Network Data

December 11, 2019

(prepared for the *Handbook of Econometrics*, Volume 7A)

Bryan S. Graham[1]

INITIAL DRAFT: JUNE 2017, THIS DRAFT: SEPTEMBER 2019

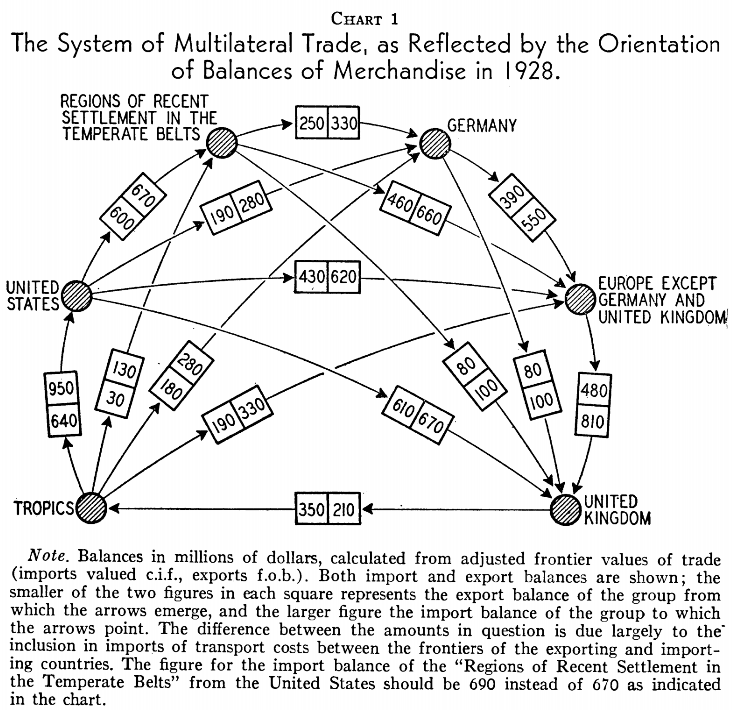
Abstract

Many economic activities are embedded in *networks*: sets of agents and the (often) rivalrous relationships connecting them to one another. Input sourcing by firms, interbank lending, scientific research, and job search are four examples, among many, of networked economic activities. Motivated by the premise that networks' structures are consequential, this chapter describes econometric methods for analyzing them. I emphasize (i) dyadic regression analysis incorporating unobserved agent-specific heterogeneity and supporting causal inference, (ii) techniques for estimating, and conducting inference on, summary network parameters (e.g., the degree distribution or transitivity index); and (iii) empirical models of strategic network formation admitting interdependencies in preferences. Current research challenges and open questions are also discussed.

[1]Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880 and National Bureau of Economic Research, e-mail: bgraham@econ.berkeley.edu, web: `http://bryangraham.github.io/econometrics/`. Financial support from NSF grants SES #1357499 and SES #1851647 is gratefully acknowledged. I am grateful for comments provided by the co-editors and other participants at a conference held at the University of Chicago in August of 2017. Portions of the material presented below benefited from conversations with Peter Bickel, Michael Jansson and Jim Powell. I am especially grateful to Eric Auerbach, Seongjoo Min, Chris Muris, Fengshi Niu and Konrad Menzel, as well as an anonymous referee, for written feedback which greatly improved the chapter. All the usual disclaimers apply.

Contents

Figure 1: World Trade Network in 1928



Notes: This figured appeared in Folke Hilgerdt's 1943 *American Economic Review* article "The case for multilateral trade". The figure shows aggregate trade balances between selected large countries and different regions of the world. The paper includes a narrative discussion of how the patterns of trade depicted in weighted digraph drawn in the figure developed historically.
Source: Reproduced from Hilgerdt (1943, Chart 1).

1 Introduction and summary

Many economic activities are embedded in *networks*: sets of agents and the (often) rivalrous relationships connecting them to one another. Firms generally buy and sell inputs not in anonymous markets, but via bilateral contracts (Kranton & Minehart, 2001). In addition to public listings, individuals gather information about job opportunities from friends and acquaintances (Granovetter, 1973). We similarly poll friends for information about new products, books, movies and so on (e.g., Jackson & Rogers, 2007; Banerjee et al., 2013; Kim et al., 2015). Banks generally meet reserve requirements through peer-to-peer interbank lending. The structure of this interbank lending network has profound implications for the vulnerability of the financial system to large negative shocks (Bech & Atalay, 2010; Gofman, 2017). Additional examples abound (cf., Jackson et al., 2017).

Although important exceptions exists, some highlighted below, economists historically avoided the study networks (see Figure 1).[2] This is now changing, very quickly, and for several reasons. First, starting in the 1990s economic theorists applied the tools of game theory to formally study network formation (e.g., Jackson & Wolinsky, 1996). In the resulting models agents add, maintain, and subtract links in order to maximize utility, with the realized network satisfying a pairwise stability equilibrium condition.[3] Second, in parallel to this theoretical work, a lively empirical and methodological literature on peer group and neighborhood effects also arose (e.g., Manski, 1993; Brock & Durlauf, 2001; Graham, 2008; Angrist, 2014). Finally, largely driven by questions in empirical industrial organization, econometricians made substantial progress on the econometric analysis of games (cf., Bajari et al., 2013; de Paula, 2013). Each of these literatures serve as foundations for material introduced below.

Outside of economics, two key initiators have been (i) the increasing availability of datasets with natural graph theoretic structure (see below for examples) and (ii) innovations in applied probability and theoretical statistics pertaining to random graph models (e.g., Diaconis & Janson, 2008). These innovations provide a foundation upon which recent work in statistics and machine learning on networks is largely based.

A consequence of these developments is the emergence of a small methodological literature on the econometrics of networks. Empirical applications with substantial network content, spurred largely by access to new datasets, arose more quickly (e.g., Fafchamps & Minten, 2002; De Weerdt, 2004; Conley & Udry, 2010; Atalay et al., 2011; Acemoglu et al., 2012; Banerjee et al., 2013; Barrot & Sauvagnat, 2016). Furthermore, these applications now span the major fields of our discipline. Nevertheless many open questions in the econometrics of networks remain. In this chapter I attempt to provide an account of recent progress as well as make suggestions for future research. My audience is both econometricians and empirical researchers.

I divide my discussion into five parts. The discussion draws from recent contributions to the analysis of networks made in probability, econometrics, and statistics (including machine learning); approximately in that order. After an initial outline of recent empirical research with a network dimension in economics, Section 3 introduces some basic probability tools that will prove useful for what follows. Several of these tools are of quite recent origin. Next, in Sections 4 to 6 I turn to the analysis of dyadic regression models. Such models go back, at least, to the pioneering work of Tinbergen (1962, Appendix VI) on gravity trade

[2]In contrast our colleagues in sociology studied networks from the outset of their discipline in its modern form. The monograph by Wasserman & Faust (1994) provides a somewhat dated introduction to this literature. See also Granovetter (1985).

[3]Other equilibrium concepts have been explored as well (cf., Bloch & Jackson, 2006).

models. Although dyadic regression is a core empirical method in international trade, as well as in certain areas of political science and development economics, a coherent inferential foundation for empirical practice is only now emerging. My discussion, in addition to covering methods of inference, discusses how to incorporate unobserved heterogeneity into dyadic regression models (Section 6). Here I appropriate and extend insights from panel data (Chamberlain, 1980, 1984, 1985; Hahn & Newey, 2004; Arellano & Hahn, 2007). This section also sketches out how to answer causal questions in dyadic settings.

Section 7 turns to the large network properties of several common network statistics. I focus on so-called *network moments*, or the frequencies with which certain low order subgraph configurations (e.g., triangles △) occur within a network. Subgraph counts, in the form of the triad census, were introduced by Holland & Leinhardt (1970) almost a half-century ago. Recent developments in probability and statistics have substantially improved our understanding of these counts (e.g., Diaconis & Janson, 2008; Bickel et al., 2011).

Subgraph counts may be of direct interest, but also serve as the building blocks of several popular network statistics, such as transitivity or moments of the degree distribution. Jackson et al. (2017) survey the mapping between different network statistics and economic phenomena and questions. My interest in network moments also stems from their value as inputs into structural model estimation in a manner akin to the way sample moments are paired with model moments in the simulated method of moments (e.g., Gourieroux et al., 1993). This idea is developed in Section 8.

The discussion of dyadic regression in Sections 4 to 6 rules out interdependencies in link formation. In dyadic models the utility two agents generate by forming a link is invariant to the presence or absence of links elsewhere in the network. Beginning with the seminal work of Jackson & Wolinsky (1996), the relaxation of this assumption is a central preoccupation of both theoretical and econometric researchers. When link formation decisions are interdependent, inefficient network structures may occur in equilibrium, making policy analysis interesting. Empirical network formation models allowing for interdependencies are also challenging to study. In a typical model many equilibrium network configurations can arise for any given parameter value; such models are incomplete (e.g., Tamer, 2003). In principle, standard tools developed in the context of economic games between a small number of agents apply. Practically speaking such methods are computationally infeasible in the many agent context of networks. Recent research proposes a variety of ways of getting around this conundrum.

Economists' interest in networks stems from the belief that their structure is consequential. For example, Loury (2002) argues that differences in social networks across Blacks and Whites drives, in part, racial inequality (cf., Graham, 2018b). Acemoglu et al. (2012) argue that the Leontief input-output structure of the economy shapes technology shock propagation. Alatas et al. (2016) show that network structure influences the flow and aggregation of information within rural villages. Theorists also study the interplay between network structure and agent behavior *on* that structure (Jackson & Yariv, 2011; Jackson & Zenou, 2015). Methodological research relating network structure to economic outcomes builds-upon the line of peer effects research initiated by Manski (1993). The paper by Bramoullé et al. (2009) is a nice, and influential, example of recent work along such lines.

This survey, however, does not review methods for the empirical analysis of behavior on networks. Instead I focus on modeling their *formation*. My motivation for this emphasis is two-fold. First, Blume et al. (2011) already survey work at the intersection of peer group effect identification and networks (cf., Blume et al., 2015; de Paula, 2017). Second, the current state of research in this area suggests that a better understanding of how networks form is a prerequisite for more credible research on their consequences.

Current research on the effects of network structure on outcomes largely treats it as exogenously given (although this is not always made explicit). This decision is one reason why research on peer effects and networks remains controversial a quarter century after Manski's foundational paper.[4] The focus maintained here, on *formation*, therefore seems to be a natural one. Ultimately, of course, the goal is to study the formation of networks and their consequences jointly, but such an integrated treatment remains largely aspirational at this stage. Although, Goldsmith-Pinkham & Imbens (2013) provide one recent "proof of possibilities" example of such an integrated approach. Qu & Lee (2015), Auerbach (2016), Badev (2017), and Johnsson & Moon (2017) represent other steps in this direction.

2 Examples, questions and notation

The analysis of datasets with natural graph theoretic structure has a long history in the other social sciences (e.g., Moreno, 1934), and more recently emerged as an area of focus within the statistics and machine learning community (e.g., Goldenberg et al., 2009; Kolaczyk, 2009). Although we were late adopters, interest in these types of datasets now also extends across virtually all fields of economics. Nevertheless, as already noted, appropriate methods for the analysis of network data are not widely available. Ad hoc and/or heuristically motivated approaches to estimation and inference abound in empirical work. Networks are characterized by complex dependencies across agents, as well as other difficult modeling, estimation and

[4] For example, Jackson et al. (2017, p. 81) argue that endogenous network formation, the tendency for the unobserved drivers of link formation and the behavior of interest to the econometrician to covary, poses a key challenge to "accurately estimating interactive effects in networked settings".

inferential challenges. These challenges are just starting be understood and solved. Before discussing methods for the analysis of network data, I briefly introduce some recent examples of empirical network research in economics. These examples also serve to introduce some basic notation.

2.1 Empirical analysis of trade flows

Figure 2 visually depicts international trade in bananas, a widely-eaten tropical fruit, in 2015. Each dot or *node* in the figure corresponds to a country. If, for example, Honduras, exports at least 50,000 tons of bananas to the United States, then there exists a *directed edge* from Honduras to the United States.[5] The exporting country (left node) is called the *tail* of the edge, while the importing country (right node) is its *head.* The set of all such exporter-importer relationships forms $G(\mathcal{V},\mathcal{E})$, a directed network or *digraph* defined on $N=|\mathcal{V}|$ vertices or agents (here countries). The set $\mathcal{V}=\{1,\ldots,N\}$ includes all agents (countries) in the network and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ the set of all directed links (exporter-importer relationships of 50,000 tons or greater) among them.[6] Let N be the *order* of the digraph and $|\mathcal{E}|$ its *size.* In what follows nodes may be equivalently referred to as vertices, agents, individuals, countries and so on depending on the context. Likewise edges may be called links, friendships, ties, arcs, relationships and so on.

There are $N=220$ countries in the banana network and hence up to $2\binom{220}{2}=48,180$ directed trading relationships among them. How might an econometrician model the presence or absence of a trading relationship from country i to j? Over fifty years ago Tinbergen (1962, Appendix VI) introduced gravity models, suitable for data of the type shown in Figure 2. In a gravity model trade between two countries, a *dyad* in network parlance, is modeled as a function of exporter and importer attributes (e.g., their gross domestic products), as well as dyad-specific covariates (e.g., physical distance between them). Generalizations of Tinbergen's approach are workhorses of modern empirical trade research (e.g., Santos Silva & Tenreyro, 2006; Helpman et al., 2008; Anderson, 2011).

Their ubiquity notwithstanding, serious open questions remain about how to estimate, and conduct inference on, the parameters of gravity trade models. Questions of particular interest here include how to account for the dependence across dyads sharing a country in common, how to incorporate country-specific (correlated) unobserved heterogeneity, and how to formalize causal policy effects in dyadic settings. As an example of the latter challenge, consider the effects of participation in multi-lateral trading agreements, such as the General Agree-

[5]In constructing this network, I binarized the underlying trade flow data to determine edge placement.
[6]Here $\mathcal{U}\times\mathcal{V}$ denotes the Cartesian product of the set $\mathcal{U}$ and $\mathcal{V}$ (i.e, $\mathcal{U}\times\mathcal{V}=\{(u,v)\,:\,u\in\mathcal{U},\,v\in\mathcal{V}\}$).

ment on Tariffs and Trade (GATT) or its successor, the World Trade Organization (WTO), on trade flows. Does trade increase across participating countries (Rose, 2004; Helpman et al., 2008)? While a mature literature on program evaluation suitable for single agent settings now exists (cf., Heckman & Vytlacil, 2007; Imbens & Wooldridge, 2009), a networked counterpart has yet to emerge.

2.2 Corporate governance

Next consider the affiliation network of (corporate board) directors and firms. This bipartite network $B(\mathcal{U}, \mathcal{V}, \mathcal{E})$ consists of two sets of agents, the set of possible directors, $\mathcal{U}$, and the set of firms, $\mathcal{V}$. Edges, $\mathcal{E}$, match directors to firms (i.e., corporate boards), and hence may only run between $\mathcal{V}$ and $\mathcal{U}$. A longstanding interest among corporate governance researchers centers on the implications of so-called board interlocks. When a single director sits on multiple corporate boards, then these corporations have interlocking directorates (Dooley, 1969). Interlocking directorships may facilitate collusion and other anti-competitive activities as well as, perhaps more positively, the diffusion of innovations in corporate governance (Davis, 1991, 1996).

Figure 3 plots the one-mode projection of the directors-to-firms bipartite network for S&P 1,500 firms in 2016. This projection generates an *undirected* network $G(\mathcal{V}, \mathcal{E})$ on the set of all firms, with an edge between any two firms sharing at least one director in common (i.e., with interlocking corporate boards). Large firms in United States are inter-connected via overlapping corporate board membership. On average firms share at least one board member in common with four other firms and over 80 percent of S&P 1,500 firms form a giant connected component of board interlocks. The board interlock network is also highly *transitive*: two firms are much more likely to share a director in common, if they also share one in common with a third firm.

Chu & Davis (2016) and Gualdani (ming) provide recent analyses of board interlocks as well as references to earlier work.

2.3 Production networks

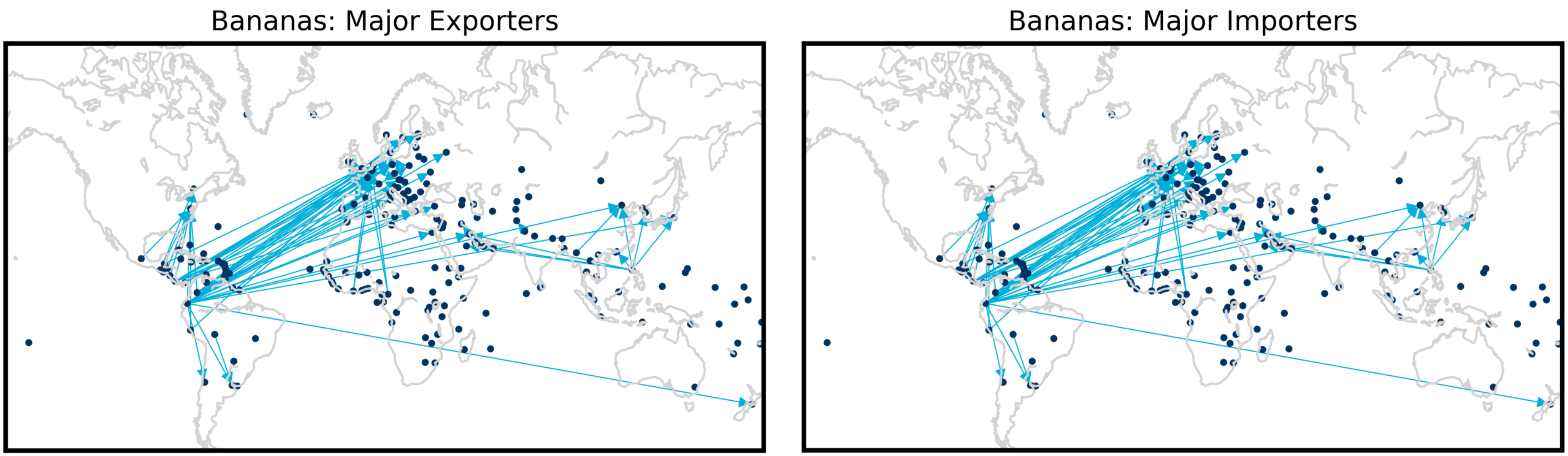
Atalay et al. (2011) study the production network of the United States economy. The sale and purchase of intermediate inputs between firms joins virtually all publicly traded corporations in the United States into one giant buyer-supplier network.

Serpa & Krishnan (2017) present evidence of productivity spillovers across firms linked together via supply chain relationships (cf., Acemoglu et al., 2016a). Acemoglu et al. (2012)

study the effect of the Leontief input-output structure of the US economy on shock propagation. Their analysis suggests that idiosyncratic technology shocks to critical input suppliers may have macro-level effects.

Bernard et al. (2018), using detailed supply-chain data from Japan, show how lowering supplier search costs allows firms to source inputs more efficiently, in turn lowering marginal production costs. The rich supply-chain data underlying the analysis of Bernard et al. (2018) is emblematic of the increasing availability of detailed supply chain network data from different countries (e.g., Dhyne et al., 2015). These datasets have the potential to dramatically improve our understanding of, for example the sources of heterogeneity in productivity across firms (e.g., Atalay et al., 2014) and the upstream and downstream implications of (horizontal) mergers (e.g., Fee & Thomas, 2004; Bhattacharya & Nain, 2011; Ahern & Harford, 2014), among many other areas of industrial organization and regulation policy.

Figure 2: World Trade in Bananas, 2015



<u>Source:</u> BACI-CEPII International Trade Database (cf., Gaulier & Zignago, 2010; De Benedictis et al., 2014) and author's calculations.
<u>Notes:</u> International trade of bananas in 2015 (HS6 code 080390). Each node in the figure represents a country (nodes are positioned at capital cities) and an edge between two nodes indicates the presence of at least 50,000 tons of directed banana flows (the head of each directed edge corresponds to the importing nation). In the left-hand panel node size is proportional to the total exports of bananas by the relevant nation, while in the right it is proportional to its total imports.

2.4 Research collaboration

Jaffe (1986), in a classic study, presented evidence of research and development (R&D) spillovers across technologically adjacent firms (Bloom et al., 2013; Acemoglu et al., 2016b). Such spillovers provide a motivation for firms to undertake collaborative R&D, a tendency which has increased over time (Hagedoorn, 2002; Tomasello et al., 2017). König et al. (2019) model the formation of R&D partnerships across firms theoretically and empirically, exploring the implications of network structure for optimal R&D subsidy policies. The structure of spillovers across firms, as well as the mechanisms whereby they form R&D partnerships, determines optimal policies.

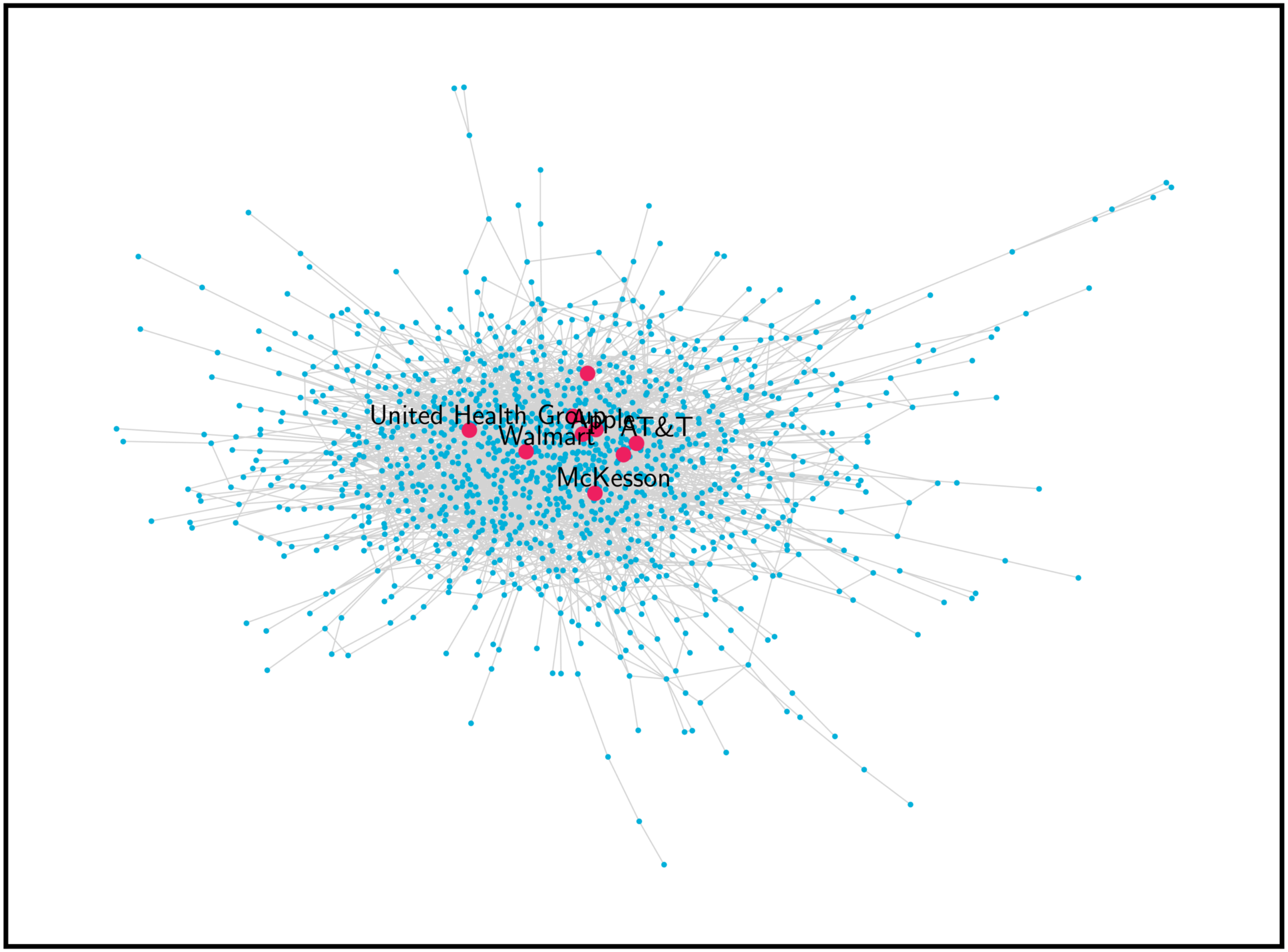
Ductor et al. (2014) study collaboration and research output among and across economists. Newman (2001) explores collaboration networks in the various sciences.

2.5 Risk-sharing across households

A classic question in development economics is whether households efficiently share risk through informal agreements (Townsend, 1994; Udry, 1994). Recently economists have directly collected information on risk-sharing relationships across households. For example, De Weerdt (2004) collected data on risk-sharing links across households in a village in Tanzania and empirically modeled the determinants of these links (cf., Fafchamps & Lund, 2003; Fafchamps & Gubert, 2007). Ambrus et al. (2014) investigate how the precise structure of links across households determines the amount of risk that can be insured, as well as the form of second best, more local, network structures.

