.** The *density* of a graph is the share of possible links that are present:

$$density(g) := \frac{1}{\binom{n}{2}} \sum_{i<j} A_{ij}.$$

It is useful to introduce some related concepts. The *degree* of a node is the number of neighbors it has and is the simplest measure of how central a node is in the network. A simple way to describe a network is to look at its *degree sequence*. This is the vector of degrees across all nodes:

$$d(g) = \sum_j A_{ij} = (d_1(g), ..., d_n(g))'.$$

Figure 2 illustrates this concept.

The *degree distribution* is the CDF of the degre sequence,

$$F_d(x) := \frac{|\{i : d_i \leq x\}|}{n}.$$

It plays a central role in parsimonious models of random matching and diffusion of behavior (Jackson and Rogers, 2007b; Jackson and Yariv, 2011; Galeotti and Rogers, 2013). Further, a large literature, mostly outside of economics, has paid considerable attention to the *shape* of the degree distribution. The researchers deal with questions as to whether there are "skinny tails" or "fat tails", as a means to examine whether there are particularly well-connected nodes in the network (e.g., see the survey by Albert and Barabasi (2002)).

Unsurprisingly, the mean of the degree distribution, the *average degree*, is a commonly used summary of how well-connected a network is

$$\bar{d}(g) := n^{-1} \sum_i \sum_j A_{ij}.$$

It is easy to see that in fact, the density is just a normalization of the average degree:

$$density(g) = \frac{1}{\binom{n}{2}} \sum_{i<j} A_{ij} = \frac{2}{n-1} \bar{d}(g).$$

A useful way to think about the density (or the expected degree) of a network is to do an asymptotic thought experiment. Imagine a sequence of distributions that generate networks, defined over a growing population of size n. Heuristically, we say that the sequence is *sparse* when the expected density converges to zero, since the graph is mostly empty. The usual case we imagine is one where the expected average degree is constant in n: individuals in large societies do not have many more links on average than in smaller societies. Notice that as long as $\mathrm{E}\bar{d} = o(n)$, under the density tends to zero.

On the other hand, we say that the sequence is *dense* if a typical node is linked to a number of nodes that is proportional to the number of potential partners in society. Dense graphs are those such that the expected average degree $\mathrm{E}\bar{d}^n \propto n$.

Throughout the literature, researchers vary as to what they mean when they say networks are sparse or dense. Sometimes sparsity refers to $\mathrm{E}\bar{d} = O(1)$, and sometimes sparsity refers to $\mathrm{E}\bar{d} = o(n)$. Correspondingly, sometimes density refers to $\mathrm{E}\bar{d} = \omega(1)$, and sometimes density refers to $\mathrm{E}\bar{d} = O(n)$. In this chapter I will be explicit about the rates when they are needed. Of course, these are all fictional constructions used for asymptotic thought experiments. After all, any realized network in the data is drawn from a finite sample.

Let us now turn to the empirical data. We can see from Table 1 that the average degree is low across all datasets (column 1). For instance, the Karnataka data has an average degree of 17 in villages that average 213 households. This is true also when we turn to the education data. The average degree in the Harvard dorm network is 10.8. Correspondingly, the density rates are rather low as well. Linking probabilities are in the 5% range, if not lower, and the highest probabilities are coming in the smallest networks. Therefore it seems reasonable that a researcher may want to use as a starting point assumptions that deliver sparse networks.

As we will discuss in later sections, sparsity can arise for a variety of natural reasons in economic settings: a lack of meeting opportunities, a budget constraint on time or resources, efficiency concerns, etc. It is important, therefore, that good network formation models respect this property of the data.

2.2. **Clustering.** The next concept concerns link correlation. If i has two neighbors j and k, what is the likelihood that j and k are linked? This is given by the *clustering coefficient*:

$$c(g) = n^{-1} \sum_i c_i(g), \; c_i(g) = \frac{\sum_{j<k} A_{ij} A_{ik} A_{jk}}{\binom{d_i(g)}{2}}.$$

A related measure looks at the global clustering coefficient. It addresses what share of all triples with ij linked and ik linked involve jk linked.

$$c^{global}(g) = \frac{\sum_{i<j<k} A_{ij} A_{jk} A_{ik}}{\sum_{i<j<k} \mathbf{1}\{A_{ij}A_{jk} + A_{ij}A_{ik} + A_{jk}A_{ik} > 0\}}.$$

Figure 2 computes clustering for three nodes, A, B, and C, which have a clustering of 0, 1 and 1/3 respectively.

Theoretical research has devoted considerable attention to addressing why agents may find it economically beneficial to have clustered links. First, there may be correlated tastes. Second, correlations in links could allow agents to sustain cooperation that they otherwise would not be able to. If i violates an agreement with j, then if they have a friend in common, k, then k could be enlisted to punish i as well (see Bloch et al. (2008); Karlan et al. (2009); Jackson et al. (2012)).

Turning to the data, we see that empirical networks exhibit high clustering, in the range of 0.2 to 0.3. Note that if every link was drawn independently with a probability given by the link density (e.g., 0.07), then the implied clustering should be 0.07. The empirical clustering rates are 3-4 times as high. This suggests that the underlying economic processes generate incentives to correlate links much more than would be expected at random.

In many ways, it should be unsurprising that real world networks have high clustering. After all, if there are incentives to link to a friend of a friend – perhaps it allows agents to better sustain cooperation – then we should see either three nodes all mutually linked or no linking at all.

3. Conditional Edge-Independence

Let us begin by looking at classes of models where each pair of individuals decide whether or not to link based only on their attributes and shocks. Conditional on these factors, the event that link ij forms is independent of the event that jk forms.

Under such an assumption, a simple model we could explore is one in which the marginal utility that i accrues by linking to j is given by

$$u_i(g + ij) - u_i(g - ij) = f(x_i, x_j) - \epsilon_{ij},$$

where x_i is a vector of individual attributes and ϵ_{ij} is a pair-level shock. Whether a link forms is given by

$$A_{ij} = \mathbf{1}\{f(x_i, x_j) - \epsilon_{ij} > 0\}. \tag{3.1}$$

For simplicity, let us take $f(\cdot, \cdot)$ to be symmetric, $\epsilon_{ij} = \epsilon_{ji}$, and ϵ_{ij} as drawn i.i.d. across the $\binom{n}{2}$ pairs with Type I extreme value. It follows that

$$\mathrm{P}(A_{ij}|x_i, x_j) = \frac{\exp f(x_i, x_j)}{1 + \exp f(x_i, x_j)}.$$

Under suitable assumptions on $f(\cdot, \cdot)$ – for instance, the probability of a link increases in the similarity of i and j, $f(x_i, x_j) = -\beta|x_i - x_j|$ – this allows for straightforward estimation of linking probabilities.[7]

In general, let us say $\epsilon_{ij} \sim \Lambda(\cdot)$ and so we have

$$\mathrm{P}(A_{ij}|x_i, x_j) = \Lambda(f(x_i, x_j)).$$

[7]One could simply do a logistic regression of A_{ij} on $|x_i - x_j|$.

Such a model does not include individual-level unobserved heterogeneity and this might be a serious limitation. After all, we think individuals who are altruistic may naturally be more likely to make friends, and this unobserved trait might systematically alter the distribution of links even conditional on observed characteristics.

To get at questions such as this, we could imagine adding in individual-level unobserved heterogeneity in the model. Consider extending the above to

$$u_i(g + ij) - u_i(g - ij) = f(x_i, x_j) + \nu_i + \nu_j - \epsilon_{ij},$$

where ν_i are unobserved fixed effects for each individual. Then the probability that a link ij forms is given by

$$p_{ij}(x_i, x_j) = \Lambda(f(x_i, x_j) + v_i + v_j).$$

The challenge for the researcher in this case is to not only estimate $f(\cdot, \cdot)$, but to do so either (a) in the presence of nuisance parameters ν_i, or (b) estimate ν_i as well in case they are of economic interest. They certainly will be if the researcher is interested in estimating counterfactual distributions of networks.

3.1. **Erdős-Rényi model.** To understand basic properties of conditional edge independent models, it is useful to remove covariates from the utility functions. It is easy to see that every link forms independently with common probability p, since

$$\mathrm{P}(A_{ij} = 1) = \mathbf{1}\{\alpha > \epsilon_{ij}\}$$

for some constant α.

This gives us the Erdős-Rényi random graph model, which is easily the most studied graph formation model (Erdős and Rényi, 1959). The Erdős-Rényi model has a number of special properties that correspond to specific restrictions on the resulting network. I focus on three implications here.

(1) The expected number of links that a given node has is $(n-1)p$.
(2) The probability that two neighbors of i, j and k, are linked is p.
(3) The probability that i, j and k are mutually linked is given by p^3.

While this is certainly not an exhaustive list of properties of Erdős-Rényi networks, it highlights the following tension that is relevant for empirical work. First, as noted above, empirical networks are often sparse. What this means is that the expected number of links that a node has is far smaller than n. When choosing how to model this, asymptotically, we typically think of a sequence $p = p_n$, which decreases in n. For example, if $p_n = d/n$, then for very large graphs, the expected degree is d irrespective of the size of the network.

Second, empirical networks are often clustered. Since clustering is essentially the share of times that jk forms given that ij and ik are links, notice that in this model, clustering has to be p_n.

Thus, if networks exhibit sparsity, so p_n is any sequence that tends to zero, then by definition there is no clustering in the limit. On the other hand, if networks exhibit non-vanishing clustering, there exists a lower bound $\underline{p}$ for the linking probability, bounded away from zero. In this case, the expected degree is at least $(n-1)\underline{p}$, which grows

linearly in n. This implies that in networks of size 1000 versus 100, individuals will be linked to 10 times as many people.

It is worth making two observations about this model. First, even if the implied network formation parameter p_n is changing in n, the researcher may be able to estimate it well in the sense that (a) $|\widehat{p}/p_n - 1| \xrightarrow{\text{P}} 0$ and (b) that, when appropriately normalized, the parameter estimate is asymptotically normally distributed

$$\sqrt{\frac{\binom{n}{2}}{p_n(1-p_n)}}(\widehat{p} - p_n) \rightsquigarrow \mathcal{N}(0,1),$$

under mild assumptions. The key assumption is that p_n does not tend to zero too fast, with the sharp threshold being that $p_n \cdot \binom{n}{2} \to \infty$.

Second, how shall we interpret p_n tending to zero when it comes to looking at microfoundations? One possible view is to think about a world where agents have (essentially) a selectivity threshold because they have a fixed budget of time. The population with which an agent interacts is indexed by ϵ_{ij} which gives them the quality of the match. In larger populations, one can screen more selectively, and this would correspond to α_n being increasingly negative along the sequence, implying that in larger networks, higher shocks are required to justify a link. Another interpretation is that individuals have a fixed time budget and therefore can only ever meet a small subset of the population n, irrespective of n.

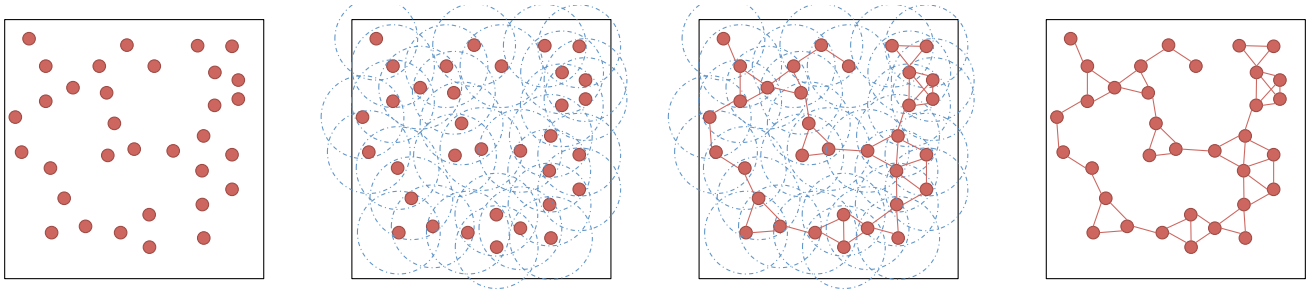
All remaining models in this section will add different layers of heterogeneity in terms of observables and unobservables. We will explore the degree to which the aforementioned trade-off between sparsity and clustering can or cannot be alleviated by these approaches.

3.2. **Random Geometric Graphs.** An alternative to the Erdős-Rényi models widely studied are random geometric graphs (Penrose, 2003). Here nodes are stochastically placed in some embedding space and then link to each other if they are close enough in this space. For instance, if nodes are formed close enough in education-wealth space, then they link, otherwise they do not. The crucial point is that this places geometric restrictions, from the embedding space, on the topologies of the networks likely to arise. For instance, for five nodes that are close enough, it may be impossible to have any subgraph between them except a clique.

A simplified version of the model is as follows. Let n nodes be assigned locations uniformly at random in $[0,1]^2$. A link exists between a pair if they are at most r_n apart:

$$A_{ij} = \mathbf{1}\{d(x_i, x_j) \leq r_n\}.$$

Having $r_n \downarrow 0$ corresponds to domain increasing asymptotics. This is depicted in Figure 3. One advantage of this perspective is that there are very well understood properties that can exploit geometry. As such, the RGGs have well-understood subgraph counts with laws of large numbers and central limit theorems for the numbers of various subgraphs that should emerge (Penrose, 2003).



(A) Nodes (B) Radii (C) Links (D) Network

FIGURE 3. This shows a random geometric graph formation process.

Notice the formal relationship between a conditional edge independence model and the random geometric graph

$$\begin{aligned} A_{ij} &= \mathbf{1}\{\alpha - \beta\,|x_i - x_j| - \epsilon_{ij} > 0\} \\ &= \mathbf{1}\{|x_i\beta - x_j\beta| + |\epsilon_i - \epsilon_j| \leq \alpha\}, \quad \text{assuming } \epsilon_{ij} := -\,|\epsilon_i - \epsilon_j| \\ &= \mathbf{1}\{d((x_i\beta, \epsilon_i)', (x_j\beta, \epsilon_j)) \leq r\}, \quad \text{setting } r := \sqrt{\alpha}. \end{aligned}$$

Under appropriate assumptions on the distributions of x and ϵ, the techniques used to study random geometric graphs and the properties of the resulting networks may be useful to help us understand how the distribution of observed and unobserved shocks may influence the globabl properties of the network in a meaningful way.

3.3. **Observed characteristics.** The most commonly used approach to model correlation in link structure is to rely on observable node characteristics (Comola and Fafchamps, 2013; Fafchamps and Gubert, 2007a,b). Since factors such as geographic location (as in trade models), education, ethnicity/caste, wealth, and so on are likely to influence whether agents form an economic relationship, it may very well be that the clustering in a network could be explained by such observable factors.

In this case, if there are covariates $x = (x_1, ..., x_n)'$ jointly drawn from a distribution $F_x(\cdot)$, one could consider

$$p(x_i, x_j) := \mathrm{P}(A_{ij}|x_i, x_j).$$

Usually this is a parametric model, but need not be. This leads to

$$\mathrm{P}_p(g|x) = \prod_{i<j} p(x_i, x_j)^{A_{ij}} (1 - p(x_i, x_j))^{1-A_{ij}}.$$

The expected graph would exhibit higher clustering when i,j, and k are such that $p(x_i, x_j)$, $p(x_k, x_j)$ and $p(x_i, x_k)$ are all high.

The extent to which this is a good or a bad approach depends on the x's that the econometrician has at her disposal. Since, in principle, given that correlations in the x's are allowed, and having access to a large number of unobservable and observable x's could always represent the right probability distribution, what really matters is how much of the correlation is described by what the researcher sees. Chandrasekhar

and Jackson (2014), for instance, look at data and show that including a large vector of covariates of Indian village households (assets, caste, geography) is not enough to generate the clustering and linking patterns in the empirical data.

To get an intuition as to what could go wrong, let us look at *Block models* (Holland, Laskey, and Leinhardt, 1983), which cover a specific parametrization of introducing covariates. Imagine that the only characteristic of an agent that matters to network formation is her group (e.g., caste, or caste × gender) and there are K communities. The adjacency matrix is formed as follows:

(1) An individual belongs to group (or community) k with probability $\pi(k)$. The community is denoted x_i.
(2) Conditional on group assignment, the probability that nodes i and j are linked depends on communities x_i, x_j: $p(x_i, x_j)$.

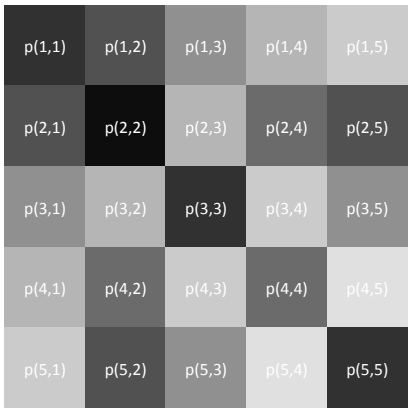


FIGURE 4. The adjacency matrix of a block model with $K = 5$, illustrated in the symmetric case. $p(k, l)$ is the linking probability for nodes from groups k and l. Darker shading indicates higher probability of linking.

Figure 4 illustrates the model. Notice that within a group an Erdős-Rényi graph $p(x_i, x_i)$ forms. Across groups, there is a fixed linking probability given group identities as well. Therefore, the intuitive limitation of such a model, holding fixed the number of groups, is that we are simply looking at interconnected Erdős-Rényi graphs. This means that unless within a group networks can be thought of as dense, the clustering-sparsity tradeoff will persist.

3.4. **Community detection.** Now we turn to a more difficult problem, wherein individuals are members of communities/groups as before but these groups themselves are unobservable. The researcher's task is to (1) estimate the community structure and then (2) recover the linking probabilities. The group membership structure can be thought of as a collection of nuisance parameters.

There is a large literature on community detection in networks spanning computer science, sociology, statistics and, more recently, economics. Surveying this literature is beyond the scope of this chapter, and I direct the reader to Clauset et al. (2004), Newman (2004), Jackson (2008), and Leskovec, Lang, and Mahoney (2010) for more in-depth reviews. The idea is straightforward, though, and has origins in sociology Airoldi, Blei, Fienberg, and Xing (2008); Lazarsfeld, Henry, and Anderson (1968)). Agents in a social network belong to some latent communities that may or may not coincide with observable characteristics. For instance, individuals in a village may share risk with others, despite being from different castes or researchers may form co-authorships across schools. A natural question is how to identify the underlying so-called communities that drive the observed network.

The logic of the identification of communities is simple but subtle. At a high level, we want to say that sets of nodes are likely to be in the same community if they have a lot of links within the set. At the same time, a natural requirement is that linking within the set should somehow exceed linking between members of different sets.

Much of the literature has focused on specifying objective functions to make precise what we mean by community and community detection. Because this involves assigning each node to a community, if there are K communities, this involves searching over a discrete space of size n^K. Therefore, the computer science literature has focused on developing simple algorithms to approximate solutions to the various objective functions that have been specified. These computational concerns are beyond the scope of this chapter.

Instead, I will discuss community detection from the perspective of statistics. After discussing a simple model of network formation based on unobserved communities, I will discuss the commonly studied Newman and Girvan objective function (frequently called NG Modularity) as well as an alternative, likelihood-based method by Bickel and Chen (2009), who present a non-parametric view of community detection models and relate them to another crucial question. They ask whether any reasonable probability distribution for a network can be covered by a model like the one specified below. The answer is, in some sense, "yes".

Specifically, as discussed by Bickel and Chen (2009) and studied extensively based on results for exchangeable arrays (Aldous (1981); Aldous (1985); Kallenberg (2006); Diaconis et al. (2008)), under very mild assumptions – basically on an unlabeled random network, the probability distribution of any realized network structure is the same across permutations – one can represent any probability distribution as

$$A_{ij} = f(\alpha, \xi_i, \xi_j, z_{ij})$$

where α, ξ_i, ξ_j, and z_{ij} are i.i.d. $U(0,1)$, $z_{ij} = z_{ji}$ and $f(\cdot)$ is a function symmetric in its second and third arguments.[8]

This provides theoretical motivation for an interesting avenue pursued by Bickel and Chen (2009) and Bickel, Chen, and Levina (2011). They look at parametric approximations to these non-parametric models. The basic idea is that every node is assigned to a community (or group) $1, ..., K_n$ independently of others. And then conditional on the

[8]Note that $f(\cdot)$ is not unique and, further, often the models do carry labels.

vector of assignment to communities, edges are generated independently based on only group-identities. The researcher, of course, does not observe groups $x = (x_1, ..., x_n)$, and therefore is interested in estimating π, and has potential nuisance parameters x. The thought experiment is one where the number of communities grow as $n \to \infty$, though they do not fully develop these results in the present work.

The model implies that the probability of seeing a given graph is

$$\mathrm{P}_{\pi,p}(g) = \sum_x \left\{ \prod_{i<j} p(x_i, x_j)^{A_{ij}} (1 - p(x_i, x_j))^{1-A_{ij}} \right\} \prod_{i \in V} \pi(x_i).$$

There are $n + \binom{K_n}{2} + K_n$ parameters of the model: $x = (x_1, ..., x_n)$ – the unobserved groups – as well as linking probabilities $p(m, l)$ for groups m, l and the group assignment probabilities, $\pi(\ell)$.

The main task, then, is to decide how to detect these groups. The first class of methods used in the literature come from computer science and focus on variations of minimum ratio cuts. The idea is simple. Given g, we can pick a collection of links into sets U and U' such that

$$\frac{\sum_{i \in U, j \in U'} A_{ij}}{|U| \cdot |U'|}$$

is minimized. The numerator is the number of links bridging members across the two groups, and the denominator is the product of the sizes of sets. This is intimately related to the conductance of the graph, which has also been used as an alternative criterion function. Nonetheless, actually solving for these cuts have long known to be NP-complete. Therefore, the literature has focused on approximation algorithms using, for instance, signs of eigenvectors associated with the adjacency matrix or related quantities and then applying repeated bisections. One limitation of this approach is that without modification, it does not adequately tell us about the group size or counts.

This observation motivates the second method, which is perhaps the most popular objective function used to partition the graph currently. The Girvan and Newman (2002) modularity is an objective function that picks community assignment to maximize the share of edges that link within a group minus the expected share that would link within a group given random location of links, holding fixed the community assignment of nodes.

It is useful to see their objective function, often called (NG)-modularity. First compute a matrix encoding the number of links between two blocks k and l

$$O_{k,l}(x|A) := \sum_{i<j} A_{ij} \mathbf{1}\{x_i = k, x_j = l\}.$$

Let $D_k := \sum_l O_{k,l}$ be the sum of degrees for those in community k. Then the objective function is given by

$$Q_{NG}(x|A) := \sum_{k=1}^{K} \frac{O_{kk}}{L} - \left(\frac{D_k}{L}\right).$$

What is conceptually wrong here? The objective function, in a sense, focuses on the diagonal of O only. Therefore it is giving up information by considering how often within group links stay within group versus the random allocation benchmark.

The approach taken by Bickel and Chen (2009) is to approach this from a likelihood perspective. Let n_k count the number of nodes in community k. Similarly, let

$$n_{a,b} := n_a n_b \text{ if } a \neq b,\ n_{a,b} := n_a(n_a - 1) \text{ if } a = b.$$

Then what is the likelihood under a known assignment of observing A? It is given by

$$Q_{LM}(x|A) := \frac{1}{2} \sum_{1 \leq a,b \leq K} \{O_{a,b} \log p(a,b) + (n_{a,b} - O_{a,b}) \log(1 - p(a,b))\}.$$

It must be the case that the likelihood is optimized by the $p(a,b)$ choices, and it is easy to see that this implies $p(a,b) = \frac{O_{a,b}}{n_{a,b}}$. Therefore,

$$Q_{LM}(x|A) := \frac{1}{2} \sum_{1 \leq a,b \leq K} n_{a,b} \left\{ \frac{O_{a,b}}{n_{a,b}} \log\left(\frac{O_{a,b}}{n_{a,b}}\right) + \left(1 - \frac{O_{a,b}}{n_{a,b}}\right) \log\left(1 - \frac{O_{a,b}}{n_{a,b}}\right) \right\}$$

gives us our (profile) likelihood.

Bickel and Chen (2009) study a general class of objective functions that nest LM and NG. Under an assumption that the underlying network formation process leads to dense enough graphs (the expected average degree grows faster than $\log n$) and that the objective function is uniquely maximized with smooth and bounded derivatives, they show that the assignments are correct with probability tending to 1 as $n \to \infty$.

Specifically, they show that for the LM objective and under some regularity conditions which NG sometimes meets,

$$\mathrm{P}(\widehat{x} = x) \to 1$$

and, further,

$$\sqrt{n}\,(\widehat{\pi} - \pi) \rightsquigarrow \mathcal{N}(0, \Sigma)$$

where Σ is a variance matrix which is unaffected by the estimation of x. They also show that p can be recovered with estimators that are consistent and asymptotically normally distributed.

What is particularly interesting is that the NG modularity can fail the identification/derivative conditions when, in fact, for every community there is higher linking probability within the community than the total sum of linking probabilities to any other community. The way that the inconsistency works is that merging two of the sparser communities can actually increase the modularity. One can see that this would be penalized by the likelihood approach. In that sense, the LM approach is a more robust objective function and, more generally, the statistical approach helps us see when and why other objective functions may fail.

Despite this, there are two main limitations to the entire enterprise. First, the computation is known to be challenging. As noted in Bickel and Chen (2009), it is NP-hard. While they and others report that certain algorithms work well, this inevitably makes implementation difficult. Second, the present theory is geared to studying dense graphs. As noted by the authors, it remains to be seen how valuable this approach could be for sparse networks, since classification will be imperfect.

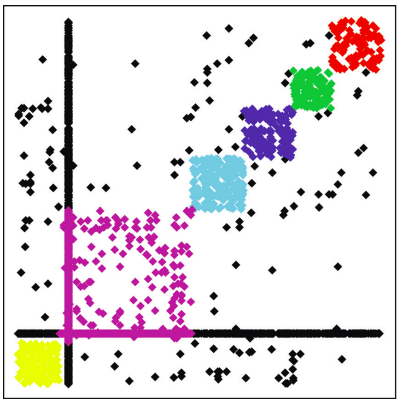


FIGURE 5. Figure 4 of Bickel and Chen (2009) reproduced. This presents the adjacency matrix, which consists of 621 individuals in a telephone communication network. The identified communities are of heterogenous sizes. Colors indicate within-group links and black indicates across-group links.

3.5. **Unobserved node-level heterogeneity.** We now move to models where individual-specific unobserved heterogeneity is introduced, at the cost of additional parametric assumptions.

Individuals are often modeled to be heterogenous. They vary preferences and beliefs, and these differences can therefore systematically change one individual's decisions relative to another's in an otherwise identical circumstance. The models we have discussed so far have skirted this issue by making all heterogeneity observable or unobserved at a group level.

Nevertheless, in certain contexts it may very well be right to assume that individuals vary in their gregariousness, which the researcher will never observe, but gregarious individuals are simply more likely to build links.

Further, it could be the case that gregariousness is correlated with certain observable attributes, and therefore the researcher faces a standard problem that arises in settings with fixed effects: systematic unobserved heterogeneity that can be correlated with observables may cause the researcher to mis-estimate how the observables influence the outcomes.

Graham (2014a) studies precisely this problem. He considers a simple model where the probability of a link ij being formed conditional on a vector of covariates is given by

$$\mathrm{P}_{\beta,\nu}(A_{ij}=1|x)=\frac{e^{x_{ij}'\beta+\nu_i+\nu_j}}{1+e^{x_{ij}'\beta+\nu_i+\nu_j}}. \tag{3.2}$$

Observe that every agent has a fixed effect ν_i that shifts the probability of a given link being formed. Since each agent is participating in $n-1$ linking decisions, we can think of each linking decision the way we would think about time in a panel data setting.

There are two goals that we might have in such a setting. First, we may want to understand how observables x_{ij} influence linking probabilities by estimating β. Since

there are n individuals, this represents an n dimensional (nuisance) parameter. The natural difficulty in this problem is that the overall number of parameters in the model grows in n since there are $\dim(\beta) + n$ parameters. However, techniques developed by Chamberlain (1982) allow us to skirt the issue by using a conditional maximum likelihood estimator (CMLE).

A second goal might be to do counterfactuals. If we want to study how a shift in the distribution of x changes the overall resulting network structure, we must actually estimate ν_i. In this case, the researcher needs to use the fact that she observes individuals making $n-1$ linking decisions in order to estimate the n fixed effects. This presents a formidable and non-standard statistical problem.

The solution to this problem lies in a literature in statistics going back to Holland and Leinhardt (1981). That literature ends with Chatterjee, Diaconis, and Sly (2010), who show how one can indeed consistently estimate all the ν_i in the above problem. I now discuss the basics of this statistical work before returning to Graham (2014a).

It is useful to address this problem from a different perspective. Let us begin with the observation that a key aspect of networks is its degree sequence: $d^n(g) = (d_1, ..., d_n)$. Therefore, a simple question to ask is whether it is possible to develop a model that, as the parameter varies over its domain, can capture all possible degree sequences. The answer is yes, and comes from a class of models introduced by Holland and Leinhardt (1981), but studied in greater detail by Chatterjee et al. (2010).

Chatterjee et al. (2010)'s first main result here is to formalize the answer to this question. We consider a sequence $\{g_n\}$ of random graphs each chosen uniformly at random given d^n. This implies that the degree sequence is a sufficient statistic.