Network structure now informs many other areas of development economics, including research on technology adoption and program take-up in rural settings (e.g., Banerjee et al., 2013; Kim et al., 2015), the productivity of small traders and firms (e.g., Fafchamps & Minten, 2002), and post-migration employment outcomes (Beaman, 2011; Munski, 2003), among other examples.

Figure 3: United States Corporate Board Interlocks, 2016



Source: Wharton Research Data Services (WRDS) - Institutional Shareholder Services (ISS) Directors dataset and author's calculations (cf., Chu & Davis, 2016).

Notes: The figure plots the largest connected component of the corporate board interlock network in 2016 among S&P 1,500 firms. The top 10 Fortune 500 firms in 2016 are the larger 'Rose Garden' colored nodes. A total of 1,216 firms belong to the largest connected component. See Newman (2010, p. 124 - 127) for details on how to construct one-mode projections of bipartite graphs.

2.6 Insurer-provider and referral networks for healthcare

Many features of the health care market naturally map into graphs. For example, physicians may have admitting privileges across multiple hospitals, insurers typically offer preferential terms to selected networks of providers, and doctors vary in the intensity with which they refer patients to one another.[7]

The welfare and economic implications of these networks are likely immense, given the magnitude of the health care sector in the United States economy. Ho (2009) represents one attempt to grapple with the network structure of healthcare markets.

2.7 Employment search

Ioannides & Loury (2004) survey the substantial literature on the interplay between social networks and job acquisition, a topic that has fascinated both sociologists and economists at least since Granovetter (1973). The growing availability of longitudinal register data from various countries provides an opportunity to study the interface between networks and inequality in the labor market more carefully.

For example, Saygin et al. (2014) use the Austrian Social Security Database to construct a co-worker network for middle aged workers in Austria. A co-worker is anyone who an individual has ever worked with previously. They find that the structure of these co-worker networks predict the ease with which workers find employment after establishment closures (i.e., mass layoffs). This paper provides a nice example of how new data may facilitate the re-visiting of a classic networks question (cf., Hensvik & Skans, 2016).

2.8 Questions

The examples outlined above represent only a small sample of recent appearances of network data in empirical economic research.[8] What do we hope to learn from this growing body of research? As noted in the introduction, empirical research on networks can usefully be divided between that which studies the *consequences* of networks and that which studies their *formation*. The premise of this chapter is that network linkages across agents are consequential. That is, I take as given that networks are important venues for shock propagation, information diffusion, learning and various types of peer interactions. Maintaining this premise justifies my focus on the econometric modeling of network formation.

[7] Barnett et al. (2011) and An et al. (2018) use patient referral patterns to map out relationships among physicians.

[8] de Paula (2017) and Jackson et al. (2017) provide additional references.

An analogy with the development of single agent models of discrete choice is useful. McFadden (1974), in a pioneering paper, initiated a research program on identifying and estimating random utility models of discrete choice. Empirical application, computation, semiparametric identification and estimation, the inclusion of unobserved choice attributes, and allowing for strategic behavior, all have been important accomplishments of this research program. These econometric models are, in turn, routinely used in virtually all areas of economics.

The goal here is analogous. Relational data are ubiquitous in economics, but econometric models for such data are not. The goal, therefore, is to develop models for these data, preferably with (i) strong microeconomic foundations, (ii) that allow for unobserved agent-level heterogeneity, and (iii) incorporate interdependencies in preferences over links. Also required are feasible methods of estimation and inference (and in this area interesting and challenging questions are abundant). The availability of econometric methods for network analysis will, in turn, allow for counterfactual policy and welfare analysis. How would a particular horizontal merger affect upstream supply chain structure? What is the effect on trade flows of Eurozone membership? Could a school principal increase friendships across races, or raise average achievement, by structuring classrooms under her purview differently?

Some readers may wish to skip Section 3 initially and instead start with Sections 4 to 6. They could then return to Section 3 before tackling Sections 7 and 8. Graph theoretic concepts and notation appears throughout the chapter. While many terms and definitions are formally defined, others are not. Missing definitions can be found in any basic graph theory textbook.

3 Basic probability tools: random graphs, graphons, graph limits and sampling

This section provides an informal introduction to key ideas from the applied probability literature on exchangeable random graphs. The main concepts are (i) exchangeable random graphs and their representation, (ii) subgraph densities or network moments, (iii) limits of sequences of exchangeable random graphs, and (iv) sampling. These ideas underlie a substantial share of recent research on the statistics of networks (e.g., Airoldi et al., 2008; Diaconis et al., 2008; Bickel & Chen, 2009; Bickel et al., 2011; Bhamidi et al., 2011; Chatterjee et al., 2011; Olhede & Wolfe, 2014; Orbanz & Roy, 2015; Gao et al., 2015).

Much of this statistics work has been motivated by research questions in computational biology and neuroscience (e.g., Picard et al., 2008). Link formation in these settings is not driven by purposeful agents. Consequently this research may initially appear rather distant from the concerns of econometricians. Nevertheless my view is that recent developments in probability and statistics have much to offer econometricians interested in networks (and also vice-versa, although making this second argument this is not on my agenda here).

The basic concepts introduced in this section appear frequently in later portions of the chapter.

3.1 Notation

Let $G(\mathcal{V},\mathcal{E})$ be a finite undirected network or graph defined on $N=|\mathcal{V}(G)|$ *vertices* or agents; here $\mathcal{V}(G)=\{1,\ldots,N\}$ denotes the set of all agents in the network.[9] Any two agents may be connected or not. The set of such links is recorded in the *edge* list $\mathcal{E}(G)=\{(i,j),(k,l),\ldots\}$, consisting of the (unordered) indices of all connected agent pairs. Call N the *order* of the network and $|\mathcal{E}(G)|$ its *size.* We can represent $G(\mathcal{V},\mathcal{E})$ by the $N\times N$ adjacency matrix $\mathbf{D}=[D_{ij}]_{i,j\in\mathcal{V}(G)}$ with ij^{th} element

$$D_{ij}=\begin{cases}1, & (i,j)\in\mathcal{E}(G)\\ 0, & \text{otherwise}\end{cases}.$$

For an undirected network, with self-ties or loops ruled out, such that $D_{ii}=0$ for $i\in\mathcal{V}(G)$, $\mathbf{D}$ is a symmetric binary matrix with a diagonal of structural zeros. I focus on undirected networks initially, but also present some results for directed networks and bipartite networks. Specific notation for these special cases will be introduced as needed.

In settings where it is useful to emphasize the order of G, I use the notation G_N. This is especially useful when considering sequences of graphs. Let $(i,j)\in\mathcal{E}(G)$ be an edge in G; sometimes I will abbreviate (i,j) as ij. The complete graph on p vertices is denoted by K_p.

Following Jackson (2008), let $G-ij$ denote the network obtained by deleting edge ij from G (if present), and $G+ij$ the network one gets after adding this link. Let $\mathbf{D}\pm ij$ denote the adjacency matrix associated with the network obtained by adding/deleting edge (i,j) from G. Let $\mathbb{D}_N$ denote the set of all $2^{\binom{N}{2}}$ possible adjacency matrices and $\mathbb{I}_N$ the set of all possible N-dimensional binary vectors.

Let $N(i)=\{j\in\mathcal{V}\,:\,ij\in\mathcal{E}\}$ be the set of agent i's *neighbors:* agents to which she is directly linked. The *degree* of agent i is given by the cardinality of this set. Equivalently agent i's degree may be computed by summing the elements of the i^{th} row of the adjacency matrix. Let ι_N be an $N\times 1$ vector of ones. The vector $\mathbf{D}_{+}=\mathbf{D}\iota_N$ is called the *degree sequence* of the

[9] If $\mathbb{X}$ is a set, then $|\mathbb{X}|$ denotes the cardinality of that set. If $\mathbf{X}$ is a matrix of reals, then $|\mathbf{X}|$ equals its (element-wise) absolute value.

network (typically we re-arrange the order of agents such that the elements of this vector are in ascending order).

I informally call a network dense if its size, or number of edges, is "close to" N^2 and sparse if its size is "close to" N. More precisely a sequence of graphs is *sparse* in the limit if the number of edges in it grows linearly with N, *dense* if this growth is quadratic.

There are $n \overset{def}{\equiv} \binom{N}{2} = \frac{1}{2}N(N-1)$ pairs of agents, or *dyads*, in a network consisting of N agents. Triples, quadruples and quintuples of agents are call *triads*, *tetrads* and *pentads* respectively. A tuple of 17 agents, which arises rather rarely in everyday empirical work, is evidently called a *septendecuple*. Not having formally studied Latin, I offer the reader no guidance on pronunciation.

Let $\sum_{i<j}$ be shorthand for $\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}$ with $\sum_{i<j<k}$ similarly defined. The *density* of a network,

$$P_N\left(\bullet\!-\!\bullet\right) \overset{def}{\equiv} \hat{\rho}_N \overset{def}{\equiv} \frac{2}{N(N-1)}\sum_{i<j} D_{ij},$$

equals the proportion of connected dyads. Let D_{i+} be the i^{th} element of the degree sequence. *Average degree,*

$$\hat{\lambda}_N \overset{def}{\equiv} (N-1)\,\hat{\rho}_N \overset{def}{\equiv} \frac{1}{N}\sum_{i=1}^{N} D_{i+},$$

equals the average number of links per agent in the network.

In what follows random variables are (generally) denoted by capital Roman letters, specific realizations by lower case Roman letters and their support by blackboard bold Roman letters. That is Y, y and $\mathbb{Y}$ respectively denote a generic random draw of, a specific value of, and the support of, Y. The abbreviations i.i.d., CLT , LLN and GGP stand for, respectively, "independent and identically distributed", "central limit theorem", "law of large numbers" and "graph generating process". For the vector $\mathbf{b}$, $\|\mathbf{b}\|_2$ denotes the Euclidean norm; for the matrix $\mathbf{B}$, $\|\mathbf{B}\|_F$ denotes the Frobenius norm. I use I_N to denote the $N \times N$ identity matrix. $\mathbb{N}$ denotes the set of natural numbers and $[Y_{ij}]_{i,j\in\mathbb{N}}$ an infinite two-dimensional array with ij^{th} element Y_{ij}.

I use the big-Omega notation $X_N = \Omega(Y_N)$ to denote that $X_N = O(Y_N)$ *and* $Y_N = O(X_N)$. The notation $\overset{D}{=}$ denotes equality in distribution, $\overset{def}{\equiv}$ a mathematical definition. Let θ be some parameter value in the space Θ. Let $S_N(\theta)$ be some statistic indexed by this parameter with population value θ_0. I let $S_N = S_N(\theta_0)$ denote the statistic evaluated at $\theta = \theta_0$. To economize on space I sometimes abbreviate $\Pr(Y = y|\, X = x)$ as $\Pr(Y = y|\, x)$ or $\Pr(y|\, x)$ and similarly for $\mathbb{E}[Y|\, x]$, $\mathbb{V}(Y|\, x)$ etc.

3.2 Exchangeable random graphs

Initially assume the unavailability of agent-specific covariates, making it natural to assume that agents are exchangeable (models with covariates, and a correspondingly weaker notion of exchangeability, feature in Sections 4, 5, 6 and 8). Let $\pi : \{1, \ldots, N\} \mapsto \{1, \ldots, N\}$ be a permutation of the node labels of $G(\mathcal{V}, \mathcal{E})$ and Π the set of all such permutations. The random graph G is *jointly exchangeable* if

$$[D_{ij}] \overset{D}{=} \left[D_{\pi(i)\pi(j)}\right] \tag{1}$$

for every permutation $\pi \in \Pi$.

In settings where node labels have no meaning, exchangeability is an implication of *a priori* researcher belief (and hence a natural modeling assumption). Consider a researcher analyzing the adjacency matrix associated with a set of friendship links among adolescents in a high school (e.g., Currarini et al., 2009), in the absence of node-specific covariates, there is no reason to change one's modeling approach after simultaneously applying a particular reshuffling of agents to *both* the rows and columns of $\mathbf{D}$ (cf., Rubin, 1981). Put differently, when node labels have no meaning, the probability attached to any isomorphism of G should be the same as that attached to G itself.

There are many interesting statistics of $\mathbf{D}$ which are invariant to simultaneous row and column permutations. Examples include a network's density, diameter and triangle ($\triangle$) count. A family of such statistics, network moments, is introduced below. Exchangeability suggests that a statistical model should attach different probabilities to networks with different values of such (permutation invariant) statistics, but the same probability to two networks which are isomorphic (which will share common values of any permutation invariant statistic).

An exchangeable model with strategic interaction

Most extant models of network formation satisfy condition (1). As an example, which will help to fix some ideas, consider the model of strategic network formation with bilateral transfers studied by Graham & Pelican (2020). Let $\nu_i : \mathbb{D}_N \to \mathbb{R}$ be a utility function for agent i, which maps networks into utility. Define the marginal utility of edge ij for agent i as

$$MU_{ij}(\mathbf{D}) = \begin{cases} \nu_i(\mathbf{D}) - \nu_i(\mathbf{D} - ij) & \text{if } D_{ij} = 1 \\ \nu_i(\mathbf{D} + ij) - \nu_i(\mathbf{D}) & \text{if } D_{ij} = 0 \end{cases}. \tag{2}$$

From Bloch & Jackson (2006), a network is *pairwise stable with transfers* if the following condition holds.

Definition 1. (Pairwise stability with Transfers) The network $G(\mathcal{V}, \mathcal{E})$ is pairwise stable with transfers if
(i) $\forall (i,j) \in \mathcal{E}(G)\,,\ MU_{ij}(\mathbf{D}) + MU_{ji}(\mathbf{D}) \geq 0$
(ii) $\forall (i,j) \notin \mathcal{E}(G)\,,\ MU_{ij}(\mathbf{D}) + MU_{ji}(\mathbf{D}) < 0$

If the network in hand is a pairwise stable one, then any links actually present generate (weakly) positive utility (on net for the two agents on each side of a link). Unobserved links, in contrast, would not generate net positive utility if present.

Graham & Pelican (2020) focus on a general family of parametric utility functions which includes, among others, the specification

$$\nu_i(\mathbf{d}|\mathbf{A}, \mathbf{B}, \mathbf{V}^*; \gamma_0) = \sum_j d_{ij}\left[A_i + B_j + \gamma_0\left(\sum_k d_{ik}d_{jk}\right) - V_{ij}^*\right] \tag{3}$$

with $\mathbf{V}^* = \left[V_{ij}^*\right]$, $\mathbf{A} = [A_i]$ and $\mathbf{B} = [B_i]$. Under (3), assuming $\gamma_0 > 0$, dyad $\{i, j\}$ will generate more utility when forming a link if they already share many links or “friends” in common (i.e., if $\sum_k d_{ik}d_{jk}$ is large). Here A_i and B_j are agent-specific “extroversion” and “popularity” parameters, the effect of which is to generate degree heterogeneity (cf., Graham, 2017). The term V_{ij}^* is an idiosyncratic dyad-specific utility shifter. Graham & Pelican (2020) leave the joint distribution of $\mathbf{A}$ and $\mathbf{B}$ unrestricted, but here I will assume that $\{(A_i, B_i)\}_{i=1}^N$ is an i.i.d. sequence which is independent of $\left\{\left(V_{ij}^*, V_{ji}^*\right)\right\}_{i,j\in\{1,\ldots,N\}, i<j}$, also assumed i.i.d.

When the utility function is of the form given in (3) the marginal utility agent i gets from a link with j is

$$MU_{ij}(\mathbf{d}|\,\mathbf{A}, \mathbf{B}, \mathbf{V}^*; \gamma_0) = A_i + B_j + \gamma_0\left(\sum_k d_{ik}d_{jk}\right) - V_{ij}^*.$$

Pairwise stability then implies, conditional on the realizations of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{V}^*$, and the value of externality parameter, γ_0, that the observed network must satisfy, for $i = 1, \ldots, N-1$ and $j = i+1, \ldots, N$

$$D_{ij} = \mathbf{1}\left(U_i + U_j + 2\gamma_0\left(\sum_k D_{ik}D_{jk}\right) \geq V_{ij}\right) \tag{4}$$

with $U_i = A_i + B_i$ and $V_{ij} = V_{ij}^* + V_{ji}^*$. Equation (4) defines a system of $\binom{N}{2} = \frac{1}{2}N(N-1)$ nonlinear simultaneous equations. Any solution to this system – and there will typically be multiple ones – constitutes a pairwise stable (with transfers) network.[10]

[10]Note that in this example existence of an equilibrium is easy to show using Tarski's (1955) fixed point

As written, model (4) is incomplete (cf., de Paula, 2013). Even if we assume that the observed network is a pairwise stable one, we have not specified a mechanism for selecting, when there are multiple ones, a specific equilibrium configuration. To complete the model, following the more careful development in Pelican & Graham (2019), let $\mathcal{N}_{\mathbf{d}}(\mathbf{V};\mathbf{U},\gamma)$ equal the probability that configuration $\mathbf{D}=\mathbf{d}$ is selected. If $\mathbf{d}$ is not an equilibrium – given $\mathbf{U}$, $\mathbf{V}$ and γ – then $\mathcal{N}_{\mathbf{d}}(\mathbf{V};\mathbf{U},\gamma)=0$. If $\mathbf{d}$ is the unique equilibrium then $\mathcal{N}_{\mathbf{d}}(\mathbf{V};\mathbf{U},\gamma)=1$. If $\mathbf{d}$ is one of several equilibria, then $0\leq\mathcal{N}_{\mathbf{d}}(\mathbf{V};\mathbf{U},\gamma)\leq 1$ etc.

For $\mathbb{D}_N$ the net of all $N\times N$ undirected adjacency matrices, we have that $\sum_{\mathbf{d}\in\mathbb{D}_N}\mathcal{N}_{\mathbf{d}}(\mathbf{V};\mathbf{U},\gamma)=1$. The conditional likelihood of observing network wiring $\mathbf{D}=\mathbf{d}$ is therefore

$$\Pr\left(\mathbf{D}=\mathbf{d}\right|\mathbf{U};\gamma)=\int_{\mathbf{v}\in\mathbb{R}^n}\mathcal{N}_{\mathbf{d}}(\mathbf{v};\mathbf{U},\gamma)\,f_{\mathbf{V}}(\mathbf{v})\,\mathrm{d}\mathbf{v}.$$

The $\binom{N}{2}$ equilibrium conditions (4) indicate that if $\mathbf{d}=[d_{ij}]$ is an equilibrium, then so is $\mathbf{d}_\pi\overset{def}{\equiv}\left[d_{\pi(i)\pi(j)}\right]$. Hence as long as the equilibrium selection mechanism is also invariant to index permutations, as is natural to require, condition (1) holds.

Under the null of no strategic interaction, $\gamma=0$, the likelihood simplifies to

$$\Pr\left(\mathbf{D}=\mathbf{d}\right|\mathbf{U};0)=\int_{\mathbf{v}\in\mathbb{R}^n}\mathcal{N}_{\mathbf{d}}(\mathbf{v};\mathbf{U},0)\,f_{\mathbf{V}}(\mathbf{v})\,\mathrm{d}\mathbf{v} \tag{5}$$

with

$$\mathcal{N}_{\mathbf{d}}(\mathbf{v};\mathbf{U},0)=\prod_{i=1}^{N-1}\prod_{j=i+1}^{N}\mathbf{1}\left(U_i+U_j\geq v_{ij}\right)^{d_{ij}}\times\mathbf{1}\left(U_i+U_j<v_{ij}\right)^{1-d_{ij}}.$$

Since $\{(V_{ij})\}_{i,j\in\{1,\ldots,N\},i<j}$ is i.i.d., if we further assume that $f_{V_{12}}(v)=e^v/\left[1+e^v\right]^2$, the logistic density, explicitly evaluating the integral in (5) yields

$$\Pr\left(\mathbf{D}=\mathbf{d}\right|\mathbf{U};0)=\prod_{i=1}^{N-1}\prod_{j=i+1}^{N}\left[\frac{\exp\left(U_i+U_j\right)}{1+\exp\left(U_i+U_j\right)}\right]^{d_{ij}}\left[\frac{1}{1+\exp\left(U_i+U_j\right)}\right]^{1-d_{ij}}, \tag{6}$$

which is the likelihood associated with the so-called β-model of Frank (1997) and Chatterjee et al. (2011).

A feature of the β-model is that links form independently *conditional* on the latent agent-specific effects $\{U_i\}_{i=1}^N$. Equation (6) consists of a product of $\binom{N}{2}$ conditionally independent

theorem.

likelihood contributions.

Evidently, this conditional independence structure is not typically a feature of the model when $\gamma > 0$, such that strategic interaction is present. To see why by means of a simple example, consider a network consisting of just three homogenous agents (i.e., $U_1 = U_2 = U_3 = 0$). Initially assume that both V_{12} and V_{13} are less then zero, but that $0 < V_{23} \leq 2\gamma_0$. This corresponds to edges $(1,2)$ and $(1,3)$ generating so much intrinsic utility that they will form irrespective of what other edges may or may not be present in the network. In contrast, the intrinsic utility attached to edge $(2,3)$ falls in an intermediate range: the edge forms if edges $(1,2)$ and $(1,3)$ are present – such that agents 2 and 3 share agent 1 as a friend in common – and does not form if they are absent. This configuration of utility shocks is depicted in the left-hand panel of Figure 4. The unique equilibrium outcome in this case is a triangle (△) network.

If, instead, V_{12} and V_{13} are both greater than 2γ, such that the $(1,2)$ and $(1,3)$ edges never form because of their low intrinsic utility (again irrespective of what other edges may or may not be present in the network), then the $(2,3)$ edge will not form either. This scenario is depicted in the right-hand panel of Figure 4. The unique equilibrium outcome in this case is an empty (∴) network.

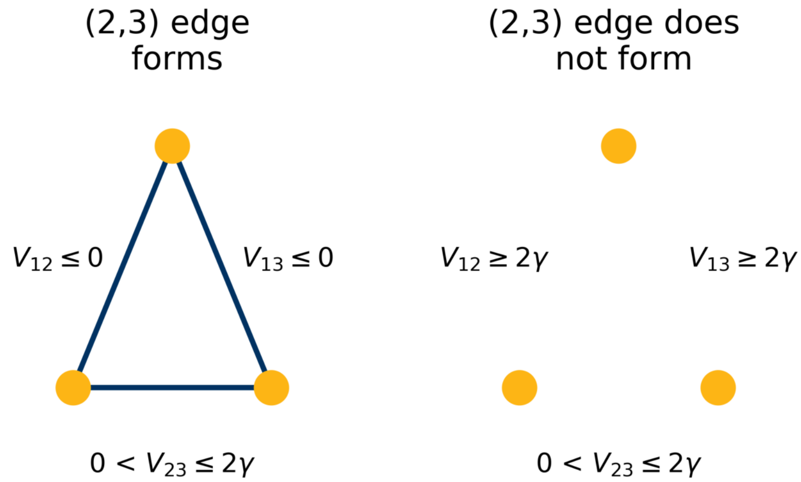
This simple example shows that D_{23} need not vary independently of D_{12} and D_{13} conditional on (U_1, U_2, U_3) in the presence of strategic interaction $(\gamma > 0)$. Such conditional independence *is* a feature of the β-model $(\gamma = 0)$. While the model is exchangeable both when $\gamma > 0$ and when $\gamma = 0$, the conditional independence of edges only obtains under the no strategic interaction null.

3.3 Conditionally independent dyad (CID) models and the graphon

Having established that a network probability model should satisfy the joint exchangeability condition (1), it is important to articulate classes of models that do so. One such family of models, suggested by the last example, are conditionally independent dyad (CID) models (Chandrasekhar, 2015; Shalizi, 2016). In these models each agent is characterized by an unobserved latent attribute, U_i. The N agents in the network in hand are viewed as independent random draws from some population, such that the $\{U_i\}_{i=1}^N$ are independently and identically distributed. Conditional on the agent-specific latent variables $\mathbf{U} = (U_1, \ldots, U_N)'$ edges form independently with

$$D_{ij} | \, U_i, U_j \sim \text{Bernoulli}\left(h\left(U_i, U_j\right)\right),$$

Figure 4: Dependent link formation



Notes: Both panels depict the unique pairwise stable equilibrium associated with the shown triple of dyad-level utility shifters V_{12}, V_{13} and V_{23} and agent-level heterogeneity parameters U_1, U_2 and U_3 identically equal to zero. In both panels the realized value of V_{23} is the same, but whether $D_{23} = 1$ or 0 varies with the realized values of V_{12} and V_{13}. If V_{12} and V_{13} are sufficiently low, then $D_{23} = 1$; if they are sufficiently high, then $D_{23} = 0$. Links are not conditionally independent given $\{U_i\}_{i=1,2,3}$.

for every dyad $\{i, j\}$ with $i < j$. Here $h(u, v) = h(v, u)$ for all $(u, v) \in \mathbb{U} \times \mathbb{U}$ is a symmetric edge probability function. In anticipation of results to come, call this function a *graphon*: short for **graph** functi**on**.