Additionally, let $f(\cdot)$ be any non-increasing function on $[0,1]$. Then there is a unique function $h : [0,1] \to \mathbb{R}$ such that

$$W(x,y) := \frac{\exp(h(x)+h(y))}{1+\exp(h(x)+h(y))},$$

$$f(x) = \int_0^1 W(x,y)dy,$$

and $\{g_n\}$ converges, in the sense described by Lovász and Szegedy (2006), almost surely to the limit represented by $W(x,y)$. This limit object is called a graphon.

A sequence of adjacency matrices can be placed on the unit square, and the limit is the case where we have each $x \in [0,1]$ corresponding to an agent, and therefore $W(x,y)$ encodes whether or not agents (x,y) are linked. Just as one would sum over every agent's friends to obtain a vector of degrees, one can integrate $W(x,y)$ over y to obtain the analogous limiting degree distribution, $f(\cdot)$.

What the result says is that if we consider a sequence of networks with a limit (in a specific technical sense that I will not cover here) given by $W(\cdot,\cdot)$, for any limit degree distribution, given by $f(\cdot)$, one can find a unique function $h(\cdot)$ that pins down $W(\cdot,\cdot)$ in the above.

Therefore, one may wonder whether we can compute the function $h(\cdot)$. Chatterjee et al. (2010) show that the answer is yes, and comes from the following, well-studied,

model called the β-model.[9] Let $\nu = (\nu_1, ..., \nu_n) \in \mathbb{R}^n$ and set

$$p_{ij} = \frac{\exp(\nu_i + \nu_j)}{1 + \exp(\nu_i + \nu_j)}.$$

Notice that this is exactly the model that Graham (2014a) later studies, only without covariates. Again, the technical challenge is that there are n parameters with only $\binom{n}{2}$ observations. This makes the problem somewhat non-standard, and a major contribution of the Chatterjee et al. (2010) paper is to address this challenge and demonstrate that the parameters can be consistently estimated.

The two main results in Chatterjee et al. (2010) are:

(1) *Consistency:* with probability $1 - O(n^{-2})$,

$$\max_{1 \le i \le n} |\widehat{\nu}_i - \nu_i| \lesssim \sqrt{\frac{\log n}{n}} \to 0.$$

(2) *No degree sequence is left out:* All possible expected degree sequences can be covered as ν varies over $\mathbb{R}^n$.

This implies that despite the challenge due to the rapidly growing number of parameters, they can consistently be estimated. Further, any limiting degree distribution can be modeled by the above for a suitable vector of parameters ν.

It is important to note that their result applies to dense networks and not to sparse networks. That means that the number of links that a typical node is a part of is on the order of n. To see why this matters, consider a simplified case where each agent had a constant order of friends. In this case, $f(x) = 0$ at every x in the statement of their first result. Next, we turn to the argument as to why one cannot recover the ν_i consistently in this case. A crucial piece of the argument is to show with probability going to one that every node's degree is close to its expected degree under the model. This requires the expected degree to grow at rate n, because otherwise, there is no growing information that allows us to precisely estimate every individual fixed effect. As noted by the authors, the arguments do not follow through for sparse networks precisely because there is not enough information in the system to estimate each parameter (for each node) tightly.

Returning to Graham (2014a), we add covariates to the problem and are interested in recovering both the fixed effects as well as the parameter driving the covariate effect. He shows the following:

(1) $\widehat{\beta} \xrightarrow{\text{P}} \beta$,
(2) with probability $1 - O(n^{-2})$, $\max_{1 \le i \le n} |\widehat{\nu}_i - \nu_i| \lesssim \sqrt{\frac{\log n}{n}}$,
(3) and for I_0 the Fisher information for β,

$$n\left(\widehat{\beta} - \beta\right) \rightsquigarrow \mathcal{N}\left(\sqrt{2}B_0, 2I_0^{-1}\right).$$

Introducing these nuisance parameters generates a bias term in the asymptotic distribution, which then can be exactly characterized and therefore subtracted out.

[9] Usually the model is written with parameter β, though we use ν to call attention to its economic interpretation as a fixed effect.

3.6. **Discussion.** In this section, I have reviewed basic models where links are conditionally independent. The fundamental tension here, which we can see from the Erdős-Rényi model, is that if we model realistic levels of clustering, then the network must be dense with expected degree being on the order of n. However, this tends to be inconsistent with the sparsity of empirical networks.

While observable covariates can get us part-way there, they are far from sufficient. It is possible, in theory, to model any distribution over networks as a function of two individual-specific latent variables and a pair-level latent variable. Estimating such a model directly from the data is infeasible, but a reasonable approach to get at this was developed by Bickel and Chen (2009), where individuals belong to communities and the number of communities grows asymptotically. Setting aside computational complexity issues, we find that a major limitation of their otherwise general result is that it depends on having a sufficiently dense network. Nonetheless, their results require that with probability approaching one, the assignment to communities is exact. Perhaps requiring a less stringent assignment criterion would allow for more sparse network formation distributions to be modeled.

A different form of unobserved heterogeneity is studied by Chatterjee et al. (2010) and Graham (2014a), who emphasize the degree to which this is economically important since fixed effects may be correlated with observables, thereby influencing the researcher's estimates about the degree to which there is homophily in covariates. Again, however, the techniques required to be able to estimate the fixed effects lean heavily on the networks being dense, with an expected degree that is on the order of n.

The density requirements of the Bickel and Chen (2009) approach is less stringent than those required by the β-model. Perhaps this is not too surprising because, while Bickel and Chen (2009) are able to recover assignment to group correctly, they reduce the dimensionality of the driving probability distribution to just a collection of groups and probabilities of linking within and across groups. In that sense, there is more information per fundamental parameter in their model, as compared to the β-model case.

Many of the models studied in the next section relax the density requirements, by directly entangling agent payoffs with choices of their neighbors. In turn, links are correlated in a more direct manner without giving up on sparsity. On the other hand, these models raise more complicated problems when it comes to statistical inference, identification, and equilibrium selection.

4. Higher Order Dependence

In this section, we look at network formation models that more directly allow for correlation in links. We will need to grapple with whether these models provide the researcher with more information, in the sense of admitting laws of large numbers and

central limit theorems, as $n \to \infty$. Finally, point identification itself may become impossible without strong assumptions on preferences or equilibrium selection.

4.1. **Exponential random graph models (ERGMs).**

4.1.1. *Dynamic, myopic best response.* There is an important relationship between strategic network formation models and potential functions, explored across a number of papers including Jackson and Watts (2001), Butts (2009), Mele (2013), and Badev (2013). We will use this as a starting point in our discussion of a class of models called exponential random graph models (ERGMs), which are among the most widely used models of network formation.

Recall the definition of a potential game. Consider a set of individuals with utility functions u and a strategy set Y. A game is a potential game if there exists a function $f(\cdot,\cdot)$ such that

$$u_i(y_i', y_{-i}) - u_i(y_i, y_{-i}) = f(y_i', y_{-i}) - f(y_i, y_{-i}).$$

for every player i and $y, y' \in Y$.

Butts (2009) and Mele (2013) study the steady-state of simple models where nodes can dynamically revise their linking decisions, given the current network, but are not forward looking. If the utility function of the players admits a potential function, the resulting steady-state distribution corresponds to a specific distribution with a sufficient statistic given by the potential function.

That is, if players choose to revise their links via a logistic choice rule, and updating opportunities arise sequentially with every pair ij being selected with positive probability, then the sequence $\{g_t\}_{t\in\mathbb{N}}$ forms a Markov chain with equilibrium distribution:

$$\mathrm{P}_{f(\cdot)}(g) = \frac{\exp(f(g))}{\sum_{g'} \exp(f(g'))}.$$

Thus, if agents are in a dynamic process where their payoffs are such that there is a potential function for the marginal utility of their links, then the stationary distribution of networks has the above form and is what the econometrician draws from.

Example **1** (Links and triangles)**.** *Imagine that an individual receives payoffs β_L for every link that she is part of and β_T for every triangle. In this case*

$$u_i(g) = \beta_L \cdot \sum_j A_{ij} + \beta_T \cdot \sum_j A_{ij}A_{jk}A_{ik}.$$

Define a function

$$f(g) := \beta_L \cdot \sum_i \sum_j A_{ij} + \beta_T \cdot \sum_i \sum_{j<k} A_{ij}A_{jk}A_{ik}.$$

It is clearly a potential function, since

$$u_i(g+ij) - u_i(g-ij) = \beta_L + \beta_T \cdot \sum_k A_{ik}A_{jk} = f(g+ij) - f(g-ij).$$

We now describe this for more general payoffs. The core argument is owed to Butts (2009) and Mele (2013). Chandrasekhar and Jackson (2014) extended the result to other subgraphs. Badev (2013) observes that adding an endogenous action to the mix does not change the fundamental result. He looks at a case where utility depends not only on links to others, but also on one's actions as well as one's neighbors' actions.

Assume that payoffs are additively separable in subgraphs that i is a member of:

$$u_i(g, x; \beta) = \sum_{g_\ell, i \in g_\ell} v(g_\ell, x_\ell; \beta_\ell),$$

where x_ℓ is a vector of covariates corresponding to the members of subgraph g_ℓ. Observe that this allows for an agent's utility to depend on "friends of friends" as desired in a number of models (Goldsmith-Pinkham and Imbens (2013); Mele (2013); Badev (2013)) by including subgraphs of the form $g_\ell = \{ij, kj\}$. However, it also allows utility to depend on more complex subgraphs such as cliques.

The links and triangles example had no covariates, and so corresponded to

$$v(\{ij\}; \beta) = \beta_L \text{ and } v(\{ijk\}; \beta) = \beta_T.$$

Time is discrete, and in each period, at most one pair of individuals meet. Every ij ex ante have some positive probability of meeting this period, though the actual meeting probability can be history-dependent.

The stability concept used is pairwise stability with transfers[10]:

- $ij \in g$ implies that $u_i(g) + u_j(g) \geq u_i(g - ij) + u_j(g - ij)$, and
- $ij \notin g$ implies that $u_i(g) + u_j(g) \geq u_i(g + ij) + u_j(g + ij)$.

The utility function and the stability concept together imply the existence of a potential function under pairwise stability with transfers:

$$f(g) := \sum_{g_\ell \subset g} 2v(g_\ell, x_\ell; \beta_\ell).$$

One can check that for any g and $ij \in g$:

$$f(g) - f(g - ij) = (u_i(g) + u_j(g)) - (u_i(g - ij) + u_j(g - ij)).$$

Network formation occurs as follows. Agents can choose to form links in pairs. They choose to add a link whenever their *mutual* gain is positive. This is important, because this allows them to bargain and make side payments (either in cash or by exchange of favors) to add links whenever those links are mutually beneficial. Allowing for this may be important from theoretical first principles.

The dynamics are as follows. At period t the process is at some network g^t. Let g^t_{-ij} be the networks excluding link ij. In every period, there is some positive probability $p(ij, g^t_{-ij})$ of each given pair being recognized – where the two relevant agents meet. This is general in that the meeting process can depend on the present graph as well as characteristics of i and j.[11]

[10]This definition is from Bloch and Jackson (2006) and is related to the definition of pairwise stability allowing for side payments that appears in the conclusion of Jackson and Wolinsky (1996).

[11]This allows their meeting probability to depend on whether they have friends in common, how many friends in common they have, their covariates, other aspects of the network in general, etc.

In period t, the chosen i and j decide whether to add or delete the link, conditional on g_{-ij}^{t-1}. In the usual way by using extreme value disturbances, one can say that the probability that the link is added/kept is a logistic function of the mutual value of the link:

$$\frac{\exp\left(u_i(g_{-ij}^t + ij) + u_j(g_{-ij}^t + ij)\right)}{\exp\left(u_i(g_{-ij}^t + ij) + u_j(g_{-ij}^t + ij)\right) + \exp\left(u_i(g_{-ij}^t - ij) + u_j(g_{-ij}^t - ij)\right)}. \tag{4.1}$$

The dynamic process generates a sequence $\{g^t\}_{t\in\mathbb{N}}$ that is an aperiodic and irreducible Markov chain over the set of networks on n nodes. Therefore, there is a unique steady-state distribution. Moreover, it is a reversible Markov chain, and the unique steady-state distribution is given by

$$\mathrm{P}(g) = \frac{\exp(f(g))}{\sum_{g'}\exp(f(g'))} = \frac{\exp\left(\sum_{g_\ell\subset g} 2v(g_\ell, X_\ell)\right)}{\sum_{g'}\exp\left(\sum_{g_\ell\subset g'} 2v(g_\ell, X_\ell)\right)}.$$

Returning to our links and triangles example, the distribution is given by

$$\mathrm{P}_{\beta_L,\beta_T}(g) = \frac{\exp\left(\beta_L \cdot S_L(g) + \beta_T \cdot S_T(g)\right)}{\sum_{g'}\exp\left(\beta_L \cdot S_L(g') + \beta_T \cdot S_T(g')\right)}$$

where $S_L(g)$ is (twice) the total number of links in the network and $S_T(g)$ the triangles. The goal of the researcher would then be to estimate β_L, β_T.

In sum, what we see is that a simple, myopic best response dynamic where the payoffs admit a potential function corresponds to an invariant distribution of networks which has the above form. Specifically, this means that the network that the econometrician observes is drawn from a distribution where the sufficient statistic is given by the potential function itself. One way to think about this is that given $f(g)$, any graph g can be drawn uniformly at random.

4.1.2. *ERGMs.* A literature spanning several disciplines (computer science, economics, sociology, and statistics) has turned to a class of models called exponential random graph models (ERGMs). See, e.g., Frank and Strauss (1986); Wasserman and Pattison (1996); Mele (2013).

The basic idea is that an ERGM specifies a vector of sufficient statistics, usually corresponding to counts of subgraphs in a network, and then formulates the probability as depending on the subgraph count with all networks with the same sufficient statistic value being drawn with equal probability. Above, I have described one approach to providing microfoundations for these models.

Let $S(g)$ denote a vector of characteristics of a newtork. For example,

$$S(g) = (S_L(g), S_T(g))' = \left(\sum_{i<j} A_{ij}, \sum_{i<j<k} A_{ij}A_{jk}A_{ki}\right)'$$

is a vector of the total number of links and the total number of triangles in the graph.

The ERGM specifies the probability of observing a network with a given vector of attributes S is given by[12]

$$\mathrm{P}_\beta(g) = \frac{\exp(\beta \cdot S(g))}{\sum_{g'} \exp(\beta \cdot S(g'))}. \tag{4.2}$$

ERGMs have become widely used for a number of reasons. First, they provide an intuitive formulation. They focus on aspects that researchers believe are important in network formation and that can encode rich types of interdependencies. Second, due to the Hammersley-Clifford theorem (Hammersley and Clifford, 1971), it turns out that any probability distribution over networks can be represented in this way. This result has very little content in practice, as it makes no claim about the complexity of the set of sufficient statistics. Trivially, any distribution can be captured by some (potentially very complicated) vector of sufficient statistics. Third, as described above, recent work provide utility-based microfoundations for such models (Butts, 2009; Mele, 2013).