Conditional on the latent agent-specific effects the likelihood of the network is

$$\Pr\left(\mathbf{D} = \mathbf{d} | \, \mathbf{U} = \mathbf{u}\right) = \prod_{i<j} h\left(u_i, u_j\right)^{d_{ij}} \left[1 - h\left(u_i, u_j\right)\right]^{1-d_{ij}}.$$

Unconditional on $\mathbf{U}$, the likelihood equals

$$\Pr\left(\mathbf{D} = \mathbf{d}\right) = \int \cdots \int \left\{ \prod_{i<j} h\left(u_i, u_j\right)^{d_{ij}} \left[1 - h\left(u_i, u_j\right)\right]^{1-d_{ij}} \right\} \prod_{i=1}^{N} f_U\left(u_i\right) \mathrm{d}u_i, \qquad (7)$$

where $f_U(u)$ is the density of U. Importantly (7) allows for dependence across dyads which share agents in common. Independence holds only conditional on the latent agent attributes (Graham, 2017). Similar independence restrictions play a prominent role in the econometrics of panel data (Chamberlain, 1984; Arellano & Honoré, 2001).

It is an easy exercise to show that (7) is compatible with the finite joint exchangeability restriction (1).

The β-model, introduced above, belongs to the family of CID models with a graphon of

$$h(u,v) = \frac{\exp(u+v)}{1+\exp(u+v)}.$$

Random threshold graphs (e.g., Diaconis et al., 2008) are also members of this family with graphon

$$h(u,v) = \mathbf{1}(F_U(u) + F_U(v) \geq \alpha),$$

and $F_U(u)$ the CDF of U.

It is important to realize that CID models constitute only a subset of all jointly exchangeable random graph models when N – the number of agents in the network – is finite. As shown by means of the example introduced above, strategic interaction in link formation can induce dependence across elements of the adjacency matrix that evidently cannot be eliminated by conditioning (see Figure 4 above). Although not all exchangeable models are CID ones, this family of models plays an outsized role in extant large sample theory for networks.

3.4 Aldous-Hoover representation theorem and the graphon

Joint exchangeability imposes more structure on the network probability distribution when there are an *infinite* number of agents. Specifically, if we strengthen (1) to hold for any permutation of a finite number of the indices of the infinite sequence $\mathbb{N} = \{1,2,3,\ldots\}$, we have a generalization of de Finetti (1931) type exchangeability of an infinite sequence, appropriate for infinite random graphs. In independent work Aldous (1981) and Hoover (1979) showed the following representation result for infinite random adjacency matrices (cf., Kallenberg, 2005).

Theorem 1. (Aldous-Hoover) *A random adjacency matrix* $[D_{ij}]_{i,j\in\mathbb{N}}$ *is jointly exchangeable if and only if there is a measurable function* $g:[0,1]^4 \to \{0,1\}$ *such that*

$$[D_{ij}] \overset{D}{=} [g(\alpha, U_i, U_j, V_{ij})]$$

for α, $\{U_i\}_{i\in\mathbb{N}}$ *, and* $\{V_{ij}\}_{i,j\in\mathbb{N},i<j}$ *independently and identically distributed* $\mathcal{U}[0,1]$ *random variables with* $V_{ij} = V_{ji}$.

Here α is a mixing parameter, analogous to the one appearing in de Finetti's (1931) classic representation theorem for exchangeable binary sequences.[11] Theorem 1 implies that if

[11] To make the connection with de Finetti (1931) transparent Aldous (1981, Lemma 1.5) also shows that an infinite sequence $\{Y_i\}_{i=1}^{\infty}$ is exchangeable if and only if there exists a measurable function f such that $[Y_i] \overset{D}{=} [f(\alpha, U_i)]$.

network agents are exchangeable for all N, then we can proceed 'as if' edges formed according to a CID model or a mixture of such models.

Exploiting the fact that the elements of $\mathbf{D}$ are binary, we can simplify Theorem 1 as follows. Averaging over V_{ij} yields

$$h(\alpha, u_i, u_j) \overset{def}{\equiv} \int_0^1 g(\alpha, u_i, u_j, v)\,\mathrm{d}v$$

from which we get the more convenient representation, for $i < j$,

$$[D_{ij}] \overset{D}{=} [\mathbf{1}(V_{ij} \leq h(\alpha, U_i, U_j))]. \tag{8}$$

This is, of course, just a conditional edge independence model (or, more precisely, a mixture of such models). In what follows I focus on inference which conditions on the empirical distribution of the data; consequently α can often safely be ignored. When this is the case I suppress the α argument in the graphon, writing $h(U_i, U_j)$. See Bickel & Chen (2009) and Menzel (2017) for additional discussion.

Theorem 1 motivates an approach to nonparametric modeling of *large* networks that proceeds 'as if' links form independently conditional on the agent-specific latent variables $\mathbf{U} = (U_1, \ldots, U_N)'$. This is convenient because CID models induce a very particular dependence structure across the rows and columns of the network adjacency matrix.

Consider, without loss of generality, agents 1, 2 and 3. In a CID model D_{12} and D_{13} may covary; the dyads $\{1, 2\}$ and $\{1, 3\}$ share the agent 1 in common and hence both links form, in part, based on the value of U_1. However D_{12} and D_{13} vary independently conditional on U_1, U_2 and U_3 (hence the conditionally independent dyad nomenclature). Links involving pairs of dyads which share no agents in common, for example D_{12} and D_{34}, form independently.

The structured pattern of dependence, independence and conditional independence associated with CID models facilitates the development of LLNs and CLTs that can be applied to statistics of the adjacency matrix. A group of statistics for which some large network distribution theory is available are network moments.

3.5 Network moments

Almost fifty years ago Holland & Leinhardt (1970) suggested that a network's architecture could be usefully summarized by its average local structure. Agent exchangeability, in conjunction with Theorem 1, also motivates an approach to network modeling based on the frequency of low order subgraph configurations (i.e., the number of edges, two stars, triangles, squares, k-stars etc).

Consider, for example, the set of all $\binom{N}{3}$ *triads* – unordered triples of agents – in a network; what fraction of these triads take two-star ∧ or triangle △ configurations? These frequencies, called *network moments* by Bickel et al. (2011), feature prominently in research by sociologists (e.g., Granovetter, 1973; Coleman, 1988; Gould & Fernandez, 1989) and computational biologists (e.g., Milo et al., 2002; Pržulj et al., 2004); albeit in the context of two largely independent and desynchronized literatures.

In economics, network moments play an increasingly important role in empirical research as well. Examples include Jackson et al. (2012), who explore, theoretically and empirically, how different triad configurations can support infrequent favor exchange between agents; Atalay et al. (2011), who calibrate a model of buyer-seller networks to the US economy by modeling its degree distribution[12]; and de Paula et al. (2018), who present conditions under which (a variant of) network moments (partially) identify preferences in a structural model of strategic network formation.

Network moments, in addition to being important summary statistics for graphs, play an important role in (i) the distribution theory for dyadic regression discussed in Sections 4 and 5, (ii) understanding the degree distribution and (iii) structural model estimation. The material which follows is dense.

Subgraphs and isomorphisms

The exact sense in which a network is summarized by its moments can be made precise using the graphon, as introduced above, and the notion of a graph limit, which will be introduced below (Diaconis & Janson, 2008; Lovász, 2012). First we require a formal definition of a subgraph. There are two definitions used by empirical network researchers.

Definition 2. (PARTIAL SUBGRAPH) Let $\mathcal{V}(S) \subseteq \mathcal{V}(G)$ be any subset of the vertices of G and $\mathcal{E}(S) \subseteq \mathcal{E}(G) \cap \mathcal{V}(S) \times \mathcal{V}(S)$, then $S = (\mathcal{V}(S), \mathcal{E}(S))$ is a *partial subgraph* of G.

A partial subgraph S of G consists of a subset of agents in G and a *subset* of all edges among $\mathcal{V}(S)$ also appearing in G. Counts of partial subgraphs are often referred to as *network motif* counts (e.g., Milo et al., 2002), although this terminology is not used consistently. The two star motif $S =$ ∧ is a partial subgraph of $G =$ △. Note that in this example S does not include the edge between agents, numbered clockwise from the top, 2 and 3.

Definition 3. (INDUCED SUBGRAPH) Let $\mathcal{V}(S) \subseteq \mathcal{V}(G)$ be any subset of the vertices of G and $\mathcal{E}(S) = \mathcal{E}(G) \cap \mathcal{V}(S) \times \mathcal{V}(S)$, then $S = (\mathcal{V}(S), \mathcal{E}(S))$ is an *induced subgraph* of G.

[12] Below I show that network moments and moments of the degree distribution are closely connected.

An induced subgraph S includes *all* edges in G connecting any two agents in $\mathcal{V}(S)$. Although $S =$  is a partial subgraph of $G =$ , it is not an induced one. Counts of induced subgraphs are often referred to as *graphlet* counts (e.g., Pržulj et al., 2004), although again not consistently so.

Consider two graphs, R and S, of the same order. Let $\varphi : \mathcal{V}(R) \to \mathcal{V}(S)$ be a bijection from the nodes of R to those of S. The bijection $\varphi : \mathcal{V}(R) \to \mathcal{V}(S)$ *maintains adjacency* if for every dyad $i, j \in \mathcal{V}(R)$ if $(i, j) \in \mathcal{E}(R)$, then $(\varphi(i), \varphi(j)) \in \mathcal{E}(S)$; it *maintains non-adjacency* if for every dyad $i, j \in \mathcal{V}(R)$ if $(i, j) \notin \mathcal{E}(R)$, then $(\varphi(i), \varphi(j)) \notin \mathcal{E}(S)$. If the bijection maintains both adjacency and non-adjacency we say it *maintains structure*.

Definition 4. (GRAPH ISOMORPHISM) The graphs R and S are *isomorphic* if there exists a structure-maintaining bijection $\varphi : \mathcal{V}(R) \to \mathcal{V}(S)$.

In what follows I use the notation $R \cong S$ to denote that "R is isomorphic to S."

Two special families of motifs/graphlets will play a prominent role in the analysis of network summary statistics presented in Section 7 below. First, a *p-cycle* is p^{th} order graphlet with nodes labeled (or relabeled) such that its edges form a cycle:

$$\mathcal{E}(S) = \{(i_1, i_2), (i_2, i_3), \ldots, (i_p, i_1)\}.$$

A p-cycle is a connected graphlet with p edges on p nodes. As one transverses a p-cycle graphlet no vertex is crossed more than once except for the first/last one. Important examples of p-cycles are triangles ($S =$ ) and 4-cycles ($S =$ ).

Second, a *tree* is a connected graph with no cycles. The number of edges on a p^{th} order tree is $p - 1$; a feature which will prove highly convenient. Important examples of trees are p-star graphlets, such as two-stars ($S =$ ) and three-stars ($S =$ ). Trees will feature in the analysis of the degree distribution given below. Trees are also called connected acyclic graphs.

Induced subgraph density

Using Definitions 3 and 4 we can formally introduce the induced subgraph density. This will be our first measure of the frequency with which a specific low-order local configuration of links appears within a network. Let S be a p^{th}-order graphlet of interest (e.g., $S =$  or $S =$ ), $\text{iso}(S)$ the group of isomorphisms of S, and $|\text{iso}(S)|$ its cardinality. It is helpful to observe that $|\text{iso}(S)|$ equals the number of (partial) subgraphs of K_p that are isomorphic to S. For example, $|\text{iso}($  $)| = 3$ since there are three ways to draw a two-star configuration on three vertices. G_N is the real world network under study.

Let $\mathbf{i}_p \subseteq \{1, 2, \ldots, N\}$ be a set of p integers. If we require that $i_1 < i_2 < \cdots < i_p$, then there are $\binom{N}{p}$ such integer sets; denote this set of integer sets by $\mathcal{C}_{p,N}$. If all that is required is that $i_k \neq i_l$ for $k \neq l$, then there are $\frac{N!}{(N-K)!}$ such integer sets; denote this set of integer sets by $\mathcal{A}_{p,N}$.

Let the vertex set of S be $\{1, \ldots, p\}$. Let $G[\mathbf{i}_p]$ denote the induced subgraph of G associated with vertex set $\mathbf{i}_p$. Since we wish to compare S and $G[\mathbf{i}_p]$ it will be convenient to relabel the latter. Let $\tilde{G}[\mathbf{i}_p]$ be a relabelling of $G[\mathbf{i}_p]$ such that $i_1 = 1, i_2 = 2, \ldots, i_p = p$ so that $kl \in \mathcal{E}\left(\tilde{G}[\mathbf{i}_p]\right)$ if $i_k i_l \in \mathcal{E}(G[\mathbf{i}_p])$. Let $\mathbf{i}_p \sim \text{Uniform}(\mathcal{A}_{p,N})$; the frequency with which $\tilde{G}_N[\mathbf{i}_p]$ equals S is then

$$P_N(S) \overset{def}{\equiv} \Pr\left(S = \tilde{G}_N[\mathbf{i}_p]\right), \ \mathbf{i}_p \sim \text{Uniform}(\mathcal{A}_{p,N}). \tag{9}$$

Call (9) the *induced subgraph density* of S in G_N. Alternatively we can write

$$P_N(S) = \frac{\Pr(S \cong G_N[\mathbf{i}_p])}{|\text{iso}(S)|}, \ \mathbf{i}_p \sim \text{Uniform}(\mathcal{C}_{p,N}) \tag{10}$$

The induced subgraph frequency of S in G_N equals the fraction of injective mappings $\varphi : \mathcal{V}(S) \to \mathcal{V}(G_N)$ that preserve both edge adjacency *and* non-adjacency. Direct computation of this fraction yields the equalities

$$\begin{aligned} P_N(S) =& \frac{N!}{(N-p)!} \sum_{\mathbf{i}_p \in A_{p,N}} \mathbf{1}\left(S = \tilde{G}_N[\mathbf{i}_p]\right) \\ =& \frac{1}{\binom{N}{p}|\text{iso}(S)|} \sum_{\mathbf{i}_p \in \mathcal{C}_{p,N}} \mathbf{1}(S \cong G_N[\mathbf{i}_p]) \\ \overset{def}{\equiv}& t_{\text{ind}}(S, G_N) \end{aligned} \tag{11}$$

In order to understand the mechanics of computing (11) it is useful to reformulate, one again, its definition. Let $\mathbf{D}_{[\mathbf{i}_p, \mathbf{i}_p]}$ be the $p \times p$ sub-adjacency matrix constructed by removing all rows and columns of $\mathbf{D}$ except those in $\mathbf{i}_p = \{i_1, \ldots, i_p\}$. We can check for whether $G[\mathbf{i}_p]$ is an isomorphism of S by inspecting the elements of the $\mathbf{D}_{[\mathbf{i}_p, \mathbf{i}_p]}$ sub-adjacency matrix.

Consider the two star triad $S =$  , we can express $\mathbf{1}(S \cong G_N[\mathbf{i}_p])$ in terms of $\mathbf{D}_{[\mathbf{i}_p, \mathbf{i}_p]}$ as

$\mathbf{1}($ $\cong G_N[\mathbf{i}_p]) = D_{i_1 i_2} D_{i_1 i_3}(1 - D_{i_2 i_3}) + D_{i_1 i_2}(1 - D_{i_1 i_3}) D_{i_2 i_3} + (1 - D_{i_1 i_2}) D_{i_1 i_3} D_{i_2 i_3}.$ (12)

We have $|\text{iso}($ $)| = 3$ with the three terms to the right of the equality in (12) equal to indicators for these three possible isomorphisms (on triad/vertex set $\{i_1, i_2, i_3\}$). In general $\mathbf{1}(S \cong G_N[\mathbf{i}_p])$ may be defined in terms of $\mathbf{D}_{[\mathbf{i}_p, \mathbf{i}_p]}$ with the number of components equal to the number of possible isomorphisms of S. There is only one isomorphism of the 

configuration, yielding a second example of

$$\mathbf{1}\left(\triangle \cong G_N\left[\mathbf{i}_p\right]\right) = D_{i_1 i_2} D_{i_1 i_3} D_{i_2 i_3}.$$

Recognizing that $t_{\text{ind}}\left(S, G_N\right)$ is a functional of the adjacency matrix of G_N allows us to easily compute its expectation when edges form according to the conditional edge independence model (8). Once again consider the two star configuration; iterated expectations and conditional independence of edges given $\mathbf{U} = \left(U_1, \ldots, U_N\right)'$ yield

$$\begin{aligned}\mathbb{E}\left[D_{i_1 i_2} D_{i_1 i_3}\left(1 - D_{i_2 i_3}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[D_{i_1 i_2} D_{i_1 i_3}\left(1 - D_{i_2 i_3}\right)\middle| \mathbf{U}\right]\right] \\ &= \mathbb{E}\left[h\left(U_{i_1}, U_{i_2}\right) h\left(U_{i_1}, U_{i_3}\right)\left[1 - h\left(U_{i_2}, U_{i_3}\right)\right]\right] \\ &= \int\int\int h\left(t, u\right) h\left(t, v\right)\left[1 - h\left(u, v\right)\right] \mathrm{d}t\mathrm{d}u\mathrm{d}v\end{aligned}$$

(and also that the value of $\mathbb{E}\left[D_{i_1 i_2} D_{i_1 i_3}\left(1 - D_{i_2 i_3}\right)\right]$ is invariant to permutations of its indices). Finally we have, recalling that $\left|\text{iso}\left(\wedge\right)\right| = 3$,

$$\mathbb{E}\left[\mathbf{1}\left(\wedge \cong G_N\left[\mathbf{i}_p\right]\right)\right] = 3 \cdot \int\int\int h\left(t, u\right) h\left(t, v\right)\left[1 - h\left(u, v\right)\right] \mathrm{d}t\mathrm{d}u\mathrm{d}v,$$

for $\mathbf{i}_p \sim \text{Uniform}(\mathcal{C}_{p,N})$. For a generic graphlet configuration we have

$$\begin{aligned}\mathbb{E}\left[t_{\text{ind}}\left(S, G_N\right)\right] &= \left|\text{iso}\left(S\right)\right|^{-1} \mathbb{E}\left[\mathbf{1}\left(S \cong G_N\left[\mathbf{i}_p\right]\right)\right] \\ &= \mathbb{E}\left[\prod_{\{i,j\} \in \mathcal{E}(S)} h\left(U_i, U_j\right) \prod_{\{i,j\} \in \mathcal{E}\left(\bar{S}\right)} \left[1 - h\left(U_i, U_j\right)\right]\right] \\ &\stackrel{def}{\equiv} P\left(S\right)\end{aligned} \tag{13}$$

where $\bar{G}$ denotes the complement of the graph G: the graph defined on the same nodes as G with an edge present if, and only if, it is not present in G. The graph sum of G and $\bar{G}$ therefore coincides with the complete graph $K_{|\mathcal{V}(G)|}$.

Call the *expectation* of $t_{\text{ind}}\left(S, G_N\right)$ the induced subgraph density of S in the *graphon* $h\left(\cdot\right)$ and write it as, in an abuse of notation, $\mathbb{E}\left[t_{\text{ind}}\left(S, G_N\right)\right] = t_{\text{ind}}\left(S, h\right) = P\left(S\right)$. Clearly $P_N\left(S\right)$ is an unbiased estimate of $t_{\text{ind}}\left(S, h\right) = P\left(S\right)$ *when the true network generating process is of the CID type.* Notice how the graphon provides a language for connecting *empirical* graphlet counts, first studied by Holland & Leinhardt (1970), with well-defined probabilistic objects. This connection will prove useful for developing a procedure for conducting inference on $P\left(S\right)$ using the sample graph G_N. Since $P\left(S\right)$ generally varies with the graphon $h\left(u, v\right)$, the idea is that by identifying $P(S)$ for enough specific configurations (e.g., $S =$ , , ,  etc.), we may be able to identify $h(u,v)$ itself (cf., Bickel et al., 2011).

Injective homomorphism density

A second notion of subgraph density also appears in some of the results which follow. Let $S \subseteq G$ denote that S is a partial subgraph of G. Using Definitions 2 and 4, we can also define what I will call, following Lovász (2012), the injective homomorphism density.[13] The homomorphism density gives the probability that *a (partial) subgraph of* $G_N[\mathbf{i}_p]$, for $\mathbf{i}_p$ chosen uniformly at random from $\mathcal{A}_{p,N}$, is equal to S. Alternatively the homomorphism density equals the fraction of injective mappings $\varphi : \mathcal{V}(S) \to \mathcal{V}(G_N)$ that preserve edge adjacency. These mappings do not need to preserve non-adjacency.[14] The *injective homomorphism density* of S in G_N equals

$$\begin{aligned} Q_N(S) &= \frac{1}{\binom{N}{p}|\mathrm{iso}(S)|} \sum_{R \subseteq K_N, R \cong S} \mathbf{1}(R \subseteq G_N) \qquad (14) \\ &= \frac{1}{\binom{N}{p}|\mathrm{iso}(S)|} \sum_{R \subseteq K_N, |V(R)|=p} \mathbf{1}(R \cong S) \prod_{\{i,j\} \in \mathcal{E}(R)} D_{ij} \\ &\overset{def}{\equiv} t_{\mathrm{inj}}(S, G_N) \end{aligned}$$

The two equivalent definitions are given to develop familiarity with notation. To understand expression (14) it is helpful to calculate the injective homomorphism density of $S =$  in $G_N =$ . There are three isomorphisms of the two star configuration such that $\binom{4}{3}|\mathrm{iso}($$)| = 4 \cdot 3 = 12$. Next consider the summation in the first line of (14). This summation is over all 3^{rd} order partial subgraphs of K_4 which are isomorphic to $S =$ . There are exactly 12 two star partial subgraphs in K_4 (three for each of its four triads), a total of 8 of these configurations are subgraphs of G_N such that $t_{\mathrm{inj}}($, $) = \frac{8}{12}$. Note that the induced subgraph density of $S =$  in $G_N =$  is just $\frac{2}{12}$.

[13]The Lovász (2012) monograph presents several different notions of a subgraph density. The two introduced here were chosen for their connection to actual empirical practice. See also Diaconis & Janson (2008).

[14]In contrast the induced subgraph density requires preservation of both adjacency and non-adjacency.

Under an Aldous-Hoover GGP we have

$$\begin{aligned}
\mathbb{E}\left[t_{\text{inj}}\left(S,G_N\right)\right] &= \frac{1}{\binom{N}{p}\left|\text{iso}\left(S\right)\right|}\sum_{R\subseteq K_N,\left|V(R)\right|=p}\mathbf{1}\left(R\cong S\right)\mathbb{E}\left[\mathbb{E}\left[\prod_{\{i,j\}\in\mathcal{E}(R)}D_{ij}\middle|U_1,\ldots,U_N\right]\right] \\
&= \mathbb{E}\left[\prod_{\{i,j\}\in\mathcal{E}(S)}h\left(U_i,U_j\right)\right] \\
&\overset{def}{\equiv} Q\left(S\right).
\end{aligned}$$

Call the expectation of $t_{\text{inj}}\left(S,G_N\right)$ the injective homomorphism density of S in the *graphon* $h\left(\cdot\right)$ and write it as $\mathbb{E}\left[t_{\text{inj}}\left(S,G_N\right)\right] = t_{\text{inj}}\left(S,h\right) = Q\left(S\right)$.

3.6 Graph limits

Let G_N be a finite exchangeable graph with adjacency matrix $\mathbf{D}$. Let

$$h_{G_N}\left(u,v\right) = \begin{cases} 1 & \text{if } \left(\lceil uN\rceil, \lceil vN\rceil\right)\in\mathcal{E}\left(G_N\right) \\ 0 & \text{otherwise} \end{cases}.$$

Observe that $h_{G_N}\left(u,v\right)$ is a valid graphon and further that

$$t_{\text{ind}}\left(S,G_N\right) = t_{\text{ind}}\left(S,h_{G_N}\right)$$

for any S of order $K \leq N$ (Chatterjee, 2017, p. 28). This equality connects the definition of the induced subgraph frequency of S in G_N, denoted by $P_N\left(S\right)$ in equation (11), with its "population" counterpart – equation (13). It also motivates the idea of the graphon as the appropriate limit object for a sequence of graphs, G_N. If the subgraph frequency

$$\lim_{N\to\infty} t_{\text{ind}}\left(S,h_{G_N}\right)$$

converges to some limit for *all* fixed subgraphs S, then we say that G_N has a limit. Lovász & Szegedy (2006) showed the natural limiting object is a graphon (i.e, heuristically, $h_{G_N}\to h$ as $N\to\infty$). Diaconis & Janson (2008) connect this finding with the Aldous-Hoover representation theorem. Collectively these results motivate an approach to summarizing a network by the frequency of different low order subgraph configurations within it; by its *average local structure*. Lovász (2012) provides a rigorous and compresensive introduction to theory of graph limits.