It should be immediately obvious that an Erdős-Rényi random graph must have a sufficient statistic which is the number of edges in the network. This is because the binomial is in the exponential family, and the sufficient statistic is the number of successes (in this case, links).

EXAMPLE **2** (Erdős-Rényi)**.** *Let $S(g) = \sum_{i<j} A_{ij}$ denote the number of links in the graph. As we saw before, the likelihood of the Erdős-Rényi model is*

$$\mathrm{P}_p(g) = p^S (1-p)^{\binom{n}{2}-S},$$

which can be written as

$$\mathrm{P}_p(g) = \exp\left(\log \frac{p}{1-p} \cdot S(g)\right) \cdot (1-p)^{\binom{n}{2}}.$$

Setting $\beta := \log \frac{p}{1-p}$ and observing $(1-p)^{-\binom{n}{2}} = \sum_{g'} \binom{\binom{n}{2}}{S(g')} p^{S'} (1-p)^{\binom{n}{2}-S'}$, yields

$$\mathrm{P}_\beta(g) = \frac{\exp(\beta \cdot S(g))}{\sum_{g'} \exp(\beta \cdot S(g'))}.$$

Similarly, the β-model we know posits that given the degree distribution, the network is drawn uniformly at random. It is easy to see that $S(g) = d(g) = (d_1(g), ..., d_n(g))'$ in this case.

EXAMPLE **3** (β-model)**.** *This is an ERGM with an n-dimensional parameter, β, with*

$$\mathrm{P}_\beta(g) = \frac{\exp(\beta \cdot S(g))}{\sum \exp(\beta \cdot S(g')} = \frac{\exp(\beta \cdot d(g))}{\prod_{i<j} 1 + \exp(\beta_i + \beta_j)}.$$

We have seen that the parameters in both of these models can be consistently estimated. However, this starts to become more of a challenge when $S(g)$ involves subgraphs with three or more nodes, such as triangles.

[12]Clearly, this formulation can depend on covariates, as well. So one can write

$$\mathrm{P}_\beta(g|x) = \frac{\exp(\beta \cdot S_x(g))}{\sum_{g'} \exp(\beta \cdot S_x(g'))},$$

where the vector of attributes can depend on x.

4.1.3. *Estimation Problems.* Let us assume the researcher believes that her data is best modeled by a distribution with a given known S. Generally, outside of models that reduce to link independence, there are two main problems with ERGMs in practice: (i) feasibility of estimation; and (ii) consistency of estimated parameter.

First, it can be very difficult to actually compute parameter estimates. The reason is the following. There are many possible networks, in fact $2^{\binom{n}{2}}$ undirected graphs. Notice that estimating the likelihood of a given network requires having an estimate of whether it is more or less likely than other possible networks. This means that either implicitly or explicitly, one needs an estimate of the denominator in (4.2). In simple models where links are conditionally independent, this is effectively trivial to do. But in the general case, this can be difficult.

Of course, computing this directly is entirely impossible. Therefore, the literature has moved to employing Markov Chain Monte Carlo (MCMC) sampling techniques (Snijders, 2002; Handcock, 2003). The basic idea is to construct a chain of networks that represents draws from the ERGM distribution. However, this approach also has its limitations. The developers of this method, as well as its practitioners, have long noted that there were degeneracy and convergence problems for many common specifications of ERGMs (usually those with links and triangles at a minimum).

Recent work by Bhamidi, Bresler, and Sly (2008) and Chatterjee and Diaconis (2013) has clarified exactly why these problems happen in the context of dense networks.[13] The MCMC procedure can only visit an infinitesimally small portion of the set of possible graphs. The crucial question, therefore, is whether a chain would actually mix in polynomial time. These papers have shown that for broad classes of ERGMs, standard MCMC procedures will take exponential time to mix *unless the links in the network are approximately independent*. But if links are approximately independent then there is no real need for an ERGM formulation to begin with! There are more direct ways to represent and estimate such a model. It seems, then, that in cases where ERGMs are really needed they cannot be accurately estimated by such MCMC techniques.

The second problem is that of the consistency of parameter estimates. If we were indeed able to compute the maximum likelihood estimate, is it the case that along an asymptotic sequence of networks as $n \to \infty$ estimates become closer and closer to the truth? This is a crucial property for any frequentist estimator to have, especially given that in most data settings, the researcher only has access to a single network or a handful of networks. It may be the case, in fact, that having more nodes may not even increase information in the system. Shalizi and Rinaldo (2012) explore these issues. They show that for any exponential family distribution, projectivity is a sufficient condition for consistency. Then they show that any ERGM with a sufficient statistic that involves a triangle violates this condition.[14]

[13]For simulation-based evidence in the sparse case, see Chandrasekhar and Jackson (2014).

[14]Chandrasekhar and Jackson (2014) approach this problem in a different way. They note that the reference distribution imposed by ERGMs is what generates the problem and point out that related classes of models with differing reference distributions can allow for consistency.

While ERGMs that have higher order link dependencies are easily interpretable in a simple random utility framework, they lead to models that are difficult, if not impossible, to estimate and parameter estimates that are not necessarily consistent as $n \to \infty$. However, the aspiration of the microeconometric exercise is simple: preferences should incorporate payoffs from not simply friends but other links, which may depend upon local network structure. The next approach deals with modeling subgraphs directly.

4.2. **Subgraph Generated Models (SUGMs).** One objective of the ERGM program was to model the correlation in links through probability models where the sufficient statistics were counts of subgraphs.

We now look at a related, but more direct approach: build networks up from subgraphs as the fundamental units. This is the path taken in Bollobás, Janson, and Riordan (2011) and Chandrasekhar and Jackson (2015). A network is constructed from building up subgraphs of various types: links, triangles, other cliques, stars, etc. These structures are layered upon each other yielding the network of interest. Figure 6 depicts the idea.

For exposition, let us consider the special case where networks are comprised of incentives to form links and incentives to form triangles. Define an *unsupported link* to be a link between two nodes that forms directly and not as part of larger subgraph, and the set of such links that form $\mathcal{LU}$. As before, a triangle is a set of three nodes that form together as a group, and the set of such triangles that form is denoted $\mathcal{T}$.

Node i forms a collection of unsupported links and triangles. These can intersect and overlap, and the resulting network that forms is given by

$$g = \{ij : ij \in \mathcal{LU} \text{ and/or } \exists k \text{ s.t. } ijk \in \mathcal{T}\}.$$

Two agents are connected if they are part of a link, or both are part of any triangle. Thus, g can be thought of as a projection of $\mathcal{LU}$ and $\mathcal{T}$.

If the researcher observed the sets $\mathcal{LU}$ and $\mathcal{T}$, then estimation would be straightforward under standard regularity conditions.[15] Often researchers only have binary information concerning whether each two individuals in the society interact. For instance, we may have data indicating which pairs of agents are "friends" based on a survey, from observing that they are linked on social media, etc.

From this perspective, the 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 $\mathcal{LU}$ and $\mathcal{T}$. If

$$A_{ij}A_{jk}A_{ik} = 1,$$

is it the case that ijk formed as a triangle; ij, jk, and ik formed as links; ij and jk formed as links and ik formed as part of a different triangle ikm; etc?

[15] In some instances, a researcher might have information on all the various groups a given individual is involved in: for example, in the case of a co-authorship network, the researcher may observe all the papers a researcher has written.

In sum, if the probability of links and triangles forming are given by some parameter vector, the researcher needs to use the observed graph g to infer how different subgraphs formed in order to estimate the parameter vector.

4.2.1. *Utility foundations.* Before describing the statistical model and the estimation methods, it is useful to look at two (distinct) random utility foundations for subgraph generated models. In the first model, agents meet other agents in pairs, triples, or other subsets and then choose, via mutual consent, whether or not to establish links. The resulting vector of relationships that have formed comprise the observed graph g.

In the second model, agents are interested in investing in partnerships for production. Productive activities can be undertaken in pairs, in groups of three, etc. Agents engage in costly search to find partners and other groups for production, and the equilibrium of this search process generates the observed graph g.

EXAMPLE **4** (Meetings)**.** *The utility an agent i receives from being in a collection of unsupported links and a collection of triangles is given by*

$$u_i(\mathcal{LU}, \mathcal{T}, x) = \sum_{\{j:\ ij \in \mathcal{LU}\}} u_{i,L}(x_i, x_j) + \sum_{\{j<k:\ ijk \in \mathcal{T}\}} u_{i,T}(x_i, x_j, x_k).$$

The utility of an unsupported link can depend on

- *a selectivity factor: $\beta_{L,0}$;*
- *factors depending on both parties' characteristics x_i, x_j: $h_L(x_i, x_j)$ (for example, $h_L(x_i, x_j) = |x_i - x_j|$); and*
- *an idiosyncratic component: $\epsilon_{i,ij}$.*

The utility to i from forming an unsupported link ij is given by

$$u_{i,L}(x_i, x_j) = \beta_{L,0} + \beta_{L,1} h_L(x_i, x_j) - \epsilon_{i,ij}.$$

An agent consents to being in ij if and only if $u_{i,L}(ij, x) \geq 0$. The unsupported link forms if both agents consent:[16]

$$u_{i,L}(x_i, x_j) \geq 0 \text{ and } u_{j,L}(x_j, x_i) \geq 0.$$

Similarly, the utility of being in a trilateral relationship – a triangle – can depend on

- *a* selectivity factor*: $\beta_{T,0}$;*
- *factors depending on the parties' characteristics: $h_T(x_i, x_j, x_k)$, (for example, $h_T(x_i, x_j, x_k) = \sum_{i<j} |x_i - x_j| / 3$); and*
- *an idiosyncratic component: $\epsilon_{i,ijk}$.*

The utility to i of being in a triangle $ijk = 1$ is given by

$$u_{i,T}(x_i, x_j, x_k) = \beta_{T,0} + \beta_{T,1} h_T(x_i, x_j, x_k) - \epsilon_{i,ijk}.$$

An agent consents to forming a triangle ijk if and only if $u_{i,T}(x_i, x_j, x_k) \geq 0$. The triangle forms if all agents consent:

$$u_{i,T}(x_i, x_j, x_k) \geq 0,\ u_{j,T}(x_j, x_i, x_k) \geq 0,\ \text{and } u_{k,T}(x_k, x_i, x_j) \geq 0.$$

[16]Notice that as $\beta_{L,0}$ decreases, the agent becomes more selective, since holding the distribution of shocks and covariates fixed, the agent requires a higher draw of the idiosyncratic component to be happy with a link.

In sum, the utility of being in a society with the set $\mathcal{LU}$ unsupported links and the set of $\mathcal{T}$ triangles is

$$\begin{aligned} u_i(\mathcal{LU}, \mathcal{T}, x) &= \sum_{\{j:\ ij \in \mathcal{LU}\}} u_L(x_i, x_j) + \sum_{\{j<k:\ ijk \in \mathcal{T}\}} u_{i,T}(x_i, x_j, x_k) \\ &= \sum_{\{j:\ ij \in \mathcal{LU}\}} (\beta_{L,0} + \beta_{L,1} h_L(x_i, x_j) - \epsilon_{i,ij}) \\ &\quad + \sum_{\{j<k:\ ijk \in \mathcal{T}\}} (\beta_{T,0} + \beta_{T,1} h_T(x_i, x_j, x_k) - \epsilon_{i,ijk}) \,. \end{aligned}$$

Individuals may not have the opportunity to meet all possible subgroups, so some subgraphs may be impossible to build. Let $\pi_L(x_i, x_j)$ denote the probability that two individuals i and j meet, and let $\pi_T(x_i, x_j, x_k)$ denote the probability that three individuals meet. Let F_L, F_T be the cumulative distribution functions for $\epsilon_{L,ij}, \epsilon_{T,ijk}$. Under independence

$$\begin{aligned} \mathrm{P}\left(ij \text{ forms as a link} | x_i, x_j\right) &= \mathrm{P}\left(ij \text{ meet} | x_i, x_j\right) \cdot \mathrm{P}\left(ij \text{ consent} |\ ij \text{ meet}; x_i, x_j\right) \\ &= \pi_L(x_i, x_j) \cdot F_L\left(\beta_{L,0} + \beta_{L,1} h_L(x_i, x_j)\right)^2 . \end{aligned}$$

and

$$\begin{aligned} \mathrm{P}\left(ijk \text{ forms as a triangle} | x_i, x_j, x_k\right) &= \mathrm{P}\left(ijk \text{ meet} | x_i, x_j, x_k\right) \cdot \mathrm{P}\left(ijk \text{ consent} |\ ijk \text{ meet}; x_i, x_j, x_k\right) \\ &= \pi_T(x_i, x_j, x_k) \cdot F_T\left(\beta_{T,0} + \beta_{T,1} h_T(x_i, x_j, x_k)\right)^3 . \end{aligned}$$

Therefore, under appropriate assumptions on functional forms, the frequency with which different subgraphs appear and how this frequency varies with member nodes' attributes, identifies the parameters of interest $(\beta_{L,0}, \beta_{L,1}, \beta_{T,0}, \beta_{T,1})$.

While this example discussed models that implicitly or explicitly involved linking costs for various subgraphs in isolation, our next approach takes on an entirely different perspective in which there is an overall cost of forming relationships. This generates a tradeoff, since individuals now need to optimally apply effort to maintain the right sort of structural relationships. The theory literature has studied models with search intensities (Currarini, Jackson, and Pin, 2009, 2010; Borgs, Chayes, Ding, and Lucier, 2010; Golub and Livne, 2010), and the next model develops an econometric analog.

EXAMPLE **5** (Search intensity)**.** *An agent with attributes x_i can put in search effort to form a clique with m nodes, Cl_m, consisting of characteristic vector x. From this match, an agent obtains a utility*

$$u(x_i, m, x).$$

For simplicity, assume $u(x_i, m, x) = u_m$. Let

$$e(x_i, m, x) \in [0, \bar{e}_m]$$

denote the search effort. For now assume $e(x_i, m, x) = e_m$.

Whether or not the clique Cl_m forms is stochastic. This probability depends on the vector of efforts of potential members,

$$p(e_1, .., e_m) = p(e_1(x_1, m, x), ..., e_m(x_m, m, x)).$$

We can assume that the probability is non-decreasing in each argument. For now set

$$p(e_1, .., e_m) = \sum_{j \in [1,...,m]} e_j$$

where by assumption $\bar{e}_m$ *is such that* $m\bar{e} < 1$.[17]

The cost of effort is $c\left((e_i(x_i, m, x))_{m,x}, x_i\right)$*, which can depend on her characteristics and her search effort for all cliques of all attribute types. Here we have*

$$c\left((e_i(x_i, m, x))_{m,x}, x_i\right) = \sum_m \phi_m \frac{e_{i,m}^2}{2}.$$

The expected utility is

$$\sum_{m, Cl_m : i \in Cl_m} u(x_i, m, x) \cdot p((e_j(x_j, m, x))_{j \in Cl_m}) - c\left((e_i(x_i, m, x))_{m,x}, x_i\right).$$

Under our simplifying assumptions, expected utility is

$$\sum_{m, Cl_m : i \in Cl_m} u_m \cdot \left(\sum_{j \in [m]} e_{j,m} \right) - \sum_m \phi_m \frac{e_{i,m}^2}{2}.$$

It follows that $e_{i,m}^* = \frac{u_m}{\phi_m}$*. Normalizing utility by the cost parameter to define* $\beta_m := \frac{u_m}{\phi_m}$ *and observing that the frequency of a clique forming is given by*

$$p_m = p(e_1^*, ..., e_m^*) = m e_{i,m}^* = m\beta_m,$$

we see how to map from the frequency of a subgraph to the structural parameter.