3.7 Sampling

In this chapter I will adopt two perspectives on "sampling". In the first we view the network in hand as the one induced by a random sample of agents from some large (i.e., infinite) population. Let G_∞ be an (infinite) exchangeable random graph. Let $\mathcal{V}$ be a random sample of agents of size N from G_∞. We assume that the observed network, G_N, coincides with the subgraph induced by this random sample of vertices:

$$G_N = G_\infty \left[\mathcal{V}\right]. \tag{15}$$

Let $\mathbf{D}_\infty = [D_{ij}]$ with $i, j \geq 1$ be the adjacency matrix of G_∞. Exchangeability implies the characterization

$$D_{ij} = \mathbf{1}\left(h\left(\alpha, U_i, U_j\right) \geq V_{ij}\right) \tag{16}$$

with α, U_i and $V_{ij} = V_{ji}$ independent $\mathcal{U}\left[0,1\right]$ random variables (cf., Aldous, 1981; Hoover, 1979). Here $h : \left[0,1\right]^3 \rightarrow \left[0,1\right]$ is symmetric in its second and third arguments.

Under (15) the elements of $\mathbf{D}$, the adjacency matrix for the network in hand, also obey the characterization (16). The "sampling distribution" of some statistic of $\mathbf{D}$, say $t_N\left(\mathbf{D}\right)$, is simply the one induced by repeated random sampling from the underlying infinite population. We calculate limit distributions by studying the sampling distribution of $t_N\left(\mathbf{D}\right)$ as $N \rightarrow \infty$.

An advantage of this first perspective it that is allows the econometrician to fully exploit the independence/dependence structure associated with the Aldous-Hoover Theorem. If the graph in hand is the one induced by a random sample of agents from some infinite exchangeable population, then we can proceed "as if"

$$D_{ij}|\, U_i, U_j \sim \text{Bernoulli}\left(h\left(U_i, U_j\right)\right) \tag{17}$$

for $i = 1, \ldots, N-1$ and $j = i+1, \ldots, N$. Although (17) is a nonparametric data generating process, it is a structured one. We can use this structure to our advantage.

An unattractive feature of this perspective is that if the density of the population graph is very low, then that of the sampled graph may be zero with high probability. To see this point heuristically assume that the population consists of N^* agents, with N^* very large. Assume that average degree, λ, is some small positive constant that does not dependent on N^*. The probability of observing an edge between the two independent random draws from the population is thus

$$\Pr\left(D_{12} = 1\right) = \frac{\frac{1}{2}\lambda N^*}{\binom{N^*}{2}} \approx \frac{\lambda}{N^*}.$$

Boole's inequality then gives a probability of observing at least one edge in our sampled network no greater than $\binom{N}{2}\lambda/N^{*}$, which will be close to zero when $N << \sqrt{N^{*}}$. When the population graph is "sparse", it is quite likely that the subgraph induced by a random sample of agents from it will be empty and hence completely uninformative. See Crane (2018, Chapter 3) for more discussion and examples.

This example raises two questions. First, how does one sample from a large sparse graph in practice? I ignore this question here, but flag it as an interesting one which merits thought. The monograph by Crane (2018) surveys extant work in this area. Second, if the sampling is fictitious (i.e., analysis is based upon the full graph), what mistakes might be made by proceeding "as if" we had randomly sampled from some (now entirely hypothetical) large graph?

To answer the second question is useful to return to an empirical example. Atalay et al. (2011) study the supply chain network of large publicly traded firms in the United States. Their network is not sampled, but rather constructed from Securities and Exchange Commission (SEC) reports filed by the entire universe of publically trade firms. If the model of network formation of interest is a conditional independent dyad (CID) one, then we are free to proceed "as if" the observed network were generated according to (17). If, instead, we view the network in hand as, for example, an equilibrium of a finite N-player supply chain formation game, then it may be difficult to justify (17); strategic interaction may induce dependence across links that cannot be conditioned away. We cannot appeal directly to the Aldous-Hoover Theorem.

As in de Finetti (1931), the Aldous-Hoover Theorem requires that the agent indices constitute an infinite sequence. However, just as the de Finetti result fails for finite sequences (e.g., Diaconis, 1977), but approximately holds when the sequence is large enough (e.g., Diaconis & Freedman, 1980), the hope is that in *large* (but finite) networks Theorem 1 remains useful (cf., Volfosky & Airoldi, 2016).

One possibility would be to assume that N is large enough such that a representation like (17) "approximately" holds. One could then conduct inference on model parameters by comparing observed network moments with model generated ones. The sampling distribution of the observed network moments would be calculated assuming an Aldous-Hoover DGP (which is appropriate for N large enough). I sketch this idea in a bit more detail in Section 8 below. Many gaps in this discussion remain. Alternatively we could proceed along the lines of Menzel (2016). In this approach we would approximate our finite player network formation game, with a limit game which is easier to deal with (see Section 8).

3.8 Adding sparsity: the Bickel & Chen (2009) model

For any finite network of unlabelled agents, exchangeability is a natural, indeed unavoidable, modeling assumption. Unfortunately its extension to infinite exchangeability, as needed for Theorem 1, has the unattractive implication that the network is either empty or dense in the limit. Specifically a (random) agent will either never form links or do so infinitely often as $N \to \infty$. Denseness and sparseness are limit properties of infinite sequences of graphs. Any empirical network is neither "dense" nor "sparse", it just is what it is. However, in most real world networks the numbers of agents and links are of similar magnitudes. This suggests that approximation results based on sequences of graphs that are sparse in the limit may be more useful than those with dense limits. Whether this is, in fact, the case remains an open question (Green & Shalizi, 2017).

One way to model sequences of graphs with sparse limits, while still preserving the analytic convenience of conditional independence across edges, was proposed by Bickel & Chen (2009). The Bickel-Chen model is the default one in the nonparametric statistics and machine learning literatures on random graphs.

Let G_N be a random network of order N generated according to (8). The expected average number of links an agent has in this network, that is *average degree*, equals

$$\lambda_N = (N-1)\,\rho_\alpha \tag{18}$$

for $\rho_\alpha = \int h(\alpha, u, v)\,\mathrm{d}u\mathrm{d}v$. Average degree (18) either tends toward infinity or is zero, depending on whether ρ_α is greater than or equal to zero.

To extend model (8) so that it can accommodate sparse graph sequences Bickel & Chen (2009) define the conditional density

$$w_\alpha(u, v) = f_{U_i, U_j | D_{ij}, \alpha}\left(u, v | \, D_{ij} = 1, \alpha\right).$$

Next observe that since $f_{U_i, U_j|\alpha}(u, v|\, \alpha) = 1$ on $[0, 1]^2$ we get can decompose the graphon as

$$h(\alpha, u, v) = \rho_\alpha w_\alpha(u, v). \tag{19}$$

With this parameterization, Bickel & Chen (2009) and Bickel et al. (2011) argue that it is natural to let $\rho_\alpha = \rho_{\alpha,N}$, but retain independence of $w_\alpha(u, v)$ from N. Suppressing the α argument (it is never identifiable), they write

$$\Pr\left(D_{ij} = 1 | \, U_i = u, U_j = v\right) = h_N(u, v) = \rho_N w(u, v). \tag{20}$$

The rate at which $\rho_N \to 0$ then controls the rate of average degree growth as N grows large. If $\lambda_N = (N-1)\rho_N \to \lambda$ with $0 < \lambda < \infty$ as $N \to \infty$, then the graph is *sparse*. If $\lambda_N = \Omega(N)$ we say the graph is *dense*, $\lambda_N = \Omega(\ln N)$ *semi-dense* etc. Many of the results presented below require that $\lambda_N = \Omega(N^{\alpha})$ for some $0 < \alpha \leq 1$, despite the fact that $\lambda_N = \Omega(1)$ might best describe real world networks (where average degree is generally low even when N is very large). In what follows I will try to highlight those few known results which can accommodate sparse graph sequences.

3.9 Further reading

Orbanz & Roy (2015) provide a non-technical introduction to the probability literature on exchangeable random arrays; the monograph by Kallenberg (2005) a more complete development. Crane (2018) also surveys this material, at a fairly accessible level, and with a somewhat contrarian point of view.

Lovász (2012) provides an overview of the theory of graph limits. Diaconis & Janson (2008) connect much of this theory to the older literature on exchangeable random arrays.

4 Dyadic regression

Jan Tinbergen's 1962 report *Shaping the World Economy*, commissioned by Twenty Century Fund, featured, along with its sculptural title, a remarkable empirical analysis of trade flows (Tinbergen, 1962). Table VI-1 in that report presented the results of a least squares fit of the logarithm of exports *from* country i *to* country j onto a constant, the (log) Gross National Product (GNP) of both countries i and j, the (log) distance between i and j, and a variety of other covariates capturing different relationships between i and j. Tinbergen's (1962) analysis was based upon a sample of $N = 18$ countries, or $N(N-1) = 306$ directed trading relationships.[15]

Table VI-1 of Tinbergen (1962) presents the results of what I will call a dyadic regression analysis. This particular analysis continues to serve as prototype for a substantial body of empirical work in international trade (Anderson, 2011). Dyadic regression analyses also appear in other areas of social science research. They have been used, to give just a few recent examples, to study the onset of war among nation states (e.g., Russett & Oneal, 2001), risk-sharing across households (e.g., De Weerdt, 2004; Fafchamps & Gubert, 2007; Attanasio et al., 2012), supply chain linkages across firms (e.g., Atalay et al., 2011, Table S3), the

[15] A second analysis, based upon a larger sample of countries, was also reported upon in Table VI-4 of the report.

formation of commercial R&D collaborations (König et al., 2019, Table 4), and co-camping behavior among hunter-gathers (Apicella et al., 2012, Tables S2 to S49).

Familiar methods of econometric analysis appropriate for single agent models, typically utilizing a random sample from the population of interest, are ill-suited for dyadic settings (cf., Cameron & Golotvina, 2005). Consequently, considerable confusion and controversy is associated with dyadic analyses in practice (e.g., Erikson et al., 2014). It is remarkable that, over a half-century after Tinbergen's (1962) pioneering analysis of trade flows across countries, and also given the considerable empirical work that has followed, a textbook treatment of estimation and inference methods for gravity and other dyadic regression models remains unavailable.

4.1 Population and sampling framework

Let $i \in \mathbb{N}$ index agents in an infinite population of interest. Associated with each agent is the observable attribute $X_i \in \mathbb{X} = \{x_1, \ldots, x_L\}$. This attribute partitions the population into $L = |\mathbb{X}|$ subpopulations which I will refer to as "types". Let $\mathbb{N}(x) = \{i : X_i = x_l\}$ be the index set for type l agents. Although L may be very large, I assume that the size of each subpopulation, $|\mathbb{N}(x)|$, is infinite with positive frequency (i.e., $\Pr(X_i = x_l) > 0$ for $l = 1, \ldots, L$).

When all observable agent attributes are discretely-valued, then $\mathbb{X}$ simply enumerates all distinct combinations of these attributes (e.g., $X = x_l$ might correspond to a Hispanic female, living in the Florida, with 12 years of schooling and two college-educated parents). More heuristically we can think of $\mathbb{X}$ as consisting of the support points of a multinomial approximation to the support of a bundle of attributes, some of which might be continuously-valued. The finite support restriction is made in order to invoke a representation result due to Crane & Towsner (2018); I do not think it is essential.

Associated with each ordered pair of agents is the scalar directed outcome $Y_{ij} \in \mathbb{Y} \subseteq \mathbb{R}$. I will refer to agent i as the "ego" of the directed dyad and agent j as its "alter". In the context of the trade example the ego agent is the exporting country, the alter the importing one. The *adjacency matrix* $[Y_{ij}]_{i,j\in\mathbb{N}}$ collects all such outcomes into an infinite random array.

From the standpoint of the econometrician, the indexing of agents within subpopulations homogenous in X_i is arbitrary: agents of the same type are exchangeable. Exchangeability of agents within subpopulations homogenous in X_i induces a particular form of exchangeability on the adjacency matrix. This form of exchangeability, in turn, induces a particular form of dependence across the rows and columns of $[Y_{ij}]_{i,j\in\mathbb{N}}$. The structure of this dependence allows for the formulation of LLNs and CLTs.

Let $\sigma_x : \mathbb{N} \to \mathbb{N}$ be any permutation of a finite number of the agent indices which satisfies the restriction

$$\left[X_{\sigma_x(i)}\right]_{i\in\mathbb{N}} = \left[X_i\right]_{i\in\mathbb{N}}. \tag{21}$$

Condition (21) constrains index permutations to occur among agents of the same type (i.e., we may permute the indices in $\mathbb{N}(x)$, but not those within, for example, $\mathbb{N}(x) \cup \mathbb{N}(x')$). Crane & Towsner (2018) call a network *relatively exchangeable* with respect to X (or X-exchangeable) if

$$\left[Y_{\sigma_x(i)\sigma_x(j)}\right]_{i,j\in\mathbb{N}} \overset{D}{=} \left[Y_{ij}\right]_{i,j\in\mathbb{N}} \tag{22}$$

for all permutations σ_x satisfying (21). X-exchangeablility is a natural generalization of joint exchangeability, as introduced in the context of the Aldous (1981) and Hoover (1979) Theorem earlier.

A insightful way to think about condition (22) is in terms of vertex colored graphs. Associate X_i with the color of a vertex; condition (22) states that all colored graph isomorphisms are equally probable. Since, when vertices of the same color are exchangeable, there is no reason to attach more or less probability to particular isomorphisms of a given vertex colored graph, any probability model for $\left[Y_{ij}\right]_{i,j\in\mathbb{N}}$ should be consistent with condition (22). As long as X_i encodes all the vertex-specific information available to the econometrician, then X-exchangeability is a nature *a priori* modeling restriction.

Let α, $\{(U_i, X_i)\}_{i\geq 1}$ and $\{(V_{ij}, V_{ji})\}_{i\geq 1, j\geq 1}$ be (sequences of) i.i.d. random variables, additionally independent of one another, and consider the random array $\left[Y^*_{ij}\right]_{i,j\in\mathbb{N}}$ generated according to the rule

$$Y^*_{ij} = \tilde{h}\left(\alpha, X_i, X_j, U_i, U_j, V_{ij}\right) \tag{23}$$

with $\tilde{h} : [0,1] \times \mathbb{X} \times \mathbb{X} \times [0,1]^3 \to \mathbb{Y}$ a measurable function (we normalize α, U_i and V_{ij} to have support on the unit interval without loss of generality). Clearly a graph generated according to (23) is X-exchangeable (cf., Crane, 2018, Chapter 8).

Here α is a mixing parameter analogous to the one appearing in de Finetti's (1931) original representation theorem. Since this parameter is unidentified, and the focus here is upon inference conditional on the realized data distribution, I will depress the dependence of $\tilde{h}$ on α, defining the notation $h\left(X_i, X_j, U_i, U_j, V_{ij}\right) \overset{def}{\equiv} \tilde{h}\left(\alpha, X_i, X_j, U_i, U_j, V_{ij}\right)$. Consistent with earlier terminology, the function $h : \mathbb{X} \times \mathbb{X} \times [0,1]^3 \to \mathbb{Y}$ will be referred to as a graphon.

Because doing so is convenient for the discussion of causal inference in dyadic settings which follows, (23) makes no presumption of independence between X_i and U_i. Of course we can always write

$$\begin{aligned} Y_{ij}^{*} &= h\left(X_i, X_j, F_{U_1|X_1}\left(U_i|\, X_i\right), F_{U_1|X_1}\left(U_j|\, X_j\right), V_{ij}\right) \\ &\overset{def}{\equiv} h^{*}\left(X_i, X_j, U_i^{*}, U_j^{*}, V_{ij}\right) \end{aligned}$$

with $U_i^{*} = F_{U_1|X_1}\left(U_i|\, X_i\right)$ equal to unit i's rank among all those units of her type. The resulting $\{U_i^{*}\}_{i\geq 1}$ sequence of 0-to-1 uniform random variables is independent of $\{X_i\}_{i\geq 1}$ by construction (cf., Graham et al., 2010).

Depending on the context, it is fine to work with either h or h^{*}, but, as explained below, the former is more useful for making causal arguments; hence I allow for dependence between the *observed* covariate vector X_i and the *unobserved* unit-specific effect U_i in what follows (akin to a correlated random effects panel data analysis). The nuances involved will become clear as we proceed.

Networks generated by (23) exhibit a very particular pattern of dependence across the rows and columns of $[Y_{ij}]_{i,j\in\mathbb{N}}$. Consider, without loss of generality, agents 1, 2, 3 and 4. The outcomes Y_{12} and Y_{34} are independent of one another; the outcomes Y_{12} and Y_{13} are, however, dependent. These two outcomes share agent 1 in common; the value of X_1 and U_1 influences both Y_{12} and Y_{13}, inducing dependence. But conditional on (X_1, X_2, X_3) and (U_1, U_2, U_3), Y_{12} and Y_{13} are independent; if we condition on the observed covariates (X_1, X_2, X_2) alone, however, they remain dependent. Finally Y_{12} and Y_{21} are dependent, this dependence holds even conditional on (X_1, X_2) and (U_1, U_2) because V_{12} and V_{21} may covary.

In words we have independence across dyads sharing no agents in common (exports from Japan to the United States and from Turkey to Germany), dependence across those sharing at least one agent in common (exports from Japan to the United States and from Japan to the United Kingdom), and "even more" dependence across dyads sharing both agents in common (e.g., exports from Japan to the United States and vice-versa).

Models with this type of dependence structure, as already noted, are called conditionally independent dyad (CID) models. The "conditionally independent" terminology reflects the fact that the outcomes Y_{12} and Y_{13}, associated with a pair of dyads sharing one agent in common, can be rendered independent of one another by conditioning on the observed covariates (X_1, X_2, X_2) *as well as* the unobserved latent attributes (U_1, U_2, U_3).

Crane & Towsner (2018), in an extension of the Aldous-Hoover representation result described earlier, show that for any X-exchangeable random array $[Y_{ij}]_{i,j\in\mathbb{N}}$ there exists another array $\left[Y_{ij}^{*}\right]_{i,j\in\mathbb{N}}$ generated according to (23) such that the two arrays have the same distribution:

$$[Y_{ij}]_{i,j\in\mathbb{N}} \overset{D}{=} \left[Y^*_{ij}\right]_{i,j\in\mathbb{N}}. \tag{24}$$

We can therefore use (23) as an 'as if' non-parametric data generating process for $[Y_{ij}]_{i,j\in\mathbb{N}}$; this will facilitate a variety of probabilistic calculations (e.g., computing conditional expectations, variances and, especially, covariances).

Let $i = 1, \ldots, N$ index a simple random sample from the target population. For each of the N sampled units the econometrician observes X_i and for each of the $\binom{N}{2}$ sampled dyads she observes (Y_{ij}, Y_{ji}). From hereon I will assume that Y_{ii} is undefined (normalized to zero for convenience). Adapting what follows to accommodate self-loops is straightforward.

4.2 Composite likelihood

Let $\left\{f_{Y_{12}|X_1,X_2}\left(Y_{12}|\, X_1, X_2; \theta\right) : \theta \in \Theta \subseteq \mathbb{R}^{\dim(\theta)}\right\}$ be a parametric family of distributions for the conditional distribution of Y_{12} given X_1 and X_2. For example, Santos Silva & Tenreyro (2006) model trade from exporter i to importer j given covariates as a Poisson random variable:

$$f_{Y_{12}|X_1,X_2}\left(y_{ij}|\, X_i, X_j; \theta\right) = \exp\left[-\exp\left[W_{ij}'\theta\right]\right] \frac{\left\{\exp\left[W_{ij}'\theta\right]\right\}^{y_{ij}}}{y_{ij}!} \tag{25}$$

with $y_{ij} = 0, 1, 2, \ldots$ and $W_{ij} \overset{def}{\equiv} w\left(X_i, X_j\right)$ a known $J \times 1$ vector of functions of X_i and X_j. As an example, if $X_i = (\ln \mathtt{GDP_i}, \mathtt{LAT_i}, \mathtt{LONG_i})'$, then setting

$$W_{ij} = \begin{pmatrix} \ln \mathtt{GDP_i} \\ \ln \mathtt{GDP_j} \\ \ln\left[(\mathtt{LAT_i} - \mathtt{LAT_j})^2 + (\mathtt{LONG_i} - \mathtt{LONG_j})^2\right]^{1/2} \end{pmatrix}$$

results in a basic gravity trade model specification.[16]

Similar to Russett & Oneal (2001), a researcher might model the conditional probability that country i attacks country j using logistic regression such that

$$f_{Y_{12}|X_1,X_2}\left(y_{ij}|\, X_i, X_j; \theta\right) = \left[F\left(W_{ij}'\theta\right)\right]^{y_{ij}} \left[1 - F\left(W_{ij}'\theta\right)\right]^{1-y_{ij}} \tag{26}$$

with $y_{ij} = 0, 1$ and $F\left(W_{ij}'\theta\right) = \exp\left(W_{ij}'\theta\right) / \left[1 + \exp\left(W_{ij}'\theta\right)\right]$.

An important feature of both (25) and (26) is that they only specify the marginal distribution

[16] In practice distance is measured using the so-called great circle formula; which accounts for the curvature of the Earth's surface.

of Y_{ij} given X_i and X_j. The econometrician is not asserting, for example, that

$$f_{Y_{12},Y_{13}|X_1,X_2,X_3}\left(y_{12},y_{13}|\,X_1,X_2,X_3;\theta\right)=f_{Y_{12}|X_1,X_2}\left(y_{12}|\,X_1,X_2;\theta\right)f_{Y_{12}|X_1,X_2}\left(y_{13}|\,X_1,X_3;\theta\right),$$

since doing so would imply independence of Y_{12} and Y_{13} given covariates; but such dependence is precisely the complication under consideration. Formulating a conditional likelihood for the entire adjacency matrix $\mathbf{Y}\overset{def}{\equiv}[Y_{ij}]_{1\leq i,j\leq N,i\neq j}$ given $\mathbf{X}\overset{def}{\equiv}[X_i]_{1\leq i\leq N}$ would require an explicit specification of the dependence structure across dyads sharing agents in common. In contrast $f_{Y_{12}|X_1,X_2}\left(Y_{12}|\,X_1,X_2;\theta\right)$, which is a model for the marginal distribution of Y_{12} alone, does not require modeling this dependence.

Let $l_{ij}(\theta)=\ln f_{Y_{12}|X_1,X_2}\left(Y_{ij}|\,X_i,X_j;\theta\right)$ and consider the estimator which chooses $\hat{\theta}$ to maximize:

$$\hat{\theta}=\arg\max_{\theta\in\Theta}\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}l_{ij}(\theta). \tag{27}$$

Because its summands are not independent of one another – at least those sharing indices in common are not – (27) does not correspond to a log-likelihood function for $\mathbf{Y}$ given $\mathbf{X}$. Instead it corresponds to what is sometimes called a composite log-likelihood (e.g., Lindsey, 1988; Cox & Reid, 2004). A composite likelihood "is an inference function derived by multiplying a collection of component likelihoods" (Varin et al., 2011, p. 5). Although (27) fails to correctly represent the dependence structure across the elements of the adjacency matrix, if it is based upon a correctly specified marginal density, $\hat{\theta}$ generally will be consistent for θ_0. This follows because the derivative of composite log-likelihood is mean zero at $\theta=\theta_0$ under correct specification of its components. While an appropriately specified composite log-likelihood typically delivers a valid estimating equation, accurate inference is more challenging, since the unmodeled dependence structure in the data does need to be explicitly taken into account at the inference stage.

4.3 Limit distribution

Consider a mean value expansion of the first order condition associated with the maximizer of (27).[17] Such an expansion yields, after some re-arrangement,

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right)=\left[-H_N\left(\bar{\theta}\right)\right]^{+}\sqrt{N}S_N\left(\theta_0\right)$$

[17] A general result on consistency of $\hat{\theta}$ could be constructed by adapting the results of Honoré & Powell (1994).

with $\bar{\theta}$ a mean value between $\hat{\theta}$ and θ_0 which may vary from row to row and the + superscript denoting a Moore-Penrose inverse. Here $S_N(\theta_0)$ is the "score" vector

$$S_N(\theta) = \frac{1}{N}\frac{1}{N-1}\sum_i \sum_{j \neq i} s_{ij}(Z_{ij}, \theta) \tag{28}$$

with $s(Z_{ij}, \theta) = \partial l_{ij}(\theta) / \partial \theta$ for $Z_{ij} = \left(Y_{ij}, X_i', X_j'\right)'$ and $H_N(\theta) = \frac{1}{N}\frac{1}{N-1}\sum_i \sum_{j \neq i} \frac{\partial^2 l_{ij}(\theta)}{\partial \theta \partial \theta'}$. If the Hessian matrix $H_N\left(\bar{\theta}\right)$ converges in probability to the invertible matrix Γ_0, as I will assume, then

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) = -\Gamma_0^{-1}\sqrt{N} S_N(\theta_0) + o_p(1)$$

so that the asymptotic sampling properties of $\sqrt{N}\left(\hat{\theta} - \theta_0\right)$ will be driven by the behavior of $\sqrt{N} S_N(\theta_0)$.

As with the composite log-likelihood criterion function, the summands of $\sqrt{N} S_N(\theta_0)$ are not independent of one another (cf., Cameron & Golotvina, 2005; Fafchamps & Gubert, 2007). A standard central limit theorem cannot be used to demonstrate asymptotic normality of $\sqrt{N} S_N(\theta_0)$. Fortunately $S_N(\theta_0)$, although not a U-Statistic, has a dependence structure similar to one. This insight can be used to derive the limit properties of $\sqrt{N}\left(\hat{\theta} - \theta_0\right)$.

Begin by re-writing $S_N(\theta_0)$ as

$$S_N = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{s_{ij} + s_{ji}}{2}, \tag{29}$$

where $s_{ij} \stackrel{def}{\equiv} s(Z_{ij}, \theta_0)$ and $S_N \stackrel{def}{\equiv} S_N(\theta_0)$. While (29) has the cursory appearance of a U-Statistic it is, in fact not one: Y_{ij}, which enters s_{ij}, varies at the dyad level, hence S_N is not a function of N i.i.d. random variables.

Let $\mathbf{U} = [U_i]_{1 \leq i \leq N}$; the projection of S_N onto the *observed* covariate matrix $\mathbf{X}$ and the *unobserved* vector of unit-specific effects $\mathbf{U}$ equals:

$$V_N \stackrel{def}{\equiv} \mathbb{E}\left[S_N \middle| \mathbf{X}, \mathbf{U}\right] = \binom{N}{2}^{-1} \sum_{i<j} \frac{\bar{s}_{ij} + \bar{s}_{ji}}{2} \tag{30}$$

with $\bar{s}_{ij} \stackrel{def}{\equiv} \bar{s}(X_i, U_i, X_j, U_j)$ and $\bar{s}(X_i, U_i, X_j, U_j) \stackrel{def}{\equiv} \mathbb{E}\left[s(Z_{ij}, \theta_0) \middle| X_i, U_i, X_j, U_j\right]$. The expression to the right of the equality in (30) follows from the Crane & Towsner (2018) representation (23) and independence of V_{ij} from $(\mathbf{X}, \mathbf{U})$.

An important observation is that the projection (30) *is* a U-statistic of order two: specifically it is a summation over all $\binom{N}{2}$ dyads that can be formed from the i.i.d. sample

$\{(X_i, U_i)\}_{1\leq i\leq N}$. Unusually our U-statistic is defined in terms of a combination of both *observed* $\{X_i\}_{1\leq i\leq N}$ and *unobserved* $\{U_i\}_{1\leq i\leq N}$ random variables.

The projection error $T_N = S_N - V_N$ consists of a summation of $\binom{N}{2}$ conditionally uncorrelated summands; hence $\mathbb{V}(T_N) = \binom{N}{2}^{-1}\mathbb{E}\left(\mathbb{V}\left(\frac{s_{12}+s_{21}}{2}\middle| X_1, U_1, X_2, U_2\right)\right) = O\left(N^{-2}\right)$ (as long as $\mathbb{V}\left(\frac{s_{12}+s_{21}}{2}\middle| X_1, U_1, X_2, U_2\right)$ does not change as $N \to \infty$). We also have that T_N and V_N are uncorrelated by construction.

Although we cannot numerically compute V_N – even if θ_0 is known – because the $\{U_i\}_{1\leq i\leq N}$ are unobserved, we can use the theory of U-statistics to characterize its sampling properties as $N \to \infty$. Decomposing V_N into a Hájek projection and a second remainder term yields (e.g., Lehmann, 1999; van der Vaart, 2000):

$$V_N = V_{1N} + V_{2N}$$

where, defining $\bar{s}^e(x,u) = \mathbb{E}\left[\bar{s}(x,u,X_1,U_1)\right]$ and $\bar{s}^a(x,u) = \mathbb{E}\left[\bar{s}(X_1,U_1,x,u)\right]$,

$$V_{1N} = \frac{2}{N}\sum_{i=1}^{N}\left\{\frac{\bar{s}_1^e(X_i,U_i) + \bar{s}_1^a(X_i,U_i)}{2}\right\} \tag{31}$$

$$\begin{aligned} V_{2N} =& \binom{N}{2}^{-1}\sum_{i<j}\left\{\frac{\bar{s}_{ij}+\bar{s}_{ji}}{2}\right. \\ &\left. -\frac{\bar{s}_1^e(X_i,U_i) + \bar{s}_1^a(X_i,U_i)}{2} - \frac{\bar{s}_1^e(X_j,U_j) + \bar{s}_1^a(X_j,U_j)}{2}\right\}. \end{aligned} \tag{32}$$

The superscript 'e' denotes 'ego' since it is the ego unit's attributes which are being held fixed in the average used to compute $\bar{s}^e(x,u)$. Similarly the 'a' denotes 'alter', since it is that unit's attributes which are held fixed when defining $\bar{s}^a(x,u)$. Conveniently V_{1N} is a sum of i.i.d. random variables to which, after scaling by $\sqrt{N}$, a CLT may be applied. Furthermore it can be shown that $\mathbb{V}(V_{2N}) = O\left(N^{-2}\right)$.

Putting these results together yields the asymptotically linear representation

$$\begin{aligned} \sqrt{N}\left(\hat{\theta} - \theta_0\right) =& -\Gamma_0^{-1}\sqrt{N}\left(V_{1N} + V_{2N} + T_N\right) + o_p(1) \\ =& -\Gamma_0^{-1}\sqrt{N}V_{1N} + o_p(1) \\ =& -\Gamma_0^{-1}\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\bar{s}_1^e(X_i,U_i) + \bar{s}_1^a(X_i,U_i)}{2}\right\} + o_p(1), \end{aligned}$$

and hence a limit distribution for $\sqrt{N}\left(\hat{\theta}-\theta_0\right)$ of

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, 4\left(\Gamma_0'\Sigma_1^{-1}\Gamma_0\right)^{-1}\right) \tag{33}$$

where $\Sigma_1 = \mathbb{V}\left(\frac{\bar{s}_1^e(X_1,U_1)+\bar{s}_1^a(X_1,U_1)}{2}\right)$. Although S_N is not a U-statistic, under the assumptions maintained here, its limit distribution coincides with that of V_N (which is a U-statistic).

Before turning to practicalities of variance estimation I will present a useful property of the kernel entering the Hájek Projection, V_{1N} above. By the usual conditional mean zero property of the score function we have that $\mathbb{E}\left[\left.s\left(Z_{12};\theta_0\right)\right|X_1=x_1, X_2=x_2\right]=0$ as long as marginal density of Y_{12} given X_1 and X_2 is correctly specified. This property can be used to show that the averages, $\bar{s}_1^e\left(X_1,U_1\right)$ and $\bar{s}_1^a\left(X_1,U_1\right)$, are also conditionally mean zero. Taking the former we have that

$$\begin{aligned}
\mathbb{E}\left[\left.\bar{s}_1^e\left(X_1,U_1\right)\right|X_1=x_1\right] &= \int\left[\int\int \bar{s}\left(x_1,u_1,x_2,u_2\right) f_{X_1,U_1}\left(x_2,u_2\right)\mathrm{d}x_2\mathrm{d}u_2\right] f_{U_1|X_1}\left(\left.u_1\right|x_1\right)\mathrm{d}u_1 \\
&= \int\left[\int\int \bar{s}\left(x_1,u_1,x_2,u_2\right) f_{U_1|X_1}\left(\left.u_2\right|x_2\right) f_{X_1}\left(x_2\right)\mathrm{d}x_2\mathrm{d}u_2\right] f_{U_1|X_1}\left(\left.u_1\right|x_1\right)\mathrm{d}u_1 \\
&= \int\left[\int\int \bar{s}\left(x_1,u_1,x_2,u_2\right) f_{U_1|X_1}\left(\left.u_1\right|x_1\right)\mathrm{d}u_1 f_{U_1|X_1}\left(\left.u_2\right|x_2\right)\mathrm{d}u_2\right] f_{X_1}\left(x_2\right)\mathrm{d}x_2 \\
&= \int \mathbb{E}\left[\left.\bar{s}\left(X_1,U_1,X_2,U_2\right)\right|X_1=x_1,X_2=x_2\right] f_{X_1}\left(x_2\right)\mathrm{d}x_2 \\
&= \int \mathbb{E}\left[\left.s\left(Z_{12};\theta_0\right)\right|X_1=x_1,X_2=x_2\right] f_{X_1}\left(x_2\right)\mathrm{d}x_2 \\
&= 0
\end{aligned}$$

where the first equality follows from the definition of $\bar{s}_1^e\left(X_1,U_1\right)$, the third from a change in the order of integration, and the second to last from iterated expectations. These calculations imply that

$$\mathbb{E}\left[\left.\bar{s}_1^e\left(X_1,U_1\right)\right|X_1,X_2\right] = \mathbb{E}\left[\left.\bar{s}_1^a\left(X_1,U_1\right)\right|X_1,X_2\right] = 0$$

and hence that $\left[\bar{s}_1^e\left(X_1,U_1\right)+\bar{s}_1^a\left(X_1,U_1\right)\right]/2$ is conditionally mean-zero given X_1 and X_2. This property will be helpful for understanding the asymptotic precision of estimates of various causal parameters introduced below.

4.4 Variance estimation

In order to conduct inference on the components of θ_0, an estimate of the variance of $\sqrt{N}\left(\hat{\theta}-\theta_0\right)$ is required. Although the distribution theory outlined above is novel[18], the history of variance estimation for "dyadic statistics" goes back at least to Holland & Leinhardt (1976). In economics, a variance estimator first proposed by Fafchamps & Gubert (2007), is widely – although not universally – used for dyadic regression analysis. In order to understand extant approaches to variance estimation, as well as to propose new ones, it is helpful to examine the structure of S_N's variance in detail.

The arguments used to derive the limit distribution of $\sqrt{N}\left(\hat{\theta}-\theta_0\right)$ above suggest that it may be insightful to think about the variance of S_N in terms of the ANOVA decomposition

$$\begin{aligned}\mathbb{V}\left(S_N\right) &= \mathbb{V}\left(\mathbb{E}\left[\left.S_N\right|\mathbf{X},\mathbf{U}\right]\right)+\mathbb{E}\left[\mathbb{V}\left(\left.S_N\right|\mathbf{X},\mathbf{U}\right)\right]\\ &= \mathbb{V}\left(V_N\right)+\mathbb{V}\left(T_N\right)\\ &= \mathbb{V}\left(V_{1N}\right)+\mathbb{V}\left(V_{2N}\right)+\mathbb{V}\left(T_N\right),\end{aligned} \tag{34}$$

where the second and third equalities follow from the decomposition for S_N developed in the previous subsection.

Let $p=1,2$ equal the number of agents dyads $\{i_1,i_2\}$ and $\{j_1,j_2\}$ share common and define the matrix Σ_p as

$$\begin{aligned}\Sigma_p \overset{def}{\equiv} \mathbb{C}\Bigg(&\frac{\bar{s}\left(X_{i_1},U_{i_1},X_{i_2},U_{i_2}\right)+\bar{s}\left(X_{i_2},U_{i_2},X_{i_1},U_{i_1}\right)}{2}, \\ &\frac{\bar{s}\left(X_{j_1},U_{j_1},X_{j_2},U_{j_2}\right)'+\bar{s}\left(X_{j_2},U_{j_2},X_{j_1},U_{j_1}\right)'}{2}\Bigg)\\ =\mathbb{C}&\left(\mathbb{E}\left[\left.\frac{s_{i_1i_2}+s_{i_2i_1}}{2}\right|X_{i_1},U_{i_1},X_{i_2},U_{i_2}\right],\mathbb{E}\left[\left.\frac{s_{j_1j_2}+s_{j_2j_1}}{2}\right|X_{j_1},U_{j_1},X_{j_2},U_{j_2}\right]'\right).\end{aligned} \tag{35}$$

When $p=1$ we have

$$\begin{aligned}\Sigma_1 &= \mathbb{C}\left(\mathbb{E}\left[\left.\frac{s_{12}+s_{21}}{2}\right|X_1,U_1,X_2,U_2\right],\mathbb{E}\left[\left.\frac{s_{13}+s_{31}}{2}\right|X_1,U_1,X_3,U_3\right]'\right)\\ &= \mathbb{C}\left(\frac{s_{12}+s_{21}}{2},\frac{s_{13}'+s_{31}'}{2}\right),\end{aligned}$$

with the second equality an implication of conditional independence of $\frac{s_{12}+s_{21}}{2}$ and $\frac{s_{13}+s_{31}}{2}$ given (X_1,X_2,X_3) and (U_1,U_2,U_3). Hence Σ_1 equals the covariance between any pair of

[18] See Tabord-Meehan (2018), Davezies et al. (2019) and, especially, Menzel (2017) for related independent work.

summands in S_N – see equation (28) above – sharing an index in common. There are many such pairs of summands in S_N (actually $\frac{1}{2}N(N-1)(N-2)$ such pairs-of-dyads). It is the preponderance of these non-zero covariances that drives their importance for understanding the sampling distribution of $\sqrt{N}\left(\hat{\theta}-\theta_0\right)$.

In a slight abuse of notation, additionally define the matrix

$$\Sigma_3 \overset{def}{\equiv} \mathbb{E}\left[\mathbb{V}\left(\left.\frac{s_{12}+s_{21}}{2}\right| X_1, U_1, X_2, U_2\right)\right]. \tag{36}$$

Calculations analogous to those use in variance analyses for U-statistics (e.g., Hoeffding, 1948; Lehmann, 1999) yield

$$\mathbb{V}\left(V_{1N}\right) = \frac{4\Sigma_1}{N} \tag{37}$$

$$\mathbb{V}\left(V_{2N}\right) = \frac{2}{N(N-1)}\left(\Sigma_2 - 2\Sigma_1\right) \tag{38}$$

$$\mathbb{V}\left(T_N\right) = \frac{2}{N(N-1)}\Sigma_3, \tag{39}$$

such that, defining the notation $\Omega \overset{def}{\equiv} \mathbb{V}\left(\sqrt{N}S_N\right)$, from (34), (37), (38) and (39):

$$\Omega = 4\Sigma_1 + \frac{2}{N-1}\left(\Sigma_2 + \Sigma_3 - 2\Sigma_1\right). \tag{40}$$

Consistent with the form of the limit distribution given in (33), the variances of V_{2N} and T_N are of smaller order. Although the contribution of these terms to the variance of $\sqrt{N}S_N$ is asymptotically negligible, their contribution for finite N need not be. As alluded to earlier, the appearance of the covariance Σ_1 as the leading term in (40) reflects the large number of non-zero covariance terms that arise when the variance operator is applied to the sum $S_N = \binom{N}{2}^{-1}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{s_{ij}+s_{ji}}{2}$. In practice, especially if $h\left(x_1, x_2, u_1, u_2, v_{12}\right)$ is nearly constant in u_1 and u_2, Σ_1 may be small in magnitude. In such cases it may be that $4\Sigma_1$ and $\frac{2}{N-1}\left(\Sigma_2 + \Sigma_3 - 2\Sigma_1\right)$ are of comparable magnitude for modest N. Using a variance estimator which includes estimates of both these terms may therefore result in tests with better size and power properties (cf., Hoeffding, 1948; Graham et al., 2014; Cattaneo et al., 2014). To construct such an estimator I propose using analog estimates of the terms appearing to the right of the equality in (40).

A benchmark analog variance estimate

A natural analog estimate of Σ_1, the leading variance term, is

$$\hat{\Sigma}_1 = \binom{N}{3}^{-1} \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^{N} \frac{1}{3} \left\{ \left(\frac{\hat{s}_{ij} + \hat{s}_{ji}}{2} \right) \left(\frac{\hat{s}_{ik} + \hat{s}_{ki}}{2} \right)' \right.$$
$$\left. \left(\frac{\hat{s}_{ij} + \hat{s}_{ji}}{2} \right) \left(\frac{\hat{s}_{jk} + \hat{s}_{kj}}{2} \right)' + \left(\frac{\hat{s}_{ik} + \hat{s}_{ki}}{2} \right) \left(\frac{\hat{s}_{jk} + \hat{s}_{kj}}{2} \right)' \right\}, \tag{41}$$

with $\hat{s}_{ij} \overset{def}{\equiv} s\left(Z_{ij}, \hat{\theta}\right)$. Equation (41) is a summation over all $\binom{N}{3} = \frac{1}{6} N(N-1)(N-3)$ triads in the dataset. Each triad ijk can be further divided into three pairs of dyads, $\{ij, ik\}$, $\{ij, jk\}$ and $\{ik, jk\}$, with each such pair sharing exactly one agent in common. Equation (41) corresponds to the sample covariance of $(\hat{s}_{ij} + \hat{s}_{ji})/2$ and $(\hat{s}_{ik} + \hat{s}_{ki})/2$ across these $3\binom{N}{3}$ pairs of dyads.

To construct an estimate of $\mathbb{V}\left(\sqrt{N} S_N\right)$ separate estimates of Σ_2 and Σ_3 are not required, only their sum is needed. Using an ANOVA decomposition we can express this sum as

$$\Sigma_2 + \Sigma_3 = \mathbb{V}\left(\mathbb{E}\left[\left. \frac{s_{12} + s_{21}}{2} \right| X_1, U_1, X_2, U_2 \right] \right) + \mathbb{E}\left[\mathbb{V}\left(\left. \frac{s_{12} + s_{21}}{2} \right| X_1, U_1, X_2, U_2 \right) \right]$$
$$= \mathbb{V}\left(\frac{s_{12} + s_{21}}{2} \right).$$

This suggests the analog estimate

$$\widehat{\Sigma_2 + \Sigma_3} = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{\hat{s}_{ij} + \hat{s}_{ji}}{2} \right) \left(\frac{\hat{s}_{ij} + \hat{s}_{ji}}{2} \right)'. \tag{42}$$

From (40), (41) and (42) we get the variance estimate

$$\hat{\mathbb{V}}\left(\sqrt{N} \left(\hat{\theta} - \theta_0 \right) \right) = \left(\hat{\Gamma}' \hat{\Omega}^{-1} \hat{\Gamma} \right)^{-1} \tag{43}$$

where

$$\hat{\Gamma} = H_N\left(\hat{\theta}\right) \tag{44}$$
$$\hat{\Omega} = 4\hat{\Sigma}_1 + \frac{2}{N-1}\left(\widehat{\Sigma_2 + \Sigma_3} - 2\hat{\Sigma}_1 \right). \tag{45}$$

Fafchamps & Gubert (2007) variance estimate

Just over a decade ago, Fafchamps & Gubert (2007) presented a variance-covariance estimator for $\hat{\theta}$ that they informally argued leads to asymptotically correct inference.[19] They proposed estimating the variance of $\sqrt{N}S_N$ by

$$\hat{\Omega}_{\text{FG}} = \frac{1}{N\left(N-1\right)^2} \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{j_1} \sum_{j_2 \neq j_1} C_{i_1 i_2 j_1 j_2} \hat{s}_{i_1 i_2} \hat{s}'_{j_1 j_2}, \tag{46}$$

where $C_{i_1 i_2 j_1 j_2} = 1$ if $i_1 = j_1$, $i_2 = j_2$, $i_1 = j_2$ or $i_2 = j_1$ and zero otherwise (here the 'or' is inclusive).[20] Equation (46) is a summation across $\binom{N}{2} \times \binom{N}{2}$ "pairs-of-pairs" or pairs of dyads. As noted above, there are $3\binom{N}{3} = \frac{1}{2}N\left(N-1\right)\left(N-2\right)$ unique pairs of dyads sharing one agent in common; but each of these pairs of dyads is counted eight different times in (46). Likewise there are $\binom{N}{2} = \frac{1}{2}N\left(N-1\right)$ pairs of dyads sharing both agents in common (i.e., straight up dyads) and each of these pairs is counted four different times in (46). From this observation we have that

$$\begin{aligned}
\hat{\Omega}_{\text{FG}} &= \frac{1}{N\left(N-1\right)^2}\left[8 \times \frac{1}{2}N\left(N-1\right)\left(N-2\right)\hat{\Sigma}_1 + 4 \times \frac{1}{2}N\left(N-1\right)\widehat{\Sigma_2 + \Sigma_3}\right] \\
&= 4\hat{\Sigma}_1 + \frac{2}{N-1}\left(\widehat{\Sigma_2 + \Sigma_3} - 2\hat{\Sigma}_1\right),
\end{aligned}$$

which exactly coincides with expression (45) above. Not only does $\hat{\Omega}_{\text{FG}}$ include a consistent estimate of the leading term in $\mathbb{V}\left(\sqrt{N}S_N\right)$, but it also includes an estimate of the asymptotically negligible higher order component.

Fafchamps & Gubert (2007) is widely-cited in the context of covariance estimation by empirical researchers, with a STATA implementation for linear and logistic dyadic regression freely available (cf., Cameron & Miller, 2014). Consequently considerable practical experience and Monte Carlo evidence on the properties of standard error estimates based on (46) exists. Among empirical researchers, the consensus is that such standard errors are often much larger than those based on the (possibly erroneous) assumption of independence across dyads.

[19] This estimator has been further explored by Cameron & Miller (2014), Aronow et al. (2017) and Tabord-Meehan (2018).

[20] My definition of $\hat{\Omega}_{\text{FG}}$ actually differs slightly from the one given by Fafchamps & Gubert (2007), due to a finite sample correction term introduced in the latter. Their expression also appears to include a notational inconsistency with N (apparently) denoting both the number of agents as well as the number of dyads (here $n = \frac{1}{2}N\left(N-1\right)$) in different components of the expression. Once these typos are corrected (46) agrees with their expression up to a finite sample correction.

Snijders & Borgatti (1999) jackknife variance estimate

Snijders & Borgatti (1999), inspired by the prior work of Frank & Snijders (1994), suggest[21] a jackknife variance estimate for $\mathbb{V}\left(\sqrt{N}S_N\right)$ of

$$\hat{\Omega}_{\text{SB}} = \left(\frac{N-2}{2}\right)\sum_i \left[S_{N,-i}\left(\hat{\theta}\right) - S_N\left(\hat{\theta}\right)\right]\left[S_{N,-i}\left(\hat{\theta}\right) - S_N\left(\hat{\theta}\right)\right]', \tag{47}$$

where $S_{N,-i}(\theta)$ is the average of the dyadic scores over the $\binom{N-1}{2}$ dyads which do not include agent i:

$$S_{N,-i}(\theta) \overset{def}{\equiv} \binom{N-1}{2}^{-1}\left[\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}\frac{s(Z_{jk};\theta)+s(Z_{kj};\theta)}{2} - \sum_{l\neq i}\frac{s(Z_{il};\theta)+s(Z_{li};\theta)}{2}\right].$$

The Snijders & Borgatti (1999) proposal, the basis of which they acknowledge was primarily intuitive, does not provide a consistent estimate of $\mathbb{V}\left(\sqrt{N}S_N\right)$, but, as I will now show, a slight modification of their proposal does.

With some manipulation we can write, defining $\hat{\bar{s}}_{1i}(\theta) \overset{def}{\equiv} \frac{1}{N-1}\sum_{j\neq i}\frac{s(Z_{ij};\theta)+s(Z_{ji};\theta)}{2}$ (in a slight abuse of notation),

$$\begin{aligned} S_{N,-i}(\theta) - S_N(\theta) &= \binom{N-1}{2}^{-1}\left[\binom{N}{2}S_N(\theta) - (N-1)\hat{\bar{s}}_{1i}(\theta)\right] - S_N(\theta) \\ &= -\frac{2}{N-2}\left[\hat{\bar{s}}_{1i}(\theta) - S_N(\theta)\right]. \end{aligned} \tag{48}$$

Observe that $\hat{\bar{s}}_{1i} \overset{def}{\equiv} \hat{\bar{s}}_{1i}\left(\hat{\theta}\right)$ would be the usual estimate of the the i^{th} summand in the Hájek projection given in (31) above (see, for example, Callaert & Veraverbeke (1981) or Cattaneo et al. (2014) and the references therein). Indeed, on the basis of the limit theory outlined above, a natural estimate of Σ_1 would be

$$\tilde{\Sigma}_1 \overset{def}{\equiv} \frac{1}{N}\sum_{i=1}^{N}\hat{\bar{s}}_{1i}\hat{\bar{s}}_{1i}'. \tag{49}$$

[21] They actually propose a jackknife estimate for the variance of S_N. I have multiplied their expression by N to get the corresponding expression for the variance of $\sqrt{N}S_N$ (see Equation (2) of Snijders & Borgatti (1999)).