Both examples presented here map two different interactions and random utility frameworks to a network wherein structural parameters are identified by the rate at which various subgraphs are present in the graph. We now turn to the problem of estimating subgraph frequency parameters which, under assumptions like those above, would allow a researcher to estimate utility parameters as well.

4.2.2. *Subgraph Generation.* In SUGMs, subgraphs are directly generated by some process. Classic examples of this are Erdős-Rényi random graphs, where links are randomly generated, and the generalization of that model, stochastic-block models, in which links are formed with probabilities based on the nodes' attributes. The generalization to SUGMs is to allow richer subgraphs to form directly, and hence to allow for dependencies in link formation. 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.

Formally, there is a a finite number of different types of nonempty subgraphs, indexed by $\ell \in \{1, \ldots, k\}$, on which the model is based. Then a SUGM on n nodes is based on some list of k subgraph types: $(G_\ell)_{\ell \in \{1,\ldots,k\}}$. Each G_ℓ is a set of possible subgraphs on m_ℓ nodes, which are identical to each other up to the relabeling of nodes. Two examples for G_ℓ are (i) all triangles, and (ii) all stars, with one central node and four other nodes. The final ingredient is a parameter vector $\beta \in \mathcal{B}$, where $\mathcal{B}$ is a subset of some finite dimensional Euclidean space.

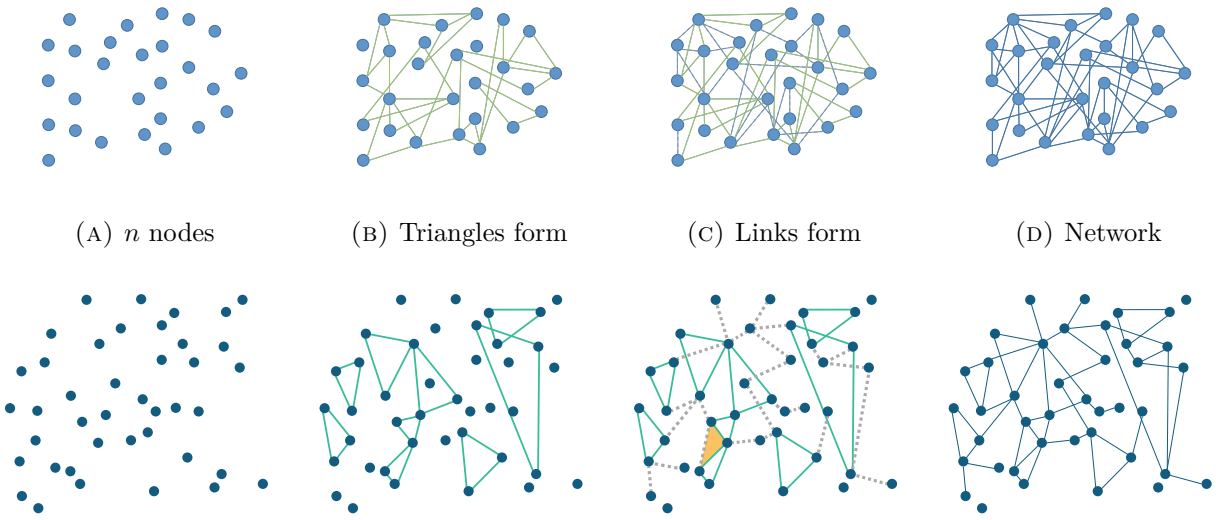
[17] This means that no agents can put in enough effort to guarantee a successful meeting.

A network g is randomly formed as follows. Each of the possible subnetworks in G_ℓ is independently formed with probability

$$p_\ell(x(g_\ell), \beta_\ell)$$

for each $\ell \in \{1, \ldots, k\}$, where $x(g_\ell)$ is a set of covariates for the nodes in the candidate subgraph g_ℓ. The resulting network, g, is the union of all the links that appear in any of the generated subgraphs.

Figure 6 describes the process for a model with (p_L, p_T). Each of the $\binom{n}{3}$ potential triangles form with probability p_T, and each of the potential unsupported links form with probability p_L. Sometimes they overlap; in this case, an extra triangle forms relative to those that were truly generated, therefore eliminating an unsupported link.



(A) n nodes (B) Triangles form (C) Links form (D) Network

(E) n nodes (F) Triangles form (G) Links form (H) Network

FIGURE 6. A figure depicting a SUGM process taken from Chandrasekhar and Jackson (2015). (A)-(D) shows the dense case and (E)-(H) shows the sparse case.

4.2.3. *Estimation.* Before we describe estimation, it is useful to clarify the asymptotic thought experiment. As noted in Section 3, since the average degree (and other features) may not grow linearly in n and could even grow slowly in n, we must allow the linking probability to vary in n.

For simplicity, consider a sequence of SUGMs without covariates, where

$$p_\ell^n = F_\ell(\beta_\ell^n),$$

where $F_\ell(\cdot)$ is the uniform distribution on [0,1], β_ℓ^n is a parameter where subgraph ℓ forms if the random variable is less than β_ℓ^n. Typically $\beta_\ell^n \downarrow 0$ at appropriate rates.

Given g, the researcher seeks to estimate β_ℓ^n (equivalently p_ℓ^n). The quality of the estimates, whether the estimator is identified, consistent, and (when normalized appropriately) asymptotically normally distributed, depends on the assumptions on the sequence $\beta^n = (\beta_1^n, ..., \beta_k^n)'$. We have already seen such a sequence and the asymptotic normality of the associated estimator when we studied the sparse Erdős-Rényi graph in equation (3.1). What we do now represents a generalization.

Incidental generation. What is the main difficulty that we face? As we can see in Figure 6, even though 20 triangles and 20 unsupported links form in the dense case, and though 9 triangles and 23 unsupported links form in the sparse case, we do not observe these sets $\mathcal{LU}$ and $\mathcal{T}$. Instead, we only observe the resulting graph g, which has 30 triangles and 4 unsupported links in the dense case and 10 triangles and 22 unsupported links in the sparse case. The generation of the extra triangles is called incidental generation.

There are two approaches to dealing with this problem. First, if the networks are sparse enough, then we can see that the error will be vanishing relative to the truly generated part, meaning that by directly counting subgraphs, one can recover the probability parameters of interest. Second, even if the networks are not sparse enough, because the rate of incidental generation varies as β varies, one can show how β is identified. We look at the general case first and then the sparse case.

General case. It is useful to define

$$S_\ell(g) := \frac{\sum_{i_1<...<i_{m_\ell}} G_{i_1,...,i_{m_\ell}}}{\binom{n}{m_\ell}}$$

as the share of subgraphs of type ℓ that exist in the graph. The vector $S(g) = (S_1(g), ..., S_k(g))'$ represents a set of moments that we will use in estimation.

When we consider our example of links and triangles, let

$$(S_L(g), S_T(g)) = \left(\sum_{i<j} A_{ij} / \binom{n}{2}, \sum_{i<j<k} A_{ij}A_{jk}A_{ik} / \binom{n}{3} \right)$$

be the fraction of links (whether isolated or in a triangle) and the fraction of possible triangles in the network g.

The researcher then estimates the minimizer of the GMM objective function.

$$\widehat{\beta} := \underset{\beta \in \mathcal{B}}{\operatorname{argmin}} \left[S(g) - \mathrm{E}_\beta S(g) \right]' W^n \left[S(g) - \mathrm{E}_\beta S(g) \right].$$

Under assumptions that allow for identification and limit the overall amount of correlation in the network, the parameters can be consistently estimated[18]

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

The estimators, when appropriately normalized, are asymptotically normally distributed

$$D_n^{1/2}(\widehat{\beta} - \beta^n) \rightsquigarrow \mathcal{N}(0, I)$$

[18] $\delta(x, y) := \sum_k \frac{|x_k - y_k|}{\max(x_k, y_k)}$ is a distance metric

for some normalizing matrix sequence (D_n).

Let us highlight briefly the logic of identification and the asymptotic normality. The identification problem potentially comes from incidental generation: when one sees a triangle present, it could have been generated as a triangle, or as a mix of other triangles and links. The question is whether it is possible for a given vector of links and triangles to have two solutions (β_L, β_T) and (β'_L, β'_T).

The main identification result for the links and triangles case is that if the unconditional probability of links is less than 1/2, then $(\beta_L, \beta_T) \neq (\beta'_L, \beta'_T)$ implies

$$\mathrm{E}_{\beta_L,\beta_T}\left[S_L(g), S_T(g)\right] \neq \mathrm{E}_{\beta'_L,\beta'_T}\left[S_L(g), S_T(g)\right].$$

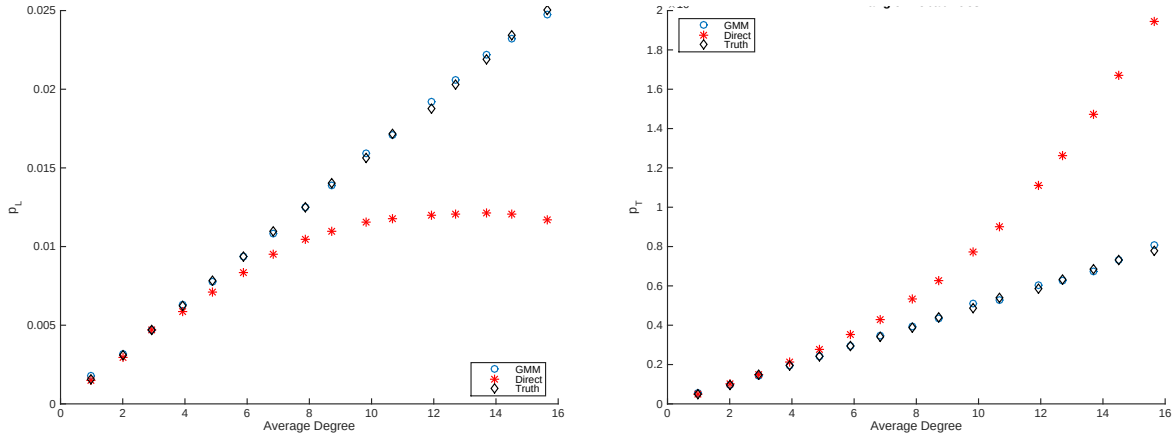
This says that under the assumption there is an injective map from the parameter to the moments, and therefore the parameters are identified.

To see this, consider the expectation of these moments where $q = q(\beta_L, \beta_T)$ is the unconditional probability of a link forming. We have

$$\mathrm{E}_{\beta_L,\beta_T}\left[S_L(g), S_T(g)\right] = \left[\beta_L + (1-\beta_L)(1-(1-\beta_T)^{n-2}), \beta_T + (1-\beta_T)(q-\beta_T)^3\right].$$

A link can be formed directly (with probability β_L) or by some triangle involving the two nodes in question (leaving $n-2$ choices for the third node). Meanwhile, a triangle forms directly (with probability β_T) or where the three links are formed by some combination of unsupported links and other triangles ($(q-\beta_T)^3$). This gives us two non-linear equations and two unknowns with a unique solution under our assumption.

The intuition for identification is as follows. While the counts of triangles and links are being used, the key observation is that how triangles are glued together – whether they share nodes or share edges – tells us about what one parameter should be, holding fixed a budget of links. In that way, the arrangement of subgraphs relative to each other in a graph can give the researcher information about the degree of incidental generation versus direct generation.



(A) Estimating p_L (B) Estimating p_T

FIGURE 7. Parameter estimates as a function of density.

Figure 7 provides simulation results to demonstrate how as sparsity declines, the direct estimation method is less suitable whereas the GMM approach works reliably.

Next, let us consider the logic as to why the parameter estimates are normally distributed. Standard arguments involve expanding the first order condition around $(\widehat{\beta} - \beta^n)$. The key piece in the normality argument comes from showing that

$$(S_\ell - \mathrm{E}S_\ell)/\sigma_{n,\ell} \rightsquigarrow \mathcal{N}(0,1)$$

for $\sigma^2_{n,\ell} := \mathrm{var}(S_\ell)$.

The reason why asymptotic normality is not obvious here is because the summands in question can be correlated. It is useful to observe that in any SUGM that has a subgraph involving at least four nodes, any two links are always correlated. Consider the model of links, triangles, and four cliques (i, j, k, l all mutually connected). Then any g_{ij} and g_{kl} are surely correlated. Further, for any collection of $ijkl$ and $mnop$ of nodes, even if they have no nodes in common, the events $\{g_{ij}g_{ik}g_{il}g_{jk}g_{jl}g_{kl} = 1\}$ and $\{g_{mn}g_{mo}g_{mp}g_{no}g_{np}g_{op} = 1\}$ are correlated. In this case

$$(S_\ell - \mathrm{E}S_\ell)/\sigma_{n,\ell} = \sigma^{-1}_{n,\ell} \sum_{i<j<k<l} G_{i,j,k,l} - \mathrm{E}[G_{i,j,k,l}]$$

where summands $G_{i,j,k,l}$ and $G_{m,n,o,p}$ are correlated irrespective of the identities of the nodes in question.

On the other hand, the covariance between any two links is low. Moreover, when we look at subgraphs of four nodes, for instance, when there are few nodes in common between $ijkl$ and $mnop$, the correlation in both subgraphs existing comes from a somewhat unlikely event: a four clique between $ijmn$ and then some subgraphs forming to fill out the four-cliques $ijkl$ and $mnop$.

If elements in an array can all be correlated with each other, if most of the correlation can be restricted to a small subset of the array, then Chandrasekhar and Jackson (2015) show that a central limit theorem applies. In the example described above there are many fewer four cliques that share nodes than those that are node-independent, allowing such a theorem to be applied. The result uses a key lemma in Stein (1986) that characterizes the distance of a random variable Y from a normal distribution (Z) in the Wasserstein metric by

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

Therefore, the goal is to show that the right-hand side tends to zero for $Y := (S_\ell - \mathrm{E}S_\ell)/\sigma_n$. This strategy has been used widely, for instance by Bolthausen (1982), to prove central limit theorems for strongly mixing random fields, as well as by Baldi and Rinott (1989) to look at data that can be described by dependency graphs. In the latter example, every observation can be correlated with some (arbitrary) k other observations but is independent of the rest. Ross et al. (2011) provides an excellent survey of the techniques. The SUGM case is somewhat more complicated, because (i) in principle, all summands can be correlated with all other summands (therefore k can equal the number of observations), making one unable to rely on independence in the argument; and (ii) unlike in the time series or spatial case, one cannot rely

on the idea that the summands can be embedded in some Euclidean space so that asymptotically each summand is infinitely far from most other summands. Ultimately, the normality result applies for SUGMs that admit considerable density as well as sparsity. For instance, in a links and triangles model, a node can have order $O(n^{1-\epsilon})$ expected degree for $\epsilon \in (0, 1]$.