After some tedious manipulation it is possible to show that

$$\begin{aligned}\tilde{\Sigma}_1 &= \hat{\Sigma}_1 + \frac{\widehat{\Sigma_2 + \Sigma_3} - \hat{\Sigma}_1}{N-1} \\ &= \hat{\Sigma}_1 + O_p\left(N^{-1}\right)\end{aligned} \tag{50}$$

with $\hat{\Sigma}_1$ and $\widehat{\Sigma_2 + \Sigma_3}$ as defined in (41) and (42) above.

Equations (48), (50) and the observation that $S_N\left(\hat{\theta}\right) = 0$ implies that the jackknife variance estimate

$$\begin{aligned}\hat{\Omega}_{\text{JK}} &\overset{def}{\equiv} \frac{(N-2)^2}{N} \sum_i \left[S_{N,-i}\left(\hat{\theta}\right) - S_N\left(\hat{\theta}\right)\right]\left[S_{N,-i}\left(\hat{\theta}\right) - S_N\left(\hat{\theta}\right)\right]' \\ &= 4\tilde{\Sigma}_1 \\ &= 4\hat{\Sigma}_1 + \frac{4\left(\widehat{\Sigma_2 + \Sigma_3} - \hat{\Sigma}_1\right)}{N-1},\end{aligned} \tag{51}$$

provides a consistent estimate of the asymptotic variance $\sqrt{N}S_N$.

Furthermore, inspired by Efron & Stein (1981) and, especially, Cattaneo et al. (2014), we can bias correct (51):

$$\hat{\Omega}_{\text{JK-BC}} \overset{def}{\equiv} \hat{\Omega}_{\text{JK}} - \frac{2}{N-1}\left(\widehat{\Sigma_2 + \Sigma_3}\right) = \hat{\Omega}_{\text{FG}} \tag{52}$$

with $\widehat{\Sigma_2 + \Sigma_3}$ as defined by (42) and the equality an implication of (50). Equation (52) implies that the Fafchamps & Gubert (2007) variance estimator, or equivalently the analog estimator proposed above, coincides with a bias corrected jackknife variance estimate. This is awesome.

4.5 Bootstrap inference

Relative to analytic variance estimation, the theory of the bootstrap for dyadic regression is comparatively less well-understood. Rewriting our dyadic regression coefficient estimate in pseudo-U-Process form yields

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left\{\frac{l_{ij}(\theta) + l_{ji}(\theta)}{2}\right\}.$$

Next let $\left\{V_i^b\right\}_{i=1}^N$ be a sequence of i.i.d. mean one random weights independent of the data.

One such sequence is drawn for each of $b = 1, \ldots, B$ bootstrap replications. In the b^{th} such replication we compute

$$\hat{\theta}_b = \arg\max_{\theta \in \Theta} \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} V_i^b V_j^b \left\{ \frac{l_{ij}(\theta) + l_{ji}(\theta)}{2} \right\}.$$

The bootstrap distribution $\left\{ \hat{\theta}_b \right\}_{b=1}^{B}$ can then be used to approximate the sampling distribution of $\hat{\theta}$. Letting V_i^b be an exponential random variable with rate parameter 1 results in a Bayesian bootstrap which is, of course, preferred. The above algorithm was proposed in the context of U-statistics by Janssen (1994). If we let V_i^b equal the number of times agent i is sampled from the set $\{1, \ldots, N\}$ across N draws with replacement, we get the proposal of Davezies et al. (2019), who show – under certain assumptions – validity for the dyadic regression case considered here.

Snijders & Borgatti (1999) proposed a bootstrap procedure for jointly exchangeable random arrays which is very close to the proposal of Davezies et al. (2019). As with their jackknife variance estimator, their development was intuitive and informal. For simplicity consider the application of their proposal for inference on the dyadic mean $\bar{Y} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left\{ \frac{Y_{ij}+Y_{ji}}{2} \right\}$. Let $i_1^b, \ldots, i_N^b$ be N indices drawn uniformly at random (with replacement) from $\{1, \ldots, N\}$. Let $\mathbf{Y}^b$ be the adjacency matrix induced by $\left\{ i_1^b, \ldots, i_N^b \right\}$. If agent j is sampled twice, say $i_1^b = j$ and $i_2^b = j$ we face the practical problem that the outcome $Y_{i_1^b i_2^b} = Y_{jj}$ is undefined. Snijders & Borgatti (1999) propose filling in such cells with independent random draws from $\{Y_{12}, Y_{21}, \ldots, Y_{N-1N}, Y_{NN-1}\}$; they note that the expected fraction of bootstrap dyads constructed from a single underlying agent in the original sample will be vanishingly small as $N \to \infty$ (suggesting that this problem may not matter for asymptotic properties). Snjiders and Borgatti's (1999) algorithm essentially coincides with the pigeon-hole bootstrap proposed by Owen (2007) for separately exchangeable random arrays (in which the problem of "zero diagonals" does not arise).

A final bootstrap procedure is proposed by Menzel (2017). He is particularly concerned with formulating a procedure that adaptively handles the possibility that there is, in fact, no dyadic correlation in the data (i.e., $\Sigma_1 = 0$). Degeneracy of this type occurs, in our regression setting, when the graphon $h(x_1, x_2, u_1, u_2, v_{12})$ is constant in both u_1 and u_2 (but also in more exotic situations where there is dyadic dependence in higher order moments, but no correlation). The arguments in Menzel (2017) suggest that the weighted bootstrap of Janssen (1994) and Davezies et al. (2019) will be inconsistent under degeneracy.

Menzel (2017) proposes several different bootstraps; what I sketch here is a simplified version of his 'BS-N' procedure (adapted to the dyadic regression case). Let

$$\hat{s}_i^e = \frac{1}{N-1}\sum_{j=1}^{N}\hat{s}_{ij},\;\; \hat{s}_j^a = \frac{1}{N-1}\sum_{i=1}^{N}\hat{s}_{ij}$$

be estimates of the average dyadic score for 'ego' i and 'alter' j. Let

$$\hat{e}_{ij} = \hat{s}_{ij} - \hat{s}_i^e - \hat{s}_j^a$$

equal the residual for $\hat{s}_{ij}$ after subtracting off these ego and alter means. Menzel (2017) actually suggests subtracting off rescaled versions of $\hat{s}_i^e$ and $\hat{s}_j^a$ when forming $\hat{e}_{ij}$. Rescaling improves the accuracy of his procedure when dyadic correlation is, in fact, absent. I omit this detail since describing it requires introducing substantial additional notation. The stylized version sketched here will be conservative under degeneracy (similar to the pigeonhole bootstrap).

Let $\left\{V_i^b\right\}_{i=1}^N$ be a sequence of i.i.d. mean *zero* random weights with unit variance (and unit third moment as well). Let $i_1^b, \ldots, i_N^b$ be N indices drawn uniformly at random (with replacement) from $\{1, \ldots, N\}$. For all $\binom{N}{2}$ dyads induced by the b^{th} such bootstrap sample construct the scores

$$\hat{s}_{i_j^b i_k^b} = \hat{s}_{i_j^b}^e + \hat{s}_{i_k^b}^a + V_{i_j^b}^b V_{i_k^b}^b \hat{e}_{i_j^b i_k^b},\;\; j = 1, \ldots, N-1 \;\&\; k = j+1, \ldots, N$$

and compute their mean

$$\hat{S}_N^b = \frac{2}{N(N-1)}\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}\left(\frac{\hat{s}_{i_j^b i_k^b} + \hat{s}_{i_k^b i_j^b}}{2}\right).$$

The variance of $\sqrt{N}\hat{S}_N^b$ across the $b = 1, \ldots, B$ bootstrap replications can be used to construct an estimate of $\Omega = \mathbb{V}\left(\sqrt{N}S_N\right)$. Menzel's (2017) preferred procedures involve an additional "model selection" step, not described here, as well as pivotizing using $\hat{\Omega}_{\mathrm{FG}} = \hat{\Omega}_{\mathrm{JK-BC}}$.

4.6 Further reading and open questions

A special case of the Fafchamps & Gubert (2007) variance estimator was first proposed by Holland & Leinhardt (1976) in the context of inference on network density; the equivalent of the dyadic mean $\mu_{Y_{12}} = \mathbb{E}[Y_{12}]$ here (estimated by $\bar{Y} = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i} Y_{ij}$). The Holland & Leinhardt (1976) variance estimate was used with some regularity in empirical social network analysis in the 1980s and 1990s (cf., Wasserman & Faust, 1994). The reference distribution was assumed to be normal, but no proof for this was available. Bickel et al. (2011) appear to be the first to have shown asymptotic normality of $\sqrt{N}\left(\bar{Y}-\mu_{Y_{12}}\right)$ under dyadic dependence. The double projection argument used to produce the $S_N = V_{1N} + V_{2N} + T_N$ decomposition used above is implicit in their work. A similar decomposition was used by Graham (2017) to show asymptotic normality of the Tetrad Logit estimator, which is described further below. The bootstrap procedure of Menzel (2017) is also based upon this decomposition. Tabord-Meehan (2018) demonstrates asymptotic normality of dyadic regression coefficients estimated by ordinary least squares. His method of proof is very different from the argument outlined here.

Cameron & Miller (2014), Aronow et al. (2017) and Tabord-Meehan (2018) provide further results on variance estimation for dyadic regression; each building upon the proposal of Fafchamps & Gubert (2007).

Menzel (2017) and Davezies et al. (2019) provide large sample theory in some generality – including for cases not covered here. Both these papers provide formal results on inference using the bootstrap as well. The presentation here is based upon Graham (2018a), a revised and expanded version of which appears as a chapter in Graham & de Paula (2020).

When dyadic correlation is weak limit theory can be non-standard. Menzel (2017) provides examples and discussion. Related issues arise in Graham et al. (2019), who study nonparametric density and regression estimation with dyadic data. Developing inference procedures with good properties across a range of (dyadic) data generating processes remains largely open.

Open research problems include extending the material summarized here to accommodate regressor endogeneity and settings where the number of regressors is comparable to, or even exceeds, the number of agents (or dyads).

5 Policy analysis

One motivation for Tinbergen's (1962) dyadic regression analysis was to evaluate the effect of preferential trade agreements on export flows. Rose (2004) explores the related question of whether membership in the General Agreement on Trade and Tariffs (GATT) or its successor, the World Trade Organization (WTO), promoted trade (see also Rose (2005)). Baldwin & Taglioni (2007) and Santos Silva & Tenreyro (2010) use gravity models to assess whether common currency zones, such as the Eurozone, promote trade. As with conventional regression analysis, a desire to assess different programs or policies underlies many dyadic regression analyses.[22]

While the logic and mechanics of program evaluation are well understood in the context of single agent models (cf., Heckman & Vytlacil, 2007; Imbens & Wooldridge, 2009), a comparable framework for causal reasoning is not, to my knowledge, available in the dyadic setting considered here. In this section I make a start at formulating such a framework. In doing so I will attempt to follow the notation and language of the standard single agent causal inference framework reviewed in, for example, Imbens & Wooldridge (2009). What follows are some initial ideas and results; much work remains to be done.

5.1 Dyadic potential response

Let $W_i \in \mathbb{W} = \{w_1, \ldots, w_K\}$ and $X_i \in \mathbb{X} = \{x_1, \ldots, x_L\}$ be a finite set of *ego* and *alter* policies. For example $\mathbb{W}$ might enumerate different export promotion policies (e.g., tax subsidies or preferential credit schemes for exporting firms), while $\mathbb{X}$ might enumerate different combinations of protectionist policies (e.g., tariff levels). The goal is to understand how different counterfactual combinations of ego and alter policy pairs map into (distributions of) outcomes.

I begin with an assumption about the form of the potential response function for (directed) dyad ij.

Assumption 1. (Dyadic Potential Response Function) *For any ego-alter pair $i, j \in \mathbb{N}$ with $i \neq j$, the potential (directed) outcome associated with adopting the pair of policies $W_i = w$ and $X_j = x$ is given by*

$$Y_{ij}(w, x) = h(w, x, A_i, B_j, V_{ij}), \; x \in \mathbb{X}, \; w \in \mathbb{W} \tag{53}$$

with $\{(A_i, B_i)\}_{i \in \mathbb{N}}$ and $\{(V_{ij}, V_{ji})\}_{i,j \in \mathbb{N}, i<j}$ both i.i.d. sequences additionally independent of each other and $h : \mathbb{W} \times \mathbb{X} \times \mathbb{A} \times \mathbb{B} \times \mathbb{V} \to \mathbb{Y}$ a measurable function.

The ego and alter effects, respectively A_i and B_i, induce dependence across any pair of potential outcomes whose corresponding dyads share at least one agent in common. This implies a type structured "interference" between units, and hence a violation of SUTVA (cf., Rosenbaum, 2007).

[22]Other examples of recent attempts to reason about causal questions with dyadic data include Schwartz and Sommers' (2014) and Goodman's (2017) analyses of whether Medcaid expansion states experienced in-migration from neighboring states which chose to forgo the Affordable Care Act's expansion of Medicaid and Mayda's (2010) and Oretega and Peri's (2013) studies of the relationship between immigration entry tightness and cross-country migration.

Since assignment to treatment is at the 'ego' or 'alter' level, setting $X_i = x$ and $W_j = w$ shapes not just the realized outcome for dyad ij, but also those of all other dyads which include either agent i or agent j. It is because of its implications for dependence across the rows and columns of $[Y_{ij}(w,x)]_{i,j\in\mathbb{N}}$ that I label Assumption 1 an "assumption". More than just notation is involved.

It is possible that Assumption 1 could be derived from a more primitive exchangeability type restriction; for example by viewing $[Y_{ij}(w,x)]_{i,j\in\mathbb{N}}$ as a jointly exchangeable random array and appealing to the Aldous-Hoover Theorem. There may be some deep subtleties involved in such an approach, so I prefer to maintain (53) as an explicit assumption in this initial exploration.

I could have also written $Y_{ij}(w,x) = h(w,x,(1,0)U_i,(0,1)U_j,V_{ij}) = h^*(w,x,U_i,U_j,V_{ij})$ with $U_i = (A_i,B_i)'$. Explicitly separating out an 'ego' and 'alter' effect, however, is conceptually useful and also facilitates, as will be demonstrated by example below, parametric modeling.

In some cases of interest the support of the ego and alter policies will coincide (i.e., $\mathbb{W} = \mathbb{X}$). Following Santos Silva & Tenreyro (2010), for example, both X_i and W_j might be indicators for Eurozone membership. This example implies the additional restriction that $X_i = W_i$ for all $i \in \mathbb{N}$, since a country belongs to the Eurozone in both their exporter (ego) and importer (alter) role. These special cases can be deduced from the more general results which follow.

5.2 Average structural function (ASF)

Dyad-level treatment effects are defined in the usual way. The effect on ij's outcome of adopting policy pair (w',x') vs. (w,x) is

$$Y_{ij}(w',x') - Y_{ij}(w,x).$$

As in the standard case, identification of such effects at the dyad-level is infeasible. This is because the econometrician only observes the outcome associated with the policy pair actually adopted. Specifically, for each of N randomly sampled units she observes the assigned or chosen ego and alter policies, $\{(W_i,X_i)\}_{i=1}^N$ and the $N(N-1)$ realized (directed) outcomes $\{(Y_{ij},Y_{ji})\}_{i<j}$, where

$$Y_{ij} \overset{def}{\equiv} Y_{ij}(W_i,X_j) \tag{54}$$

equals (directed) dyad ij's realized outcome. No counterfactual outcomes are observed.

Although dyad-level treatment effects are not identified, averages of such effects over agents and/or dyads are (under certain assumptions). Here I will focus on identifying average treatment effect (ATE) type parameters. Consider the following thought experiment: (i) draw an ego unit at random from the target population and exogenously assign it policy $W_i = w$, (ii) independently draw an alter unit at random and assign it policy $X_j = x$. The (ex ante) expected outcome associated with this directed dyad, so configured, is

$$
\begin{aligned}
m^{\text{ASF}}(w,x) &\overset{def}{\equiv} \mathbb{E}\left[Y_{12}(w,x)\right] \\
&= \int\int\int h(w,x,a,b,v)\, f_{A_1}(a)\, f_{B_2}(b)\, f_{V_{12}}(v)\, \mathrm{d}a\mathrm{d}b\mathrm{d}v \\
&\overset{def}{\equiv} \int\int\int \bar{h}(w,x,a,b)\, f_{A_1}(a)\, f_{B_2}(b)\, \mathrm{d}a\mathrm{d}b,
\end{aligned} \tag{55}
$$

where the second '$\overset{def}{\equiv}$' in (55) follows from defining $\bar{h}(w,x,a,b) \overset{def}{\equiv} \mathbb{E}\left[h(w,x,a,b,V_{12})\right] \overset{def}{\equiv} \bar{Y}_{ij}(w,x)$. Note also that $\mathbb{E}\left[h(w,x,a,b,V_{12})\middle|\, A_1 = a, B_2 = b\right] = \mathbb{E}\left[h(w,x,a,b,V_{12})\right]$ by independence of A_1, B_2 and V_{12} (Assumption 1).

Differences of the form $m^{\text{ASF}}(w',x') - m^{\text{ASF}}(w,x)$ measure the expected effects of different combinations of policies on the directed dyadic outcome. If $W_i \in \{0,1\}$ and $X_i \in \{0,1\}$ are both binary indicators for GATT/WHO membership, as in Rose (2004), then the contrast

$$
m^{\text{ASF}}(1,1) - m^{\text{ASF}}(0,0) \tag{56}
$$

gives differences in export flows between a random pair of countries in the GATT/WHO vs. non-GATT/WHO states of the world. This is an average treatment effect (ATE) type parameter, adapted to the dyadic setting.

The dyadic setting also raises new questions. For example the double difference

$$
m^{\text{ASF}}(1,1) - m^{\text{ASF}}(0,1) - \left[m^{\text{ASF}}(1,0) - m^{\text{ASF}}(0,0)\right] \tag{57}
$$

measures complementarity in a binary policy/treatment across the two agents in the dyad.

Other estimands beside the ASF may be of interest. The difference of sample means

$$
\frac{1}{N-1}\sum_{j \neq i}\left[Y_{ij}(1,X_j) - Y_{ij}(0,X_j)\right]
$$

measures the average effect – for unit i alone – of adopting ego policy $W_i = 1$ versus $W_i = 0$; the average is over the status quo distribution of alter polices. Additionally averaging over ego units gives

$$\frac{1}{N}\frac{1}{N-1}\sum_{i}\sum_{j\neq i}\left[Y_{ij}\left(1,X_{j}\right)-Y_{ij}\left(0,X_{j}\right)\right].$$

This equals the average effect, across all units in the sample, of adopting ego policy $W_i = 1$ versus $W_i = 0$, again given the status quo distribution of alter policies. The population counterparts of these two sample averages may also be of interest.

For the purposes of illustration, assume that $\mathbb{W} = \mathbb{X} = \{0, 1\}$. A parametric form for $Y_{ij}(w, x)$ that will be helpful for both understanding extant empirical work and interpreting some of the assumptions which follow is:

$$Y_{ij}\left(w,x\right)=\alpha+w\beta+x\gamma+wx\delta+A_{i}+B_{j}+V_{ij}. \tag{58}$$

Response (58) implies that treatment effects are constant across units, for example,

$$Y_{ij}\left(1,0\right)-Y_{ij}\left(0,0\right)=\beta,$$

which is constant in $i \in \mathbb{N}$. Under (58) we also have estimand (56) equaling $\beta + \gamma + \delta$ and (57) equal to δ.

The average structural function (ASF) estimand is a leading case and will be emphasized here. However, as I hope the brief sketch above makes clear, other estimands merit exploration and, I conjecture, will involve interesting identification, estimation and inference issues.

5.3 Identification under exogeneity

In order to identify the ASF I will assert the existence of the observable proxy variables, R_i and S_i, respectively for the ego and alter effects A_i and B_i. These proxy variables will satisfy two key restrictions, the first of which is:

Assumption 2. (Redundancy) *For $R_i \in \mathcal{R} \subseteq \mathbb{R}^{\dim(R)}$ a proxy variable for A_i, and $S_i \in \mathcal{S} \subseteq \mathbb{R}^{\dim(S)}$ a proxy variable for B_i, we have that*

$$\mathbb{E}\left[\left.Y_{ij}\left(w,x\right)\right|W_{i},X_{j},A_{i},B_{j},R_{i},S_{j}\right]=\mathbb{E}\left[\left.Y_{ij}\left(w,x\right)\right|W_{i},X_{j},A_{i},B_{j}\right],$$

for any $w \in \mathbb{W}$ and $x \in \mathbb{X}$.

Assumption 2 is a redundancy assumption of the type introduced by Wooldridge (2005); it simply asserts that R_i and S_j have no predictive power (in the conditional mean sense) for the dyadic potential outcome $Y_{ij}(w,x)$ conditional on the latent ego and alter attributes A_i and B_j. Adapting Wooldridge's (2005) example, it asserts that ego and alter Armed Forces Qualification Test (AFQT) scores, R_i and S_j, do not predict Y_{ij} conditional on the unobserved cognitive abilities, A_i and B_j. Assumption 2 is a weak requirement since we are free to conceptualize the latent attributes, A_i and B_j, such that R_i and S_j are clearly redundant.

Assumption 3. (Strict Exogeneity) *The ij ego-alter treatment assignment (W_i, X_j) is independent of V_{ij} conditional on the latent ego A_i and alter B_j effects:*

$$V_{ij} \perp (W_i, X_j)|\, A_i = a, B_i = b,\, a \in \mathbb{A},\, b \in \mathbb{B}. \tag{59}$$

While conditional independence assumptions feature prominently in the causal inference literature, Assumption 3, which involves conditioning on *unobservables*, has no clear analog in the standard program evaluation model. The closest analog of this assumption I can think of is Chamberlain's (1984) definition of strict exogeneity of a time-varying regressor conditional on a latent (time-invariant) unit-specific effect in the context of panel data. To see the parallel return to parametric potential response function (58) and note that (54) and (59) imply that

$$\mathbb{E}\left[Y_{ij}|\, W_i, X_j, A_i, B_j\right] = \alpha + W_i\beta + X_j\gamma + W_iX_j\delta + A_i + B_j \tag{60}$$

since Assumption 3 gives $\mathbb{E}\left[V_{ij}|\, W_i, X_j, A_i, B_j\right] = \mathbb{E}\left[V_{ij}|\, A_i, B_j\right]$ and $\mathbb{E}\left[V_{ij}|\, A_i, B_j\right] = \mathbb{E}\left[V_{ij}\right]$ by independence of $\{(A_i, B_i)\}_{i=1}^{N}$ and $\{(V_{ij}, V_{ji})\}_{i<j}$ (setting $\mathbb{E}\left[V_{ij}\right] = 0$ is a normalization). Equation (60) looks a lot like the definition of strict exogeneity in Chamberlain (1984, Equation 1.2 on p. 1248). Equation (60) implies, for example, that

$$\mathbb{E}\left[Y_{ij} - Y_{il} - (Y_{kj} - Y_{kl})|\, W_i, X_j, A_i, B_j\right] = (W_i - W_k)\,(X_j - X_l)\,\delta,$$

such that "within-tetrad" variation identifies δ. Similar to how within-group variation in a strictly exogenous regressor identifies its corresponding coefficient in the panel context.

Under Assumption 3 we have the density factorization

$$\begin{aligned} f_{V_{12},A_1,W_1,B_2,X_2}\left(v_{12}, a_1, w_1, b_2, x_2\right) =& f_{V_{12}|A_1,W_1,B_2,X_2}\left(v_{12}|\, a_1, w_1, b_2, x_2\right) \\ & \times f_{A_1,W_1}\left(a_1, w_1\right) f_{B_2,X_2}\left(b_2, x_2\right) \\ =& f_{V_{12}|A_1,B_2}\left(v_{12}|\, a_1, b_2\right) f_{A_1,W_1}\left(a_1, w_1\right) f_{B_2,X_2}\left(b_2, x_2\right) \\ =& f_{V_{12}}\left(v_{12}\right) f_{A_1,W_1}\left(a_1, w_1\right) f_{B_2,X_2}\left(b_2, x_2\right) \end{aligned}$$

with the first equality an implication of units 1 and 2 being independent random draws, the second equality following from Assumption 3, and the third from independence of $\{(A_i, B_i)\}_{i=1}^{N}$ and $\{(V_{ij}, V_{ji})\}_{i<j}$ (i.e., Assumption 1).

This factorization clarifies that the effect of Assumption 3 is to ensure that all "endogeneity" in treatment choice is reflected in dependence between W_i and A_i and/or B_j and X_j. Conditional on these two latent variables, variation in treatment is "idiosyncratic" or exogenous.

To deal with dependence between W_i and A_i, and B_j and X_j, I make a familiar selection of observables type assumption.