Sparse case. While in the general treatment, we typically had $\beta_\ell^n \downarrow$ for parameters that did not correspond to direct links, the rate at which this was required was not particularly stringent. In particular, it allowed for a tremendous amount of incidental generation, and in fact, allowed for dense graphs (e.g., so long as less than half the links were present in expectation in the links and triangles model).

As we discussed previously, empirical network data is rather sparse. It turns out that in this case, our problem becomes much simpler. We can compute that it is exponentially more likely to have a directly formed subgraph than one formed as a combination (incidentally generated). In turn, this implies that the measurement error problem is small, so direct counting / regression approaches will work.

To make the required convergence rates concrete, in the links and triangles model we need

$$p_L = o(n^{-1/2}) \text{ and } p_T = o(n^{-3/2}).$$

This means that every node can participate in $o(\sqrt{n})$ unsupported links and $o(\sqrt{n})$ triangles, which as we know from the stylized facts, is more than sufficient to model empirical data.

To see how this would be estimated in practice, let us stay with the links and triangles SUGM with covariates where the probability of a triangle or a link can depend on covariates of the nodes:

$$p_T^n(x_T; \beta_T) = \frac{\exp(\beta_{0,T}^n + \beta_{1,T}\left(|x_i - x_j| + |x_j - x_k| + |x_i - x_k|\right))}{1 + \exp(\beta_{0,T}^n + \beta_{1,T}\left(|x_i - x_j| + |x_j - x_k| + |x_i - x_k|\right))}, \text{ and}$$

$$p_L^n(x_L; \beta_L) = \frac{\exp(\beta_{0,L}^n + \beta_{1,L}\left(|x_i - x_j|\right))}{1 + \exp(\beta_{0,L}^n + \beta_{1,L}\left(|x_i - x_j|\right))}.$$

Under an assumption that preserves sparsity – for instance the constants $\beta_{0,\ell}^n \to -\infty$ at appropriate rates – the parameters β_ℓ can be estimated consistently with estimators that are asymptotically normally distributed. Further, the estimation procedure is rather direct. In the links and triangles example:

(1) Consider all ijk, and conduct a logistic regression of whether a triangle exists on covariates of the triple. The coefficients are $\widehat{\beta}_T$.
(2) Eliminate all nodes ij where ij was a member of some triangle. With the reduced dataset, conduct a logistic regression of whether an edge exists on covariates of the pair. The coefficients are $\widehat{\beta}_L$.

4.3. **Sequential Arrival and Preferential Attachment.**

A commonly studied class of random network models is *preferential attachment*. First introduced by Barabasi and Albert (1999), the model proceeds as follows.

Given a seed graph, in every period, some nodes are born. These nodes link with some probability to previously existing nodes with probabilities proportional to the previously existing nodes' degrees. This generates a rich-get-richer phenomenon, where more popular nodes differentially become more likely to gain links. By tuning the number of nodes born each period, the number of edges born each period, etc., one can vary properties of the resulting distribution of graphs.

The main result of the literature is the ability to generate a power law tailed degree distribution. This is in contrast with (sparse) Erdős-Rényi graphs, which generate Poisson (thin) tails.

There are several difficulties in employing these models. Until recently, we knew little about whether parameter estimates were consistent. Nonetheless practitioners made use of either a non-linear least squares estimator an estimator by Hill (1975). A key contribution to this literature was made in Kolotilin (2013) and fills this gap. The paper shows that for a generalized class of preferential attachment models, the formation parameters can consistently estimated by a certain class of GMM estimators. At the same time the non-linear least squares and Hill estimators exhibit severe biases. However, the question as to whether these consistent GMM-based estimates are asymptotically normally distributed remains open. Kolotilin (2013) provides some simulations suggesting this may be the case, but there are no formal results in the literature.

In a seminal paper, Christakis et al. (2010) developed a sequential arrival model in this vein. In every period one pair of agents have the opportunity to form a link and conditional on the opportunity the link is formed if both agents consent. As in the aforementioned models (e.g., Butts (2009), Mele (2013), and Badev (2013)), agents are myopic and therefore engage in myopic best response.

A simplified version of the utility function they use is given by

$$u_i(g^{t-1} + ij) = \beta_0 + \beta_1 x_j - \beta_2 |x_i - x_j| + \beta_d d_j(g^{t-1}) + \beta_T \mathbf{1}\{\gamma_{g^{t-1}}(i,j) = 2\} + \epsilon_{ij}.$$

To make notation transparent $d_j(g^{t-1})$ is the degree in graph g^{t-1} of j whereas $\gamma_{g^{t-1}}(i,j)$ is the distance in the graph g^{t-1} from nodes i to j. Thus, individuals receive utility from being linked to more popular individuals and from completing a triangle.

A key difference in this model relative to those that led to stationary distributions (e.g., Butts (2009); Mele (2013); Badev (2013)) is that their posited process is a one-time match process. The $\binom{n}{2}$ pairs meet in some unobserved sequence and therefore the econometrician has to integrate this sequence out. They use Bayesian MCMC methods to obtain draws from the posterior distribution of interest.

A related sequential arrival model in the literature is Koenig (2012). This paper studies a preferential attachment style network formation model. Agents arrive in a sequence and form links with pre-existing nodes. The agents have utilities that correspond, essentially, to having incentive to link to the highest degree partner. In that way the model closely resembles the Barabasi and Albert (1999) model discussed before. The crucial departure from this literature is that he tinkers with the observation radius, meaning how much a newborn node i can see when making her decision. There are two regimes he considers. First, the agent at time t may see a large portion of the

pre-existing network (essentially all its predecessors and their links, similar to the usual Barabasi and Albert (1999) model). Second, the agent may only see a selected subset of nodes, their degrees and their neighbors' degrees. By assumption the agent's payoff depends on her own degree and her neighbors' degrees and as such, and chooses to link with the highest degree nodes in her observation radius. The paper focuses on (i) how the observation radius affects the patterns of networks that emerge through this process and (ii) how the network patterns also vary with how much neighbors' degrees affects one's incentive to link with them (*noise* in linking decision). The author is able to generate patterns that both fit empirical data better than standard preferential attachment models which usually assume that a node born in period t sees the entire pre-existing network.

As the literature currently stand, neither of the models described here are known to have estimators that are consistent for large n. Results such as these are open questions.

4.4. **Strong Homophily.** There is a set of models in the literature that exploits a strong form of homophily to generate consistently estimable parametric utility functions. Instead of assuming additive separability over subgraph payoffs, they place less restriction on the local payoff structure.

However, the price paid for releasing the restrictiveness here is (i) more structure on which nodes are allowed, under the model, to link to which other nodes, (ii) intensive computational requirements, and (iii) partial identification. This approach has been taken in Goldsmith-Pinkham and Imbens (2013), Boucher and Mourifié (2012), and Leung (2014). Below, we review a simplified version of these models, using the notation closest to Leung (2014).

4.4.1. *Overview.* In this model, nodes have deterministic locations, $l_i \in \mathbb{R}^d$. Additionally, they have random characteristics x_i. Finally, there are pairwise shocks ϵ_{ij}. Let $w = \{(l_i, x_i, \epsilon_{ij})\}_{i,j \in V}$. As before, a general parametric model to write the utility of an equilibrium graph is

$$u_i(g, w; \beta).$$

As in previous sections, the literature focuses on pairwise stability. Define

$$U_{ij}^n(g, w, \beta) = u_i(g + ij, w, \beta) - u_i(g - ij, w, \beta).$$

This encodes the marginal value at the equilibrium graph of having link ij.

4.4.2. *Estimation requirements.* In order to identify β, the econometrician will have some moment functions, $\psi_i^n(\beta)$, derived from the game. A crucial step, therefore, in conducting asymptotic inference is establishing a central limit theorem for the moments:

$$\frac{1}{\sqrt{n}} \sum_i \psi_i^n(\beta) \rightsquigarrow \mathcal{N}(0, \Sigma).$$

The main result is that if we assume that i and j are located far from each other, it is almost never worthwhile to form a direct link. Thus, one can show that the system $\{\psi_i^n : i \in V_n, n \in \mathbb{N}\}$ is an α-mixing random field.

To see what this means, consider two sets of nodes $V_A, V_B \subset V_n$. Let E_A and E_B be events in the σ-algebras generated by $\{\psi_i : i \in V_A\}$ and $\{\psi_i : i \in V_B\}$. Then

$$|\mathrm{P}(E_A \cap E_B) - \mathrm{P}(E_A)\mathrm{P}(E_B)|$$

looks at the probability of both events occurring versus the product of the probabilities that each occurs. If the events were independent, this is zero. Loosely, this is a measure of the degree of interdependence between random variables on V_A and V_B.

The α-mixing coefficient measures this. Formally,

$$\alpha_{k,l}(r) := \sup_{n \geq 1} \sup\{|\mathrm{P}(E_A \cap E_B) - \mathrm{P}(E_A)\mathrm{P}(E_B)| : E_j \in \mathcal{F}_j, j \in \{A, B\}, |V_A| \leq k, |V_B| \leq l, d(V_A, V_B) \geq r\}$$

where $d(V_A, V_B)$ is the Hausdorff distance in l-space between the sets. If we look at all events taking place among nodes in V_A and V_B, with sizes k and l, if the sets are least r-distance apart, then the degree of interdependence over all events is at most $\alpha_{k,l}(r)$. A sufficient condition for a central limit theorem has this coefficient going to zero sufficiently fast as $n \to \infty$.

4.4.3. *Driving assumptions.* The next step is to make assumptions on preferences such that the resulting distribution over links generated by strategic interaction will satisfy some mixing properties.

Assume that nodes arrive in communities. $C(i)$ is a set of nodes that are at most r_0 apart from each other. The distance between two nodes in "location" space, an admittedly clumsy term, is

$$d(i, j) := \|l_i - l_j\|.$$

Additionally, parametrize the marginal utility of having link ij as

$$U_{ij}(g, w; \beta) := U_{ij}(\underbrace{d(i,j)}_{\text{distance}}, \underbrace{z(g, x, \epsilon)}_{\text{endogenous}}, \underbrace{f(x, \epsilon)}_{\text{exogenous}}; \beta).$$

Note that $z = z(g, x, \epsilon)$ is a function that allows for endogenous effects of the network since it takes g as an argument. $f(x, \epsilon)$ encodes the direct effect of exogenous factors. In principle, endogenous effects allows for arbitrary correlation between links. To make progress, four key assumptions are made.[19]

A1. *Strong homophily:* Irrespective of the endogenous network effects on utility, it is typically not worthwhile to directly link to someone very far away:

$$\lim_{d(i,j) \to \infty} U_{ij}(d(i, j), \bar{z}, f(x, \epsilon); \bar{\beta}) < 0.$$

for $f(x, \epsilon) \leq \bar{f}$, and the probability of $f(x, \epsilon) > \bar{f}$ is sufficiently low.

If $f(x, \epsilon)$ is bounded, then A1. makes it impossible for certain nodes to link. Also observe that it is necessary to bound the endogenous effect. So very far nodes only link due to purely exogenous factors, and this is rare.

[19]For simplicity, let z and β be unidimensional and bounded by $\bar{z}, \bar{\beta}$ and let the marginal utility increase in these parameters.

A2. *Increasing domain:* All but finitely many nodes are within some minimum radius r_0 of any given node i:

$$\sup_{n\geq 1}\max_{i\in V_n}|j\in V_n:\ d(i,j)<r_0|<\infty.$$

This means that communities are small relative to the overall network.

A3. *Independent equilibrium selection/no coordination:*[20] For any given draw of the primitives x,ϵ, sets of nodes that aren't connected under any equilibrium independently form subnetworks.

Note that this is satisfied if equilbrium selection is degenerate. It is also satisfied under myopic best response dynamics.

4.4.4. *Main results.* Let us sketch the proof of α-mixing from A1-A3. Consider an event

$$E:=\{i\leftrightarrow j, d(C(i),C(j))=r\},$$

that i and j are path connected in the graph when they are members of groups that satisfy $d(C(i),C(j))=r$. We want to look at the probability of events $E_i:=\{\psi_i\in H\}$ and the probability that $E_j=\{\psi_j\in H'\}$,

$$|\mathrm{P}(E_i\cap E_j)-\mathrm{P}(E_i)\mathrm{P}(E_j)|\leq \underbrace{|\mathrm{P}(E_i\cap E_j|E^c)-\mathrm{P}(E_i|E^c)\mathrm{P}(E_j|E^c)|}_{=0 \text{ by } \mathbf{A3}.}\mathrm{P}(E^c)+3\mathrm{P}(E).$$

If i and j are not path-connected, by A3, the events are independent. So we just need to know how likely it is that i and j are path connected in the graph if they are at least r apart.

$\mathrm{P}(E)$ is a complicated event. There are many possibilities across many paths, potentially on the order of $n!$. A strong but sufficient condition provided in Leung (2014) illustrates that if groups of nodes are further and further apart along the sequence – that is, if there is drift – then one can drive this rate to zero. This brings us to the fourth assumption. I provide a simplified version here.

A4. *Diversity:* For every i and r, there exists a set of nodes S containing i such that any $k\in S$ satisfies $d(i,k)<r$ and $d(k,l)\geq 9\log r$ for $l\notin S$.[21]

This provides a sufficient condition to argue that

$$\mathrm{P}(E)\leq\phi(r)\to 0 \text{ as } r\to\infty,$$

which completes the α-mixing argument.

To see why A4. is important, consider bounding the probability that there exists some path between i and j:

$$\mathrm{P}(E)=\mathrm{P}\left(\underbrace{\bigcup_{\ell=1}^{n-2}}_{\#\text{ nodes in path}}\ \bigcup_{k_1,\ldots,k_\ell\in V_n}\left\{A_{ik_1}=A_{k_1,k_2}=\cdots=A_{k_\ell,j}=1\right\}\right). \tag{4.3}$$

[20]Set of rationalizable networks is non-empty and the true data generating process is rationalized by the model and the selection mechanism.

[21]This applies when errors have exponential tails only.

Typically, we would take a union bound, but in this case we would have more than $n!$ terms, which would not be useful at all. A4 allows us to turn a number of these terms off, permitting the union bound.