Assumption 4. (Conditional Independence) *An ego's (alter's) treatment choice varies independently of their latent effect A_i (B_j) given the observed proxy R_i (S_j):*

$$A_i \perp W_i | \, R_i = r, \; r \in \mathcal{R} \subseteq \mathbb{R}^{\dim(R)} \tag{61}$$

$$B_i \perp X_i | \, S_i = s, \; s \in \mathcal{S} \subseteq \mathbb{R}^{\dim(S)}. \tag{62}$$

Assumption 4 is a standard one in the context of single agent program evaluation problems, asserting – for example – that A_i and W_i vary independently within subpopulations homogenous in the proxy variable R_i. Extensive discussions of selection-on-observables type assumptions like these, including assessments of their appropriateness in different settings of interest to empirical researchers, can be found in Blundell & Powell (2003), Heckman & Vytlacil (2007), Imbens & Wooldridge (2009) and Imbens & Rubin (2015). Their invocation here can raise new issues, but, for the most part familiar approaches to reasoning apply; see Graham et al. (2018) for a related discussion.

Assumptions 1 to 4, plus an additional support condition described below, are sufficient to show identification of the ASF. To develop the argument first let

$$q\left(w, x, r, s\right) = \mathbb{E}\left[Y_{ij} | \, W_i = w, X_j = x, R_i = s, S_j = s\right] \tag{63}$$

be the dyadic proxy variable regression (PVR). Under Assumptions 1 through 4 the PVR relates to $\bar{Y}_{12}(w,x) = \bar{h}(w,x,A_1,B_2)$ as follows:

$$
\begin{aligned}
q(w,x,r,s) =& \mathbb{E}\left[h(W_i,X_j,A_i,B_j,V_{ij}) \middle| W_i = w, X_j = x, R_i = r, S_j = s\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[h(W_i,X_j,A_i,B_j,V_{ij}) \middle| W_i = w, X_j = x, A_i, B_j, R_i = r, S_j = s\right] \right.\\
& \left. | W_i = w, X_j = x, R_i = r, S_j = s\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[h(W_i,X_j,A_i,B_j,V_{ij}) \middle| W_i = w, X_j = x, A_i, B_j\right] \right.\\
& \left. | W_i = w, X_j = x, R_i = r, S_j = s\right] \\
=& \mathbb{E}\left[\bar{h}(w,x,A_i,B_j) \middle| W_i = w, X_j = x, R_i = r, S_j = s\right] \\
=& \int_a \int_b \bar{h}(w,x,a,b) f_{A|R}(a|r) f_{B|S}(b|s) \mathrm{d}a\mathrm{d}b \\
=& \mathbb{E}\left[\bar{Y}_{12}(w,x) \middle| R_1 = r, S_2 = s\right]. \qquad (64)
\end{aligned}
$$

where the first equality follows from Assumption 1 and equation (54), the second from iterated expectations, the third from the redundancy condition (Assumption 2), the fourth from Assumption 3, independence of $\{(A_i, B_i)\}_{i=1}^N$ and $\{(V_{ij}, V_{ji})\}_{i<j}$ and the definition of $\bar{h}$, and the fifth from selection on observables (Assumption 4).

Equation (64) gives the identification result

$$
\begin{aligned}
\mathbb{E}_R\left[\mathbb{E}_S\left[q(w,x,R_i,S_j)\right]\right] =& \int_r \int_s \left[\int_a \int_b \bar{h}(w,x,a,b) f_{A|R}(a|r) f_{B|S}(b|s) \mathrm{d}a\mathrm{d}b\right] \qquad (65)\\
& \times f_R(r) f_S(s) \mathrm{d}r\mathrm{d}s \\
=& \int_a \int_b \bar{h}(w,x,a,b) f_A(a) f_B(b) \mathrm{d}a\mathrm{d}b \\
=& \mathbb{E}\left[\bar{Y}_{12}(w,x)\right] \\
=& m^{\mathrm{ASF}}(w,x).
\end{aligned}
$$

Since $q(w,x,r,s)$ is only identified at those points where $f_{R|W}(r|x) f_{S|X}(s|x) > 0$, while the integration in (65) is over $\mathcal{R} \times \mathcal{S}$, we require a formal support condition:

$$
\mathbb{S}(w,x) \stackrel{def}{\equiv} \left\{r, s \,:\, f_{R|W}(r|w) f_{S|X}(s|x) > 0\right\} = \mathcal{R} \times \mathcal{S}. \qquad (66)
$$

When W_i and X_j are discretely-valued, with a finite number of support points, as assumed here, (66) can be expressed in a form similar to the overlap condition familiar from the program evaluation literature (e.g., Heckman et al., 1997; Imbens & Wooldridge, 2009).

Assumption 5. (Overlap) *For (w,x) the ego-alter treatment combination of interest*

$$p_w(r)\,p_x(s) \geq \kappa > 0 \text{ for all } (r,s) \in \mathcal{R}\times\mathcal{S}$$

where $p_w(r) \overset{def}{\equiv} \Pr(W_i = w|\, R_i = r)$ *and* $p_x(s) \overset{def}{\equiv} \Pr(X_i = x|\, S_i = s)$.

We have shown.

Theorem 2. *Under Assumptions 1 through 5 the ASF is identified by*

$$m^{\mathrm{ASF}}(w,x) = \int\int q(w,x,r,s)\, f_R(r)\, f_S(s)\,\mathrm{d}r\mathrm{d}s. \tag{67}$$

Theorem 2 shows that the ASF is identified by double marginal integration over the dyadic proxy variable regression function. Double marginal integration also features in Graham et al. (2018), in the context of identifying an average match function (AMF), and Brown & Newey (1998), in their discussion of efficient expectation estimation under independence restrictions. However the random array structure present here is absent in both these examples, which accounts for many of the differences in underlying arguments.

5.4 Estimation of the average structural function

Let $q(w,x,r,s;\gamma)$ be a (flexibly) parametric model for the dyadic proxy variable regression function. For example, if the outcome of interest is export flows, we might specify that

$$q(w,x,r,s;\gamma) = \exp\left(t(Q_i)'\gamma\right),$$

with $Q_i = (W_i', X_i', R_i', S_i')'$ and $t(Q_i)$ a finite (and pre-specified) set of basis functions (preferably including interactions of terms in the treatment variables – W, X – and proxy variables – R, S). We can estimate γ use the Poisson dyadic regression estimator described in Section 4. Proceeding in this way delivers an asymptotically linear representation for $\sqrt{N}(\hat{\gamma}-\gamma_0)$ of

$$\sqrt{N}(\hat{\gamma}-\gamma_0) = -\Gamma_0^{-1}\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\bar{s}_1^e(Q_i,U_i;\gamma_0)+\bar{s}_1^a(Q_i,U_i;\gamma_0)}{2}\right\} + o_p(1) \tag{68}$$

with $U_i = (A_i, B_i)'$, Γ_0 the probability limit of the Hessian matrix associated with the dyadic Poisson composite log-likelihood, and $\bar{s}_1^e(Q_i,U_i;\gamma_0)$ and $\bar{s}_1^a(Q_i,U_i;\gamma_0)$ as defined on page 40 (with Q_i playing the role of X_i).

With an estimate of γ in hand, form the fitted values $\{q(w,x,R_i,S_j;\hat{\gamma})\}_{i<j}$ and, invoking

Theorem 2, compute the analog estimate

$$\hat{m}^{\text{ASF}}(w,x;\hat{\gamma}) = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{q(w,x,R_i,S_j;\hat{\gamma}) + q(w,x,R_j,S_i;\hat{\gamma})}{2}. \tag{69}$$

To present the limit distribution of $\hat{m}^{\text{ASF}}(w,x;\hat{\gamma})$ I impose a regularity condition on the proxy variable regression function:

Assumption 6. *(i) $\gamma \in \mathbb{C} \subseteq \mathbb{R}^{\dim(\gamma)}$ with $\mathbb{C}$ compact, (ii) $q(w,x,r,s;\gamma)$ is twice continuously differentiable in γ, and (iii) the expectations $\mathbb{E}\left[\left|q(w,x,R_1,S_2;\gamma) + q(w,x,R_2,S_1;\gamma)\right|\right]$, $\mathbb{E}\left[\left\|\frac{\partial q(w,x,R_1,S_2;\gamma)}{\partial \gamma} + \frac{\partial q(w,x,R_2,S_1;\gamma)}{\partial \gamma}\right\|_2\right]$ and $\mathbb{E}\left[\left\|\frac{\partial^2 q(w,x,R_1,S_2;\gamma)}{\partial \gamma \partial \gamma'} + \frac{\partial^2 q(w,x,R_2,S_1;\gamma)}{\partial \gamma \partial \gamma'}\right\|_F\right]$ are finite.*

Under this assumption we have the following Lemma.

Lemma 1. (ASF Estimation) *Under Assumption 6, with $\hat{\gamma}$ a $\sqrt{N}$ consistent estimate of γ_0, we have that*

$$\begin{aligned}\sqrt{N}\left(\hat{m}^{\text{ASF}}(w,x;\hat{\gamma}) - m^{\text{ASF}}(w,x;\gamma_0)\right) =& \frac{2}{\sqrt{N}} \sum_{i=1}^{N} \psi_0(w,x,R_i,S_i;\gamma_0) \\ &+ M_0(w,x)\sqrt{N}(\hat{\gamma} - \gamma_0) + o_p(1)\end{aligned} \tag{70}$$

where

$$\begin{aligned}\psi_0(w,x,R_1,S_1;\gamma) &= \frac{q^e(w,x,R_1;\gamma) + q^a(w,x,S_1;\gamma_0)}{2} - m^{\text{ASF}}(w,x;\gamma) \\ M_0(w,x) =& \frac{1}{2}\mathbb{E}\left[\frac{\partial q(w,x,R_1,S_2;\gamma_0)}{\partial \gamma'} + \frac{\partial q(w,x,R_2,S_1;\gamma_0)}{\partial \gamma'}\right]\end{aligned}$$

with

$$\begin{aligned}q^e(w,x,r;\gamma) &= \mathbb{E}_S\left[q(w,x,r,S;\gamma)\right] \\ q^a(w,x,s;\gamma) &= \mathbb{E}_R\left[q(w,x,R,s;\gamma)\right].\end{aligned}$$

Proof. The result follows from Assumption 6 and an application of Lemma 1 in Appendix A. □

Lemma 1 and equation (68) yields an asymptotically linear representation for

$\sqrt{N}\left(\hat{m}^{\text{ASF}}(w,x;\hat{\gamma})-m^{\text{ASF}}(w,x;\gamma_0)\right)$ of

$$\begin{aligned}\sqrt{N}\left(\hat{m}^{\text{ASF}}(w,x;\hat{\gamma})-m^{\text{ASF}}(w,x;\gamma_0)\right)=&\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\psi_0(w,x,R_1,S_1;\gamma_0)\\&-M_0(w,x)\Gamma_0^{-1}\\&\times\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\bar{s}_1^e(Q_i,U_i;\gamma_0)+\bar{s}_1^a(Q_i,U_i;\gamma_0)}{2}\right\}\\&+o_p(1).\end{aligned} \tag{71}$$

Under correct (enough) specification of the composite likelihood, which will typically follow if the parametric form of the the PVR function is itself correctly specified, both $\bar{s}_1^e(Q_1,U_1;\gamma_0)$ and $\bar{s}_1^a(Q_1,U_1;\gamma_0)$ will be conditional mean zero given Q_1, hence the first and second terms in (71) will be uncorrelated with each other such that a CLT will imply a limit distribution of

$$\sqrt{N}\left(\hat{m}^{\text{ASF}}(w,x;\hat{\gamma})-m^{\text{ASF}}(w,x;\gamma_0)\right)\xrightarrow{D}\mathcal{N}\left(0,4\Xi_0(w,x)+4M_0(w,x)\left(\Gamma_0'\Sigma_1^{-1}\Gamma_0\right)^{-1}M_0(w,x)'\right)$$

with $\Xi_0(w,x)=\mathbb{V}(\psi_0(w,x,R_1,S_1;\gamma_0))$ and $\Sigma_1=\mathbb{V}\left(\frac{\bar{s}_1^e(Q_i,U_i;\gamma_0)+\bar{s}_1^a(Q_i,U_i;\gamma_0)}{2}\right)$.
The first term in the asymptotic variance reflects the econometrician's imperfect knowledge of the distribution of the proxy variables $(R_i',S_i')'$. The second term reflects the asymptotic penalty associated with not knowing the conditional distribution of Y_{12} given W_1,X_2,R_1,S_2. See Graham (2011) and Graham et al. (2018) for more expansive discussions in related contexts (see also Chamberlain (1992)).

In order to conduct inference an asymptotic variance estimate is required. Estimation of covariance matrix $\mathbb{V}\left(\sqrt{N}(\hat{\gamma}-\gamma_0)\right)=\left(\Gamma_0'\Sigma_1^{-1}\Gamma_0\right)^{-1}$ can proceed using one of the methods described in Section 4. The $\Xi_0(w,x)$ term may be estimated by

$$\hat{\Xi}(w,x)=\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}(w,x,R_i,S_i;\hat{\gamma})\hat{\psi}(w,x,R_i,S_i;\hat{\gamma})'$$

where $\hat{\psi}(w,x,R_i,S_i;\hat{\gamma})=\frac{1}{N-1}\sum_{j\neq i}\frac{q(w,x,R_i,S_j;\hat{\gamma})+q(w,x,R_j,S_i;\hat{\gamma})}{2}-\hat{m}^{\text{ASF}}(w,x;\hat{\gamma})$. The Jacobian, $M_0(w,x)$, is naturally estimated by

$$M_0(w,x)=\frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{1}{2}\left[\frac{\partial q(w,x,R_i,S_j;\hat{\gamma})}{\partial\gamma'}+\frac{\partial q(w,x,R_j,S_i;\hat{\gamma})}{\partial\gamma'}\right].$$

In practice, for reasons analogous to those discussed in Section 4, it may be preferable to replace the estimate of Σ_1 with one for Ω (as defined in equation (40)) and use a "Fafchampfs and Gubert" type estimate of $\mathbb{V}\left(\sqrt{N}\hat{m}^{\text{ASF}}\left(w,x;\gamma_0\right)\right)$ in place of $\hat{\Xi}\left(w,x\right)$.

5.5 Further reading and open questions

I am aware of no extant work on causal inference in the setting considered here. There is a large, and rapidly growing, literature on causal inference and interference, some of which makes connections to networks (e.g., Athey et al., 2018); VanderWeele & An (2013) provide a review of some relevant research.

The approach to estimation outlined above builds upon the dyadic regression material already introduced. A natural extension would to replace the parametric proxy variable regression function estimate with a non-parametric one (perhaps estimated using machine learning procedures). Inverse probability weighting (IPW) type estimators are also easily constructed (cf., Graham et al., 2018). I conjecture that augmented inverse probability weighting estimators (AIPW), exhibiting double robustness type properties, could also be constructed. The maximal asymptotic precision with which $m^{\text{ASF}}\left(w,x;\gamma_0\right)$ may be estimated under Assumptions 1 through 5 is also unknown. This semiparametric efficiency bound calculation, as in other network problems with likelihoods that don't easily factor into independent components, does not appear to be straightforward.

6 Incorporating unobserved heterogeneity

In its natural to associate the agent-specific U_i and U_j terms appearing in the Crane & Towsner (2018) representation result for X-exchangeable networks with unobserved correlated heterogeneity. In Section 4 I introduced methods for parametric estimation of the dyadic regression function $q\left(x,x'\right) \overset{def}{\equiv} \mathbb{E}\left[\left.Y_{ij}\right|X_i=x,X_j=x'\right]$. The relationship between $q\left(x,x'\right)$ and the graphon $h\left(x,x',u,u',v\right)$ depends, of course, on the dependence structure between X_i and U_i. Assumptions about this dependence structure played a prominent role in identifying the average structural function (ASF) in Section 5. In both Sections 4 and 5, however, the focus was on direct modeling of the conditional mean of Y_{ij} given observed covariates.

In this section I wish to explore the advantages of a modeling approach which directly specifies a parametric form for the graphon. This idea, at least implicitly, goes back to the work of Holland & Leinhardt (1981) and van Duijn et al. (2004).

The analysis in Sections 4 and 5 requires that the researcher directly specify the correct parametric form of the dyadic regression function. In contrast, the exact structure of (conditional) dependence across dyads sharing agents in common was left unspecified. To understand how such dependence might arise, it is useful to specify a structural *correlated* random effects model, analogous to those familiar from single-agent discrete choice panel data settings (e.g., Chamberlain, 1980, 1984).

6.1 A parametric dyadic potential response function

For the purposes of illustration, I will focus on modeling a directed binary outcome variable. The generalization to non-binary outcomes is straightforward. Refer to the dyadic potential response function introduced in Assumption 1. Consider the following parametric form for this function

$$\begin{aligned} Y_{12}\left(w_1, x_2\right) &= \mathbf{1}\left(t^e\left(w_1\right)' \beta_0^e + t^a\left(x_2\right)' \beta_0^a + \omega\left(w_1, x_2\right)' \gamma_0 + A_1 + B_2 + V_{12} > 0\right) \\ &= h\left(w_1, x_2, A_1, B_2, V_{12}\right) \end{aligned} \tag{72}$$

with

$$\left(V_{12}, V_{21}\right) \mid Q_1, Q_2, A_1, B_1, A_2, B_2 \sim \mathcal{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \zeta \\ \zeta & 1 \end{array}\right)\right) \tag{73}$$

and independently distributed across dyads. As in Section 5, X_i and W_j correspond to the chosen ego and alter treatments; A_i and B_j are unobserved ego and alter heterogeneity, which may be correlated with these treatment choices, and R_i and S_j are proxy variables (recall that $Q_i = (W_i', X_i', R_i', S_i')'$). The vectors $t^e\left(w_1\right)$, $t^a\left(x_2\right)$ and $\omega\left(w_1, x_2\right)$ consist of known basis functions in the underlying treatment variables. In the case where both W_i and X_j are binary we would set $t^e\left(w_1\right) = w_1$, $t^a\left(x_2\right) = x_2$ and $\omega\left(w_1, x_2\right) = w_1 x_2$.

Next posit the correlated random effects specification for the joint distribution of the ego and alter heterogeneity

$$\left.\begin{array}{c} A_i \\ B_i \end{array}\right| W_i, X_i, R_i, S_i \sim \mathcal{N}\left(\left(\begin{array}{c} \alpha_0^e + k^e\left(R_i\right)' \delta_0^e \\ \alpha_0^e + k^a\left(S_i\right)' \delta_0^a \end{array}\right), \left(\begin{array}{cc} \sigma_A^2 & \rho\sigma_A\sigma_B \\ \rho\sigma_A\sigma_B & \sigma_B^2 \end{array}\right)\right), \tag{74}$$

with $k^e\left(R_i\right)$ and $k^a\left(S_i\right)$ vectors of known functions of the proxy variables. Note that (72) and (74) jointly imply the selection on observables, Assumption 4 introduced earlier. Redundancy and strict exogeneity, respectively Assumptions 2 and 3, also hold in this set-up.

Averaging over A_i and B_j gives a dyadic proxy variable regression function of

$$q\left(W_i, X_j, R_i, S_j; \eta_0\right) = \Phi\left(T_{ij}' \eta_0\right) \tag{75}$$

for $\eta_0 = \left(1 + \sigma_A^2 + \sigma_B^2\right)^{-1/2} \left(\alpha_0^e + \alpha_0^a, \left(\beta_0^e\right)', \left(\beta_0^a\right)', \gamma_0', \left(\delta_0^e\right)', \left(\delta_0^a\right)'\right)'$ and

$$T_{ij} = \left(1, t^e\left(W_i\right), t^a\left(X_j\right), \omega\left(W_i, X_j\right)', k^e\left(R_i\right), k^a\left(S_j\right)\right)'.$$

It is possible to estimate η_0 along the lines outlined in Section 4 above. Alternatively one could attempt to directly maximize the integrated likelihood implied by (72), (73) and (74). This would be computationally non-trivial since the integral does not easily factor. van Duijn et al. (2004) and Zijlstra et al. (2009) develop this approach using Markov Chain Monte Carlo (MCMC) methods.

6.2 Triad probit: a correlated random effects estimator

An intermediate approach, which is more efficient than the basic dyadic regression estimator introduced earlier, and additionally recovers more features of the graph generation process, is what I will call *triad probit*. Triad probit is also a composite likelihood estimator. Instead of modeling the dyadic outcome, $Y_{12}| Q_1, Q_2$, marginally however, it is composed of component likelihoods for the joint outcome $\left(Y_{12}, Y_{21}, Y_{13}, Y_{31}\right)| Q_1, Q_2, Q_3$. That is I model the outcome configuration associated with a *pair-of-dyads* sharing one agent in common. An overall criterion function is constructed by summing over the component log-likelihoods, so constructed, for all $3\binom{N}{3}$ pairs-of-dyads sharing one agent in common.[23]

The probability of the event $Y_{12} = y_{12}, Y_{21} = y_{21}, Y_{13} = y_{13}, Y_{31} = y_{31}$ given the parameters and regressors is

$$\Pr\left(Y_{12} = y_{12}, Y_{21} = y_{21}, Y_{13} = y_{13}, Y_{31} = y_{31}| Q_1, Q_2, Q_3\right) = \int_{\mathcal{A}_{12}} \int_{\mathcal{A}_{21}} \int_{\mathcal{A}_{13}} \int_{\mathcal{A}_{31}} \phi_4\left(\mathbf{t}| \Sigma\right) \mathrm{d}\mathbf{t} \tag{76}$$

with $\phi_4\left(t| \Sigma\right)$ the density of a tetra-variate normal distribution with mean zero and covari-

[23] This approach is related to the pairwise likelihood estimator for models with crossed random effects discussed by Bellio & Varin (2005) and Cattelan & Varin (2013).

ance matrix Σ. The intervals of integration are given by

$$\mathcal{A}_{ij} = \begin{cases} \left(-\infty, T_{ij}'\eta_0\right) & \text{if } y_{ij} = 1 \\ \left[T_{ij}'\eta_0, \infty\right) & \text{if } y_{ij} = 0 \end{cases},$$

with the covariance matrix, which is in correlation form (a scale normalization), taking the form

$$\Sigma = \Sigma\left(\zeta, \sigma_A, \sigma_B, \rho\right) = \begin{pmatrix} 1 & \frac{\zeta+2\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} & \frac{\sigma_A^2}{1+\sigma_A^2+\sigma_B^2} & \frac{\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} \\ \frac{\zeta+2\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} & 1 & \frac{\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} & \frac{\sigma_B^2}{1+\sigma_A^2+\sigma_B^2} \\ \frac{\sigma_A^2}{1+\sigma_A^2+\sigma_B^2} & \frac{\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} & 1 & \frac{\zeta+2\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} \\ \frac{\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} & \frac{\sigma_B^2}{1+\sigma_A^2+\sigma_B^2} & \frac{\zeta+2\rho\sigma_A\sigma_B}{1+\sigma_A^2+\sigma_B^2} & 1 \end{pmatrix}.$$

The integral (76) does not have a closed form expression. Fortunately a large econometrics and statistics literature suggest various methods for its numerical evaluation; see, for example, Keane (1994) and Chib & Greenberg (1998).

Let $l_{123}^{*}\left(\theta\right)$ equal the logarithm of (76) with $\theta = \left(\eta', \zeta, \sigma_A, \sigma_B, \rho\right)'$. To induce symmetry in the criterion function summands I form the average

$$l_{ijk}\left(\theta\right) = \frac{1}{3}\left[l_{ijk}^{*}\left(\theta\right) + l_{jik}^{*}\left(\theta\right) + l_{kij}^{*}\left(\theta\right)\right].$$

The triad probit estimate $\hat{\theta}_{\text{TP}}$ of θ_0 is the maximizer of the sum of the $l_{ijk}\left(\theta\right)$ kernels over all $\binom{N}{3}$ triads in the network:

$$L_N\left(\theta\right) = \binom{N}{3}^{-1} \sum_{i<j<k} l_{ijk}\left(\theta\right). \tag{77}$$

Note that (4) sums over all $3\binom{N}{3}$ pairs-of-dyads sharing one agent in common. It does this by summing over all $\binom{N}{3}$ triads in the network and, for each such triad, summing over the three pairs-of-dyads sharing an agent in common that can be constructed from it.

The criterion (77) is not a U-process-minimizer, although, as in the other contexts introduced above, it shares similarities with one. The results of Honoré & Powell (1994) do not immediately characterize the asymptotic sampling properties of $\hat{\theta}_{\text{TP}}$. Nevertheless arguments similar to those outlined in Sections 4 and 5 above can be applied to also analyze $\hat{\theta}_{\text{TP}}$.