Assume $d(i,j) = r$. Then $i \leftrightarrow j$ implies that there exists k with $d(i,k) < r$ is directly connected to some node l such that $d(k,l) \geq 9 \log r$. This means that if two nodes r apart are connected, then some pair of nodes on that path must be $9 \log r$ apart. Formally,

$$\mathrm{P}(E) \leq \mathrm{P}\left(\bigcup_{\substack{k:d(i,k)<r, \\ l \in V_n, \\ d(k,l) \geq 9 \log r}} \left\{ A_{kl} = 1 \right\} \right). \tag{4.4}$$

This significantly reduces the number of events in the union. We can then apply the union bound:

$$\mathrm{P}\left(\bigcup_{\substack{k:d(i,k)<r, \\ l \in V_n, \\ d(k,l) \geq 9 \log r}} \left\{ A_{kl} = 1 \right\} \right) \leq (\#k : d(i,k) < r) \max_k \sum_{l=1}^{n} \mathrm{P}(A_{kl} = 1 \,|\, d(k,l) \geq 9 \log r).$$

Intuitively, as the sum ranges over nodes l increasingly distant from k, the summands on the RHS should be decaying to zero by homophily. Furthermore, as $9 \log r \to \infty$, the entire summation decays to zero. Some restrictions ensure that this decay is fast enough to overtake the growth of $(\#k : d(i,k) < r)$ as r increases. Then the RHS can be bounded by some $\phi(r) \to 0$ as $r \to \infty$, thus bounding $\mathrm{P}(i \leftrightarrow j)$.

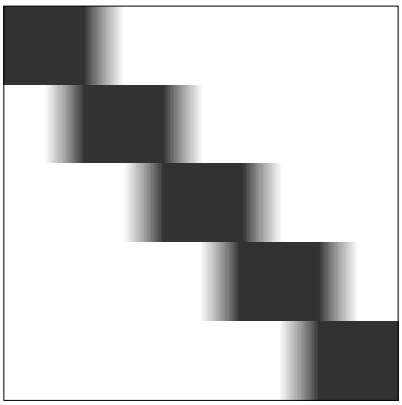


FIGURE 8. A schematic of the adjacency matrix of a network under strong homophily. Darker shades indicate greater likelihoods of link formation.

In sum, exploiting strong homophily allows us to construct models of sparse networks and derive limit theorem by showing mixing properties. While the preferences can be left somewhat unrestricted locally (geographically), allowing for multiple equilibria and motivating partial identification approaches, by construction the interactions have to be localized. Figure 8 presents a schematic of the adjacency matrix implied if nodes were embedded in one dimension. Imagine that nodes were arranged geographically on a line. Nodes are more likely to establish links to their local, geographic, neighbors whereas links between distant nodes are unlikely. While the highly local link patterns are unrestricted, there are strong conditions placed on the off-diagonal terms beyond the main block. Just as in a random geometric graph, distant nodes simply cannot link, so too is the case here. Connectedness and other graphical properties of these models remain open for future research.

4.5. **Network Types.** de Paula, Richards-Shubik, and Tamer (2014) look at identifying preferences in network formation games. As in several preceding models, they assume that agents strategically form links under pairwise stability. The logic of the approach taken here is to consider local network structures, which they call "types," and to show how the relative proportion of various types map to preference parameters. Identification in this framework will be partial because there is a multiplicity of equilibria, so generally, there will be a set of parameters that are consistent with the model and data.

4.5.1. *Driving assumptions.*

A1. *Local payoffs:* Only links up to distance D are payoff relevant, and degree is bounded by L.

This allows control over the local structures that will become payoff relevant. Observe that it imposes a strong sparsity, that the expected average degree is $O(1)$, which may not be a binding restriction in the data (Section 2). On the other hand, this approach by construction will struggle to generate heavy tailed degree distributions.

A2. *Preference shocks:* There is one preference shock $\epsilon_{il}(x)$ for each potential direct connection $l = 1, ..., L$ and characteristic $x \in \mathcal{X}$. Further, $supp(\mathcal{X})$ is finite, and the shock is independent of x and is drawn from a known distribution (up to unknown finite dimensional parameter).

A substantive restriction imposed here is on unobserved heterogeneity. Suppose agents j and k have the same observables. Then an agent i is different between using her lth link on j or k.

While the utility functions studied in the paper are general (required to satisfy A1 and A2), the leading example should be familiar to the reader (e.g., Christakis et al. (2010); Goldsmith-Pinkham and Imbens (2013); Mele (2013)):

$$u_i(g, x) = \sum_j A_{ij} \left(f(x_i, x_j) + \epsilon_{ij}(x_j) \right) + \beta_T \sum_{j<k\neq i} A_{ik} A_{jk} + \beta_{FoF} \left(|\cup_{j \in N_i} N_j| - d_i(g) \right),$$

where $N_i = \{j : A_{ij} = 1\}$. To maintain A1, they impose the restriction that $\max_i d_i(g) \leq L$. Payoffs depend on direct friends and their traits, the number of triangles, and the number of friend-of-friends who aren't direct friends.

A3. *Value of indirect links close to value of closure:* The value of mutual friends is not large relative to the value of friends of friends: $\beta_T \leq \frac{L}{L-1}\beta_{FoF}$ and $\beta_{FoF} > 0$.

This assumption is not used in the general case, but relevant when studying the utility specified above. It imposes that the value of a triad is almost the same as a friend of friend. For instance, if the maximal degree is 11, then the utility from having a triangle can be at most 10% more than the utility of having a friend of friend. How reasonable this assumption is, of course, depends on context. It may be appropriate where closure is not important, and less relevant when closure is key (e.g., favor exchange or risk sharing without commitment).

Two novel concepts introduced in this paper are network types and preference classes. The *network type* describes the local network up to distance D from the reference node. The type is a pair $t = \left(A^{loc}, v\right)$ where A^{loc} is a local square adjacency matrix and v is a vector of characteristics whose length is the size of A^{loc}.

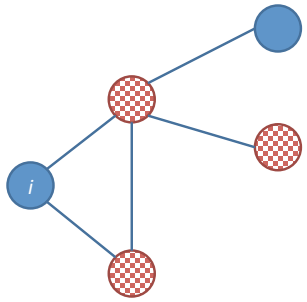


FIGURE 9. An example of a network type with $D = 2$ and $L > 2$. Characteristics are either blue (solid) or red (checkered) and $v_1 = Blue$.

The nodes of A^{loc} consist of the reference node, its direct potential neighbors (L elements), its second order neighbors ($L(L-1)$ elements), through its $D-1^{th}$ order neighbors ($L(L-1)^{D-1}$). Therefore, A^{loc} is a square of size $1 + L\sum_{d=1}^{D}(L-1)^{d-1}$ and v has this length. v_1 encodes i's characteristics, and v_k encodes the characteristics x_k of the k^{th} node in A^{loc}. The set of network types is denoted $\mathcal{T}$.

A *preference class* is a subset $\mathcal{H} \subset \mathcal{T}$ of types defined as follows. Consider mapping characteristics and shocks $(x, \epsilon) \mapsto H(x, \epsilon) \in 2^{\mathcal{T}}$ to all types that satisfy

$$u\left(A^{loc}, v; \epsilon\right) \geq u\left(A^{loc}_{-j}, v; \epsilon\right)$$

for $j = 1, ..., L$. This means that it is the subset of types such that for some characteristic and shock vector, the reference agent wants to keep all her links in the local substructure. Essentially, preference classes are types that can be rationalized in the data for some collection of covariates and shocks.

The next two conditions are necessary for the observable distribution of network types to correspond to a pairwise stable network that could be generated under a given set of preference parameters and shocks. Let $\alpha_H(t)$ denote the share of nodes in preference class H that are of network type t. This is called an *allocation parameter*.

A4. *Existing links shouldn't be dropped:* All existing links are pairwise stable. For any type t and preference class H

$$t \notin H \implies \alpha_H(t) = 0.$$

Next, we define μ_{v_1} the share of nodes that have characteristic $x = v_1$. As all network types in a given preference class by construction have the same characteristic for the reference node, we can define the probability of an ϵ that would support all types in H

$$\mathrm{P}_{H|v_{1(t)}} := \mathrm{P}\left(\epsilon : \ H(x, \epsilon) = H | x = v_1\right).$$

A5. *No mutually beneficial links among far nodes:* For any two type pairs t, s where the nodes are at distance of at least $2D$, where adding a link between the nodes now generates two type pairs $\bar{t}, \bar{s}$:

$$\left(\mu_{v_1(t)} \sum_{\check{H} \in \mathcal{H}} \mathrm{P}_{\check{H}|v_1(t)} \alpha_{\check{H}}(t) \cdot \mathbf{1}\left\{\bar{t} \in \check{H}\right\}\right) \cdot \left(\mu_{v_1(s)} \sum_{\tilde{H} \in \mathcal{H}} \mathrm{P}_{\tilde{H}|v_1(s)} \alpha_{\tilde{H}}(s) \cdot \mathbf{1}\left\{\bar{s} \in \tilde{H}\right\}\right) = 0.$$

4.5.2. *Main results.* The first main result is that given a distribution of preferences in the population, if there exists a pairwise stable network where the proportion of agents of type t is equal to π_t for every $t \in \mathcal{T}$, then there exists a vector of allocation parameters α satisfying assumptions 4 and 5 such that

$$\pi_t = \frac{1}{\mu} \sum_H \mu_{v_1(t)} \mathrm{P}_{H|v_{1(t)}} \alpha_H(t) \text{ for every } t \in \mathcal{T}.$$

The logic of the proof is as follows. For A4, we can prove the contrapositive: $\alpha_H(t) > 0 \implies t \in H$. This follows because under pairwise stability, all links are such that it is better for each participant to keep the link. This implies (by definition) that every node is in a network type that is within her preference class.

To prove A5 is satisfied, start with any pair of types t, s. Then consider the (new) pair of types that would exist if links were added between the two candidate nodes $(\bar{t}, \bar{s})$. The argument will show that the share of at least one of these new types is zero in the population.

They also show that under preference classes such as the one described above, A4 and A5 are necessary and sufficient for pairwise stable networks to exist. This argument relies on A3, which ensures that there is a balance between the value of a triangle and the value of a friend of friend.

The second main result concerns computation. The key observation is that A5 is stated as a product and, therefore, can be written as a quadratic function of allocation parameters. They have to define a Q such that A5 can be written as $\alpha' Q \alpha = 0$ at the optimal $\alpha_H(t)$. Q is defined in the following way: with elements such that both types t and s involved in $Q_{t,s}$ would prefer to link to each other if they were at length $2D$ apart.

The result is that given a structural parameter vector θ yielding $\mathrm{P}(\cdot)$, a network with type shares $\{\pi_x(t)\}$ satisfies A4 and A5 if and only if

$$\min_{\{\alpha_H(t):\ t \in H\}} \alpha' Q \alpha$$

such that

(1) $\sum_{t \in H} \alpha_H(t) = 1, \ \forall H$
(2) $\alpha_H(t) \geq 0, \forall t, H$
(3) $\sum_H \mathrm{P}_{H|x} \alpha_H(t) = \pi_x(t), \ \forall t$

is equal to zero.

To see the argument, observe that A5 holds if and only if

$$\sum_{\bar{H} \in \mathcal{H}} \sum_{\tilde{H} \in \mathcal{H}} \alpha_{\bar{H}}(t) \cdot \mathbf{1}\left\{\bar{t} \in \bar{H}\right\} \alpha_{\tilde{H}}(s) \cdot \mathbf{1}\left\{\tilde{s} \in \tilde{H}\right\} = 0.$$

Meanwhile, A4 is encoded in the constraint that allocation parameters are only over the preference classes.

4.6. **Discussion.** In this section I have reviewed several classes of econometric models of network formation that allow us to identify (sets of) parameters and, at times, establish when the estimators are consistent with a known asymptotic distribution as $n \to \infty$. One common aspect of almost every model that we have looked at is that payoffs are localized in some way. Payoffs were explicitly limited to the structure of the subgraph(s) that a reference node is a member of, and different models took different stands on the degree to which, ex ante, certain nodes were allowed to link with others.

The different modeling choices involved trading off restrictions on what sort of adjacency matrices were admitted against the degree of structure placed on equilibrium selection. For instance, the models of strong homophily imposed and network types imposed very little on equilibrium selection. But this forced very specific network structures: a near-block diagonal structure in the case of strong homophily, and a strongly sparse structure with limited value on triadic closure in the case of network types. Subgraph generated models allowed for more flexible modeling of network structure but required more separability and myopia in the microfoundations. The basic tension is that without limiting the degree of interdependence, through direct payoffs, strategic incentives, or otherwise, even a large network will not have enough information to either identify the formation parameters or estimate them consistently.

A role for experiments. Research on the econometrics of network formation has plenty to gain by interacting with empirical research. Experiments in the field, or even more in the lab, may teach us about what individuals take into account as they form networks (Burger and Buskens, 2009; Kearns, Judd, and Vorobeychik, 2012). How localized are individuals' preferences? How myopic are decisions? How separable are payoffs in subnetwork structure? Given the complexity of the modeling decisions and the sensitivity of the macrostructure of the resulting networks to these choices, it may be particularly important to take lessons from experimental economics as researchers develop further models.

The value of panel data. In this chapter, I have focused on estimating network formation when the researcher has access to an edge list from a single large network as well as a vector of covariates for each agent. Another strand of the literature has exploited panel data. I outline the ideas briefly in three seminal papers but a detailed discussion is beyond the scope of this chapter.

The first context in which panel data is useful is when the researcher wants to identify which agents influence others through the network when the network is unobserved. Manresa (2014) takes on this challenge by considering an environment in which agents' behavior depends on their neighbors' behavior, but the econometrician does not actually see the network.[22] A very simplified version of this framework is

$$y_{i,t} = \alpha + \beta \sum_j \frac{A_{ij}}{d_i} x_{j,t} + \delta x_{i,t} + \epsilon_{i,t}.$$

The econometrician observes $(y_{i,t}, x_{i,t})_{(i,t)\in V\times T}$, and the goal is to estimate (α, β, δ) as well as $\mathbf{A}$. If the underlying network is sparse, in the sense that the maximal degree is small relative to the number of time periods in the panel, the network structure can be recovered by using a pooled LASSO estimator (Tibshirani, 1996; Meinshausen and Bühlmann, 2006; Yuan and Lin, 2007; Meinshausen and Yu, 2009).[23] The intuition is that the panel provides information as to when correlations in x's of two agents lead to correlation in outcomes and, provided there is enough sparsity in the system, one can sort through the $\binom{n}{2}$ potential links and identify which are present using LASSO approaches.

The second context in which panel data is useful is when the researcher has a second-stage equation which relates outcomes to neighbors' outcomes as well as a first-stage equation that considers network formation. Both Goldsmith-Pinkham and Imbens (2013) and Graham (2014b) look at how unobserved heterogeneity can confound a system of equations an econometrician is interested in estimating. The intuition is that unobserved heterogeneity can make it so that certain agents are likely to interact and form a link. Moreover, the unobserved heterogeneity may directly influence outcomes, but by virtue of being unobserved, may bias the researcher's estimate of a peer effect. Goldsmith-Pinkham and Imbens (2013) develop a method to identify such endogeneity in a panel setting. Graham (2014b) shows that when a researcher has access to four periods of network data, by looking at how the network transitions, one can identify various parameters of network formation despite the unobserved heterogeneity.

5. Conclusion

In this chapter, I have surveyed econometric models of network formation that are estimated using data from a single, large network observed for one period. My goal has been to acquaint readers with network formation models used across a variety of disciplines. I have tried to highlight how well various models reflect patterns in empirical data and the trade-offs they present when thinking about microfoundations.

Empirical network data tend to be very sparse and highly clustered. This represents a challenge to the modeler, because both sparsity and clustering imply that the network

[22] Another issue that arises in the context of econometrics of network formation has to do with sampled networks (Chandrasekhar and Lewis, 2013). In a panel setting, the Manresa (2014) results allow such problems to be circumvented. A discussion of the econometrics of sampled networks is beyond the scope of this chapter.

[23] These results also have a connection to the literature on Gaussian graphical models.

carries limited information. Since links are rare, this implies that there are fewer effective observations, which can limit the degree to which a researcher can model heterogeneity. Further, because links are correlated, the researcher could in principle have far less than $\binom{n}{2}$ independent units of information.

Models that treat links as conditionally edge independent need to allow for correlation in links either through heterogeneity of nodes or through unobserved community assignment. The difficulty these models encounter is that they require a certain degree of network density in order to have enough information to identify the heterogeneity parameters of interest, but this may not be consistent with the sparsity of empirical data. Further, in many contexts, economic payoffs may not simply be bilateral, but instead depend on multilateral linking decisions. Meanwhile, models that more directly model correlation in links through payoffs that depend on other subnetwork structures local to the node in question have to deal with a trade-off between how much structure they impose in terms of equilibrium selection and how much structure they impose on the shape of the network (e.g., through forcing a block diagonal structure or strongly limiting the value of triadic closure).

The challenges presented here require the researcher to become familiar with a broad set of tools. As we have seen, the tools used in modeling network formation can come from a variety of topics spread across several disciplines (e.g., statistics, computer science, sociology, statistical mechanics). Further, researchers working in this space may be able to draw inspiration from empirical work. For instance, experiments that vary aspects of the environment and observe the resulting patterns of interaction may provide some insight about what modeling assumptions are likely to be more relevant when developing models.

References

Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008): "Mixed membership stochastic blockmodels." *Journal of Machine Learning Research*, 9, 1981–2014. 3.4

Albert, R. and A. Barabasi (2002): "Statistical mechanics of complex networks," *Reviews of modern physics*, 74, 47–97. 1, 2.1

Aldous, D. J. (1981): "Representations for partially exchangeable arrays of random variables," *Journal of Multivariate Analysis*, 11, 581–598. 3.4

——— (1985): *Exchangeability and related topics*, Springer. 3.4

Badev, A. (2013): "Discrete Games in Endogenous Networks: Theory and Policy," *mimeo: University of Pennsylvania*. 2, 4.1.1, 4.1.1, 4.3

Baldi, P. and Y. Rinott (1989): "On normal approximations of distributions in terms of dependency graphs," *The Annals of Probability*, 1646–1650. 4.2.3

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

Banerjee, A. V., A. Chandrasekhar, E. Duflo, and M. O. Jackson (2014): "Gossip: Identifying Central Individuals in a Social Network," *Available at SSRN 2425379*. 2

Barabasi, A. and R. Albert (1999): "Emergence of scaling in random networks," *Science*, 286, 509. 4.3

Bhamidi, S., G. Bresler, and A. Sly (2008): "Mixing time of exponential random graphs," *Arxiv preprint arXiv:0812.2265*. 4.1.3

Bickel, P., A. Chen, and E. Levina (2011): "The method of moments and degree distributions for network models," *Annals of Statistics*, 39, 2280–2301. 3.4

Bickel, P. J. and A. Chen (2009): "A nonparametric view of network models and Newman–Girvan and other modularities," *Proceedings of the National Academy of Sciences*, 106, 21068–21073. 3.4, 3.4, 5, 3.6

Bloch, F., G. Genicot, and D. Ray (2008): "Informal insurance in social networks," *Journal of Economic Theory*, 143, 36–58. 2.2

Bloch, F. and M. Jackson (2006): "Definitions of equilibrium in network formation games," *International Journal of Game Theory*, 34, 305âĂŞ318. 10

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2010): "Identification of social interactions," . 1

Bollobás, B., S. Janson, and O. Riordan (2011): "Sparse random graphs with clustering," *Random Structures & Algorithms*, 38, 269–323. 4.2

Bolthausen, E. (1982): "On the central limit theorem for stationary mixing random fields," *The Annals of Probability*, 1047–1050. 4.2.3

Borgs, C., J. Chayes, J. Ding, and B. Lucier (2010): "The Hitchhiker's Guide to Affiliation Networks: A Game-Theoretic Approach," *ArXiv:1008.1516v1*. 4.2.1

Boucher, V. and I. Mourifié (2012): "My friend far far away: Asymptotic properties of pairwise stable networks," *Available at SSRN*, 2170803. 4.4

Bresnahan, T. F. and P. C. Reiss (1991): "Entry and competition in concentrated markets," *Journal of Political Economy*, 977–1009. 1

BURGER, M. J. AND V. BUSKENS (2009): "Social context and network formation: An experimental study," *Social Networks*, 31, 63–75. 4.6

BUTTS, C. (2009): "Using Potential Games to Parameterize ERG Models," *working paper, UCI*. 4.1.1, 4.1.1, 4.1.2, 4.3

CAI, J., A. DE JANVRY, AND E. SADOULET (2012): "Social networks and the decision to insure," *Available at SSRN 2161686*. 2

CHAMBERLAIN, G. (1982): "Multivariate regression models for panel data," *Journal of Econometrics*, 18, 5–46. 3.5

CHANDRASEKHAR, A. AND R. LEWIS (2013): "Econometrics of sampled networks," Stanford working paper. 22

CHANDRASEKHAR, A. G. AND M. JACKSON (2014): "Tractable and Consistent Random Graph Models," *SSRN Working Paper: http://ssrn.com/abstract=2150428*. 3.3, 4.1.1, 13, 14

——— (2015): "A network formation model based on subgraphs," . 4.2, 6, 4.2.3

CHANDRASEKHAR, A. G., C. KINNAN, AND H. LARREGUY (2014): "Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field," . 1

CHATTERJEE, S. AND P. DIACONIS (2013): "Estimating and understanding exponential random graph models," *The Annals of Statistics*, 41, 2428–2461. 4.1.3

CHATTERJEE, S., P. DIACONIS, AND A. SLY (2010): "Random graphs with a given degree sequence," *Arxiv preprint arXiv:1005.1136*. 3.5, 3.6

CHRISTAKIS, N., J. FOWLER, G. IMBENS, AND K. KALYANARAMAN (2010): "An Empirical Model for Strategic Network Formation," *NBER Working Paper*. 4.3, 4.5.1

CLAUSET, A., M. E. NEWMAN, AND C. MOORE (2004): "Finding community structure in very large networks," *Physical review E*, 70, 066111. 3.4

COMOLA, M. AND M. FAFCHAMPS (2013): "Testing unilateral and bilateral link formation," *The Economic Journal*. 3.3

CURRARINI, S., M. JACKSON, AND P. PIN (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77, 1003–1045. 4.2.1

——— (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. 4.2.1

DE PAULA, A., S. RICHARDS-SHUBIK, AND E. TAMER (2014): "Identification of Preferences in Network Formation Games," Tech. rep., Mimeo, Carnegie Mellon University. 4.5

DE WEERDT, J. AND S. DERCON (2006): "Risk-sharing networks and insurance against illness," *Journal of Development Economics*, 81, 337–356. 2

DIACONIS, P., S. HOLMES, AND S. JANSON (2008): "Threshold graph limits and random threshold graphs," *Internet Mathematics*, 5, 267–320. 3.4

ERDŐS, P. AND A. RÉNYI (1959): "On random graphs," *Publ. Math. Debrecen*, 6, 156. 3.1

FAFCHAMPS, M. AND F. GUBERT (2007a): "Risk sharing and network formation," *The American Economic Review*, 97, 75–79. 3.3

——— (2007b): "The formation of risk sharing networks," *Journal of Development Economics*, 83, 326–350. 3.3

FRANK, O. AND D. STRAUSS (1986): "Markov graphs," *Journal of the American Statistical Association*, 832–842. 4.1.2

GALEOTTI, A. AND B. W. ROGERS (2013): "Strategic immunization and group structure," *American Economic Journal: Microeconomics*, 5, 1–32. 2.1

GIRVAN, M. AND M. E. NEWMAN (2002): "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, 99, 7821–7826. 3.4

GOLDENBERG, A., A. X. ZHENG, S. E. FIENBERG, AND E. M. AIROLDI (2010): "A survey of statistical network models," *Foundations and Trends® in Machine Learning*, 2, 129–233. 1

GOLDSMITH-PINKHAM, P. AND G. IMBENS (2013): "Social Networks and the Identification of Peer Effects," *Journal of Business and Economic Statistics*, 31:3, 253–264. 4.1.1, 4.4, 4.5.1, 4.6

GOLUB, B. AND Y. LIVNE (2010): "Strategic Random Networks: Why Social Networking Technology Matters," *SSRN working paper*. 4.2.1

GRAHAM, B. S. (2014a): "An empirical model of network formation: detecting homophily when agents are heterogenous," . 3.5, 3.5, 3.5, 3.6

——— (2014b): "Methods of Identification in Social Networks," . 1, 4.6

HAMMERSLEY, J. AND P. CLIFFORD (1971): "Markov fields on finite graphs and lattices," . 4.1.2

HANDCOCK, M. (2003): "Assessing degeneracy in statistical models of social networks," . 4.1.3

HILL, B. M. (1975): "A simple general approach to inference about the tail of a distribution," *The annals of statistics*, 3, 1163–1174. 4.3

HOLLAND, P. W., K. B. LASKEY, AND S. LEINHARDT (1983): "Stochastic blockmodels: First steps," *Social networks*, 5, 109–137. 3.3

HOLLAND, P. W. AND S. LEINHARDT (1981): "An exponential family of probability distributions for directed graphs," *Journal of the american Statistical association*, 76, 33–50. 3.5

JACKSON, M. (2008): *Social and economic networks*, Princeton: Princeton University Press. 6, 3.4

JACKSON, M., T. BARRAQUER, AND X. TAN (2012): "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102, 1857–1897. 1, 2.2

JACKSON, M. AND B. ROGERS (2007a): "Meeting strangers and friends of friends: How random are social networks?" *The American economic review*, 97, 890–915. 6

——— (2007b): "Relating network structure to diffusion properties through stochastic dominance," *The BE Journal of Theoretical Economics*, 7, 1–13. 2.1

JACKSON, M. AND A. WATTS (2001): "The Existence of Pairwise Stable Networks," *Seoul Journal of Economics*, 14(3), 299–321. 4.1.1

JACKSON, M. AND A. WOLINSKY (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44–74. 10

JACKSON, M. AND L. YARIV (2011): *Handbook of Social Economics*, San Diego: North Holland, chap. Diffusion, strategic interaction, and social structure. 2.1

KALLENBERG, O. (2006): *Probabilistic symmetries and invariance principles*, Springer Science & Business Media. 3.4

KARLAN, D., M. MOBIUS, T. ROSENBLAT, AND A. SZEIDL (2009): "Trust and Social Collateral*," *Quarterly Journal of Economics*, 124, 1307–1361. 2.2

KEARNS, M., S. JUDD, AND Y. VOROBEYCHIK (2012): "Behavioral experiments on a network formation game," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, ACM, 690–704. 4.6

KOENIG, M. (2012): "The formation of networks with local spillovers and limited observability," . 4.3

KOLOTILIN, A. (2013): "Estimation of a Scale-Free Network Formation Model," *Available at SSRN*. 4.3

LAZARSFELD, P. F., N. W. HENRY, AND T. W. ANDERSON (1968): *Latent structure analysis*, Houghton Mifflin Boston. 3.4

LEIDER, S., M. M
"OBIUS, T. ROSENBLAT, AND Q. DO (2009): "Directed Altruism and Enforced Reciprocity in Social Networks*," *Quarterly Journal of Economics*, 124, 1815–1851. 2

LESKOVEC, J., K. J. LANG, AND M. MAHONEY (2010): "Empirical comparison of algorithms for network community detection," in *Proceedings of the 19th international conference on World wide web*, ACM, 631–640. 3.4

LEUNG, M. (2014): "A Random-Field Approach to Inference in Large Models of Network Formation," *Stanford Working Paper*. 4.4, 4.4.4

LOVÁSZ, L. AND B. SZEGEDY (2006): "Limits of dense graph sequences," *Journal of Combinatorial Theory, Series B*, 96, 933–957. 3.5

MANRESA, E. (2014): "Estimating the Structure of Social Interactions Using Panel Data," . 4.6, 22

MEINSHAUSEN, N. AND P. BÜHLMANN (2006): "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, 1436–1462. 4.6

MEINSHAUSEN, N. AND B. YU (2009): "Lasso-type recovery of sparse representations for high-dimensional data," *Annals of Statistics*, 37, 246–270. 4.6

MELE, A. (2013): "A Structural Model of Segregation in Social Networks," *working paper*. 4.1.1, 4.1.1, 4.1.2, 4.1.2, 4.3, 4.5.1

MOBIUS, M., T. PHAN, AND A. SZEIDL (2010): "Treasure Hunt," *working paper*, http://misrc.umn.edu/wise/papers/p1-11.pdf. 2

NEWMAN, M. E. (2004): "Fast algorithm for detecting community structure in networks," *Physical review E*, 69, 066133. 3.4

NGATIA, M. (2011): "Social interactions and individual reproductive decisions," *Unpublished Working Paper*. 2

PENROSE, M. (2003): *Random Geometric Graphs*, Oxford University Press. 3.2

ROSS, N. ET AL. (2011): "Fundamentals of Stein's method," *Probab. Surv*, 8, 210–293. 4.2.3

SHALIZI, C. AND A. RINALDO (2012): "Consistency under Sampling of Exponential Random Graph Models," *ArXiv 1111.3054v3*. 4.1.3

Snijders, T. (2002): "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 3, 240. 4.1.3

Stein, C. (1986): "Approximate computation of expectations," *Lecture Notes-Monograph Series*, 7, i–164. 4.2.3

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

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. 4.6

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. 4.1.2

Yuan, M. and Y. Lin (2007): "Model selection and estimation in the Gaussian graphical model," *Biometrika*, 94, 19–35. 4.6