A quick outline of these arguments goes as follows. Let $S_N\left(\theta\right) = \binom{N}{3}^{-1} \sum_{i<j<k} s_{ijk}\left(\theta\right)$ with $s_{ijk}\left(\theta\right) = \frac{\partial l_{ijk}(\theta)}{\partial\theta}$. Also define $\Gamma_0 = \mathbb{E}\left[\frac{\partial^2 l_{ijk}(\theta)}{\partial\theta\partial\theta'}\right]$ and, as earlier, $\Sigma_q = \mathbb{E}\left[s_{i_1i_2i_3}s_{j_1j_2j_3}'\right]$ to be the covariance of $s_{i_1i_2i_3}$ and $s_{j_1j_2j_3}$ when they share $q = 0, 1, 2, 3$ indices in common.

Calculation then gives

$$\mathbb{V}\left(\sqrt{N}S_N\left(\theta\right)\right) = 9\Sigma_1 + \frac{18}{N-1}\left(\Sigma_2 - 2\Sigma_1\right) + \frac{6}{\left(N-1\right)\left(N-2\right)}\left(\Sigma_3 + 3\Sigma_1\right) \tag{78}$$

which suggests, under regularity conditions, the limiting distribution

$$\sqrt{N}\left(\hat{\theta}_{\mathrm{TP}} - \theta_0\right) \xrightarrow{D} N\left(0, 9\Gamma_0^{-1}\Sigma_1\Gamma_0^{-1}\right). \tag{79}$$

Associated with the triad probit is a proxy variable regression function estimate of

$$q\left(W_i, X_j, R_i, S_j; \hat{\eta}_{\mathrm{TP}}\right) = \Phi\left(T_{ij}'\hat{\eta}_{\mathrm{TP}}\right)$$

from which an estimate of the ASF (or differences thereof) can be directly constructed according to equation (69). This corresponds (essentially) to a dyadic generalization of the average partial effect (APE) estimator introduced by Chamberlain (1984) in the context of a correlated random effects probit panel data model.

6.3 Fixed effects approaches

The models introduced above, while allowing for dependence in outcomes across dyads sharing agents in common, restrict its structure. In contrast, Graham (2017) provides a fixed effects analysis of a model where a undirected binary dyadic outcome is determined according to

$$Y_{ij} = \mathbf{1}\left(\left[t\left(X_i\right) + t\left(X_j\right)\right]'\beta_0 + \omega\left(X_i, X_j\right)'\gamma_0 + A_i + A_j - V_{ij} \leq 0\right), \tag{80}$$

with V_{ij} standard logistic and independent across dyads. Specifically he studies identification and estimation of γ_0, leaving the joint distribution of X_i and A_i unrestricted (without restrictions on this distribution β_0 is unidentified (cf., Hausman & Taylor, 1981; Arellano & Bover, 1995). The parameter of interest, γ_0, indexes the strength of any homophilous sorting on the observables agent attributes in X_i, while $\{A_i\}_{i=1}^N$ indexes unobserved *degree-heterogeneity*. Since real world network degree distributions often have high variance (and in particular fat right tails), incorporating degree heterogeneity may be important in practice (Barabási & Albert, 1999; Barabási & Bonabau, 2003). Graham (2017) shows how failing to accommodate degree heterogeneity may attenuate measured homophily (i.e., bias estimates of γ_0).

Conditional on $\mathbf{X} = \left(X_1, \ldots, X_N\right)'$ and $\mathbf{A} = \left(A_1, \ldots, A_N\right)'$, the likelihood for the adjacency matrix $\mathbf{D}$ factors into $\binom{N}{2}$ conditionally independent components. Absorbing $t\left(X_i\right)'\beta_0$ into the individual effect A_i, the model consists of the finite dimensional parameter of interest, γ_0, and the N incidental heterogeneity parameters, $\mathbf{A}_0$. Let $K = \dim(\gamma_0)$; in this model the number of parameters, $K + N$, is a function of the order of the network. Since this number grows with N, the model is non-standard (cf., Holland & Leinhardt, 1981; Chatterjee et al., 2011).

Graham (2017) analyzes the large network properties of two estimates of γ_0. The first estimate, leveraging the implicit "large-N, large-T" structure of dense networks, is the joint maximum-likelihood one, which also simultaneously estimates the incidental parameters $\mathbf{A}_0 = (A_{01}, \ldots, A_{0N})'$. The second exploits the exponential family structure of the model and conditions on a sufficient statistic for $\mathbf{A}_0$. Both estimates have antecedents in the literature on panel data.

Joint estimators

Let T_{ij} be an $N \times 1$ vector with a one in the i^{th} and j^{th} elements and zeros elsewhere. The joint-MLE coincides with the logit fit of Y_{ij} onto $\omega(X_i, X_j)$ and T_{ij} for all $i < j$.[24] Although this estimator involves $K+N$ parameters, it is based upon a criterion function with $\binom{N}{2} = O(N^2)$ summands. This feature is similar to joint maximum likelihood estimation in a panel data setting where both N and T are allowed to grow. Here each of N agents make $N-1$ linking decisions; the latter is analogous to "T" in the "large-N, large-T" panel data setting. As the number of agents in the network grows, so to does the number of link decisions observed for each of them. This feature of the model allows for consistent estimation of both γ_0 and $\mathbf{A}_0$, although, as in the panel data case, there is a bias in the limit distribution of $\hat{\gamma}$ which must be corrected in order to undertake asymptotically valid inference (Hahn & Newey, 2004; Arellano & Hahn, 2007).[25]

Graham's (2017) assumptions imply that the limiting network will be dense. Yan et al. (2018) show that it is possible to weaken his assumptions somewhat, but it appears impossible to accommodate asymptotic sequences with sparse limits. In Monte Carlo experiments the joint MLE works poorly in networks with low density. Researchers are advised to be cautious when applying this estimator to low density networks.

Dzemski (2018) and Yan et al. (2018) study joint estimation of a directed version of (80). The former paper presents a method of testing for reciprocity in links as well as for neglected transitivity.

[24]Graham (2017) outlines a more convenient nested-fixed-point approach to estimation based upon an insight due to Chatterjee et al. (2011).

[25]A technical difficultly involving the inverse Hessian arises in the network setting. A similar challenge is also present in panel data models with time effects (Fernández-Val & Weidner, 2016).

Conditional estimators

Under the logistic assumption, the likelihood associated with (80) is a member of the exponential family. It turns out that the degree sequence of the network is a sufficient statistic for $\mathbf{A}_0$ (Snijders, 2002). A conditional maximum likelihood estimator could be constructed, however, unlike in the panel case considered by Chamberlain (1980), the likelihood does not nicely factor into independent components. It would also be non-trivial to evaluate and maximize the conditional likelihood (cf., Blitzstein & Diaconis, 2011).

Graham (2017) instead builds a criterion involving tetrads – quadruples of agents. A tetrad is the smallest subgraph that is not completely determined by its degree sequence. For example, there are three isomorphisms of the two edge graphlet  on four vertices, each with an identical subgraph degree sequence of $(1,1,1,1)'$. If $\gamma_0 = 0$, then conditional on the event that a randomly sampled tetrad takes one of these three forms, any one of them occurs with an equal probability of one third. Deviations from this benchmark are possible when $\gamma_0 \neq 0$, depending on the configuration of covariates across agents in the sampled tetrad. Graham's (2017) conditional estimator, which he calls tetrad logit, is based upon this insight.

The large network properties of the tetrad logit estimate of γ_0 may be derived in a way roughly analogous to that of the dyadic regression estimators introduced above. The analysis in Graham (2017), however, allows for sequences of graphs which are sparse in the limit. This affects the rate-of-convergence of the tetrad logit estimate. Conveniently its limit distribution remains normal under both dense and sparse sequences.

Jochmans (2018) provides a conditional analysis, including several worked empirical examples, of a directed analog of tetrad logit. Nadler (2015) proposes a related estimator for bipartite networks and presents an empirical application.

6.4 Further reading and open questions

Varin et al. (2011) survey the statistics literature on composite likelihoods. A standard reference on U-Process minimizers is Honoré & Powell (1994). Many of the results presented in this section, as well as the previous ones, utilize ideas coming from the theory of composite likelihood and U-Process minimizers. Connections to panel data have also featured prominently; here I recommend Chamberlain (1980), Chamberlain (1984), Arellano & Honoré (2001), and Arellano & Hahn (2007).

The triad probit estimator introduced above has a rate of convergence equal to $\sqrt{N}$. In the simplest setup the the tetrad logit estimator has a faster $\sqrt{\binom{N}{2}}$ rate of convergence. This is peculiar because, invoking intuitions familiar from panel data, one would generally expect an estimate based upon an integrated/random effects likelihood to be more efficient than one based upon a conditional/fixed effects likelihood. Here the two estimators have different rates of convergence with, perhaps, a ranking reverse of what one might expect *a priori*.

van Duijn et al. (2004) use MCMC methods to (essentially) maximize the network likelihood implied by (72), (73) and (74). Their approach to inference is Bayesian; it would be interesting to formally study the maximum integrated likelihood estimator proper (as opposed to the triad probit composite likelihood estimator introduced here). What is the rate of convergence associated with the true random effects maximum likelihood estimator (MLE)? Likewise, tetrad logit, while inspired by conditional likelihood ideas, is not a conditional MLE (it is akin to a conditional composite MLE). Graham (2017) describes the conditional MLE, but does not formally analyze it. Such a formal analysis could be insightful. More generally we know very little about efficiency in even the simplest of network problems.

The introduction of heterogeneity in this section is restrictive in nature. It allows for what Graham (2017) calls degree heterogeneity. Methods for incorporating assortative matching on latent agent-specific attributes would also be useful. For inspiration see, for example, Krivitsky et al. (2009). Recent ideas from panel data may be useful here too; especially the work on discrete heterogeneity done by Bonhomme & Manresa (2015). Ideas from the stochastic block literature – which is not surveyed in this chapter – might also be useful for incorporating richer heterogeneity structure into econometric models for dyadic outcomes.

7 Asymptotic distribution theory for network statistics

Wasserman & Faust (1994) exposit a large post World War II literature on the computation and interpretation of different statistics of the adjacency matrix. Researchers routinely report statistics like reciprocity, transitivity, moments of the degree sequence, and diameter when presenting real world network data. Measures of statistical uncertainty almost never accompany these reports. The leading approach to assessing whether a reported network statistic is unusual is to informally compare it with its expected value under an Erdös-Renyi null or, alternatively, a reference sample of real world networks (e.g., Milo et al., 2002; Newman, 2010; Graham, 2015).[26] Informal simulation-based approaches to "inference" abound.

[26]Blitzstein & Diaconis (2011) present an elegant approach based on comparing statistics of the network in hand to those of the reference set of all graphs with the same degree sequence (i.e., a β-model null).

Large network approaches to hypothesis testing only recently emerged (e.g., Bobollas et al., 2007; Picard et al., 2008; Bickel et al., 2011). This is currently an active research area (e.g., Gao & Lafferty, 2017; Green & Shalizi, 2017; Menzel, 2017), with many open questions. To be fair, work on the distributional properties of network statistics under specific graph generation processes, generally the Erdös-Renyi one or close variants, was undertaken earlier. This work arose largely in response to the seminal papers by Holland & Leinhardt (1970, 1976). Examples include the work of Frank (1979; 1980; 1988), Wasserman (1977) and Nowicki (1991). The last reference is a useful survey of such analyses.

This section presents results on the large network distribution of induced subgraph frequencies (and various statistics constructed from them). I begin, in Subsection 7.1, with a detailed analysis of triad counts and their application to inference on the transitivity index or global clustering coefficient (e.g., Kolaczyk, 2009, p. 96). This is a classic, practically important, and pedagogically valuable, example. Results on counts of trees and cycles of any order are available in the Appendix. In Subsection 7.2, I turn to moments of the degree distribution, an area of intense focus in applied work (e.g., Barabási & Bonabau, 2003; Atalay et al., 2011; Acemoglu et al., 2012).

Not all common network statistics are covered by the results presented in this section. Statistics such as diameter and average path length, for example, have, to my knowledge, unknown sampling properties. Subsection 7.3 discusses open questions.

The work surveyed in this section dates to the papers by Holland & Leinhardt (1970, 1976). More recent contributions, generally by statisticians, were often motivated by examples from computational biology (e.g., Picard et al., 2008). An especially important contribution is the paper by Bickel et al. (2011). This section draws heavily from the work by Bickel and coauthors. Related ideas were used in the discussion of dyadic regression in Section 4. Recent work on strategic models of network formation, where econometricians play the leading role, arose separately. However, in Section 8 I argue that ideas from research on subgraph counts could be valuable there as well. Specifically for structural estimation of strategic network formation models.

The results in this section are based on the following hypothetical repeated sampling experiment. Let $G_{\infty,N}$ be an infinite exchangeable random graph of interest. The network in hand, G_N, is the one induced by a random sample of N vertices from $G_{\infty,N}$. Let $h_N(u,v)$ denote the Aldous-Hoover graphon characterizing the infinite graph $G_{\infty,N}$ from which the econometrician samples N agents independently at random. Note I suppress dependence of this graphon on the mixing parameter, α, since I seek to conduct inference conditional on it (i.e., conditional on the empirical distribution of $[D_{ij}]_{i,j\in\mathbb{N},i<j}$).

Using the observed network, G_N, we construct the statistic $t_N(G_N)$. The sampling distribution of this statistic is the one induced by repeated sampling of N agents from the underlying infinite graph $G_{\infty,N}$. To derive a limit distribution I assume there is a sequence of infinite random graphs $\{G_{\infty,N}\}$ – indexed by N – such that

$$h_N(u,v) = \rho_N w(u,v)$$

with ρ_N (possibly) approaching zero as $N \to \infty$. In this way I pair a sequence of increasingly larger "sampled" networks with a corresponding sequence of infinite networks that are allowed to become increasingly sparser. With this set-up we can study the distribution of $t_N(G_N)$, appropriately scaled, as $N \to \infty$.

As noted earlier, the above thought experiment does not mirror how empirical networks are constructed in practice. Typically one of two cases obtain. In the first, the network under study really is a very large graph (e.g., the Facebook graph) and the econometrician really does sample from it. However, due to spareness, sampling is rarely conducted as described above. Instead snowball sampling, edge sampling, path sampling etc. are typically used (Crane, 2018). Understanding how to consistently estimate network statistics and their sampling distributions under these more exotic data collection schemes is an interesting topic for future research. In the second case the econometrician works with the complete graph on some finite population of vertices. In this cases the idea of sampling from an infinite graph is a thought experiment used to get results that are hopefully useful in practice. It is this latter, rather commonplace case, which I have in mind here.

There is a subtlety in this second case, already touched upon in Section 3 in the context of my discussion of the Aldous-Hoover Theorem. A jointly exchangeable random graph with a *finite* number of agents need not have a probability law with a conditionally independent dyad (CID) structure. The pattern of dependence across links in such a network may be more complicated than that implied by the Aldous-Hoover representation. I conjecture, by speculative extrapolation based upon the example introduced in Section 3, that this is especially the case when agents form links strategically. We know, however, that, for N large enough, joint exchangeability will deliver a probability law for the network that is of the Aldous-Hoover form. This suggests that, to derive limit theory, it is reasonable to proceed in the way I do here; but there are missing steps in the argument. Menzel (2016) represents the only attempt I am aware of to struggle with these issues in a disciplined way. A more rigorous pairing of the game theoretic models of network formation of interest to many economists, with the theory of graph limits would be a high priority topic for future research.

7.1 Large network estimation of the transitivity index

In the social sciences, hypothesis formulation often involves graphlet counts (e.g., Holland & Leinhardt, 1970; Bearman et al., 2004; Choi & Wu, 2009; Jackson et al., 2012; Isakov et al., 2019).[27] Graphlet counts are also used to construct important network statistics like the transitivity index. It is this last statistic that is studied in this subsection.

After introducing some notation and definitions, I apply the basic approach outlined by Bhattacharya & Bickel (2015, Proposition 6) to calculate variance expressions for induced subgraph counts of two-stars () and triangles (). While this is a relatively straightforward extension, it does require some carefully constructed notation.[28] Asymptotic normality of these counts, appropriately scaled, follows from their results. An analysis of transitivity in the Nyakatoke risk-sharing network studied by De Weerdt (2004) illustrates the practical application of these ideas.

A special case of a CID model is the Erdös-Renyi graph generation process (i.e., $h(u,v) = \rho$ for some $0 < \rho < 1$ and all $(u,v) \in [0,1]^2$). The behavior of subgraph counts under this GGP were studied by Nowicki and co-authors in the late 1980s and early 1990s (Nowicki & Wierman, 1988; Janson & Nowicki, 1991; Nowicki, 1991). It turns out that this case exhibits a form of degeneracy. Specifically, the leading terms in the variance expressions presented below are identically zero under the Erdös-Renyi graph generation process. Subgraph frequencies remain asymptotically normal in this case, but with a faster rate of convergence. A separate treatment of this case is provided below.

Notation and estimation

Recall from Section 3 that the induced subgraph frequency of S in G_N is

$$P_N(S) = \frac{1}{\binom{N}{p}\,|\mathrm{iso}(S)|} \sum_{\mathbf{i}_p \in \mathcal{C}_{p,N}} \mathbf{1}\left(S \cong G_N[\mathbf{i}_p]\right). \tag{81}$$

Under the maintained sampling scheme it is easy to see that (81) is an unbiased estimate of $P(S) = t_{\mathrm{ind}}(S,h)$, the "population" induced subgraph density.

Consider the two-star () and triangle () triad configurations. Applying (81) gives the

[27] In practice it is easier to derive results for homomorphism frequencies and, not coincidentally, the theory of graph limits generally works with homomorphisms.

[28] One could even argue that these expressions are already implicit in Holland & Leinhardt (1976), although they did not explore the properties of their expressions under sparse versus dense graph sequences, nor did they analyze rates of convergence. Indeed, Wasserman & Faust (1994, p. 580), referring to the covariance calculations of Holland & Leinhardt (1976), comment that they "can be time-consuming to calculate (and maybe even difficult to comprehend)".

estimates

$$P_N(\wedge) = \binom{N}{3}^{-1} \frac{1}{3} \sum_{\mathbf{i}_3 \in \mathcal{C}_{3,N}} \left[D_{i_1 i_2} D_{i_1 i_3} (1 - D_{i_2 i_3}) + D_{i_1 i_2} (1 - D_{i_1 i_3}) D_{i_2 i_3} \right. \tag{82}$$

$$\left. + (1 - D_{i_1 i_2}) D_{i_1 i_3} D_{i_2 i_3} \right]$$

$$P_N(\triangle) = \binom{N}{3}^{-1} \sum_{\mathbf{i}_3 \in \mathcal{C}_{3,N}} D_{i_1 i_2} D_{i_1 i_3} D_{i_2 i_3}. \tag{83}$$

From (82) and (83) we can construct an estimate of the *transitivity index* or global clustering coefficient:

$$\mathrm{TI_N} = \frac{3 \times (\#\text{ of triangles})}{(\#\text{ of two-stars}) + 3 \times (\#\text{ of triangles})} = \frac{P_N(\triangle)}{P_N(\wedge) + P_N(\triangle)} = \frac{Q_N(\triangle)}{Q_N(\wedge)}. \tag{84}$$

Under an Erdös-Renyi graph generation process it is easy to show that (84) should be close to the density of the network (e.g., Graham, 2015). Gao & Lafferty (2017) develop a test based on this idea. If, suitably normalized, the limit distribution of the vector $(P_N(\wedge), P_N(\triangle))'$ can be characterized, then delta methods can be used to conduct large network inference on transitivity. This idea is developed in detail below.

Distribution theory for induced subgraph counts may also be useful for structural model estimation via the method of (simulated) minimum distance. In this approach model parameters are estimated by matching model-implied values of subgraph counts with their empirical counterparts. Sampling uncertainty in such estimates, stems from the corresponding uncertainty about the reduced form subgraph counts being matched. This idea is developed more completely in Section 8.

Graphlet Stitchings

In developing an interpretable expression for the variance of graphlet counts, it is helpful to introduce something I will call a *graphlet stitching*.[29]

Let R and S be two p^{th} order subgraphs of interest to the econometrician. Furthermore, let $\mathbf{i}_p$ and $\mathbf{j}_p$ be two p-tuples drawn independently at random from $\mathcal{C}_{p,N}$ (as defined in Section 3.5 above). The (scaled) covariance of the events "$G_N[\mathbf{i}_p]$ is isomorphic to R" and "$G_N[\mathbf{j}_p]$

[29] After completing the initial draft of this Chapter I discovered independent work by Green & Shalizi (2017) that develops a closely related concept which they call "merged copy sets". Graphlet stitchings, as I define them, are more suited to my specific needs; although both approaches lead to the same answer in the end. The basic idea is already implicit in Bhattacharya & Bickel (2015) (and really even Holland & Leinhardt (1976)). Essentially the same idea is also used in Graham (2017) to derive large network theory for Tetrad Logit.

is isomorphic to S", when there are q integers/vertices common to $\mathbf{i}_p = \{i_1, i_2, \dots, i_p\}$ and $\mathbf{j}_p = \{j_1, j_2, \dots, j_p\}$, is

$$\Sigma_q(R, S) = \Xi(\mathcal{W}_{q,R,S}) - P(R)P(S) \tag{85}$$

where $P(R)$ is the induced subgraph density defined in equation (11) and

$$\Xi(\mathcal{W}_{q,R,S}) \overset{def}{\equiv} \frac{\mathbb{E}\left[\mathbf{1}(R \cong G_N[\mathbf{i}_p])\,\mathbf{1}(S \cong G_N[\mathbf{j}_p])\right]}{|\mathrm{iso}(R)|\,|\mathrm{iso}(S)|} \tag{86}$$

Here $\mathcal{W}_{q,R,S}$ is notation for a *set* of what I call *graphlet stitchings*. In order to understand the structure of $\Xi(\mathcal{W}_{q,S,R})$ further we need a formal definition.

Definition 5. (Graphlet Stitching) Let $W_{q,R,S}$ be the graph union of R and S, labelled isomorphisms of two graphlets of interest, if
(i) $\mathcal{V}(R) \subseteq \mathcal{V}(G)$ and $\mathcal{V}(S) \subseteq \mathcal{V}(G)$;
(ii) $|\mathcal{V}(R)| = |\mathcal{V}(S)| = p$ vertices each;
(iii) $|\mathcal{V}(R) \cap \mathcal{V}(S)| = q$ vertices in common;
(iv) identical structure across all vertices in common (i.e., $(i,j) \in \mathcal{E}(R) \Leftrightarrow (i,j) \in \mathcal{E}(S)\ \forall\, i,j \in \mathcal{V}(R) \cap \mathcal{V}(S)$),
then $W_{q,R,S}$ is a graphlet stitching of R and S.

Next define the *set* of all feasible stitchings of R and S which satisfy Definition 5 as $\mathcal{W}_{q,S,R}$. When R and S belong to the same isomorphism class write $\mathcal{W}_{q,S,S} = \mathcal{W}_{q,S}$.

Requirement (iv) of Definition 5 is constraining. It implies, for example, that some pairs of labelled two-stars cannot be stitched together. For example $R = (\{1,2,3\}, \{(1,2),(1,3)\})$ and $S = (\{1,2,4\}, \{(1,4),(2,4)\})$ cannot be logically stitched together because the $(1,2)$ edge is present in R but not S. This violates requirement (iv) of Definition 5. Note also that the set $\mathcal{W}_{q,S,R}$ may contain elements which are isomorphic to one another.

For simplicity consider the vertices $1, 2, \dots, p, p+1, \dots, 2p-q$ in G_N.[30] If R, defined on vertices $1, 2, \dots, p$, is isomorphic to the subgraph of G_N induced by vertices $\{1, \dots, p\}$ and S, defined on vertices vertices $p-q, \dots, 2p-q$, is isomorphic to the subgraph of G_N induced by vertices $\{p-q, \dots 2p-q\}$, then it must be the case that the union of these two induced subgraphs is an element of $\mathcal{W}_{q,S,R}$. This gives the equality

$$\Xi(\mathcal{W}_{q,S,R}) = \sum_{W \in \mathcal{W}_{q,S,R}} \frac{\Pr\left(W = G_N[\{1,\dots,p\}] \cup G_N[\{p-q,\dots,2p-q\}]\right)}{|\mathrm{iso}(R)|\,|\mathrm{iso}(S)|}. \tag{87}$$

[30]Since G_N is induced by a random sample of vertices, vertices $1, 2, \dots, p, p+1, \dots, 2p-q$ correspond to a random $2p-q$ tuple.

Note that the graph union of $G_N[\{1,\ldots,p\}]$ and $G_N[\{p-q,\ldots,2p-q\}]$ may differ from the subgraph induced by the union of the two overlapping vertex sets:

$$G_N[\{1,\ldots,p\}] \cup G_N[\{p-q,\ldots,2p-q\}] \neq G_N[\{1,\ldots,2p-q\}].$$

This is because the union of $G_N[\{1,\ldots,p\}]$ and $G_N[\{p-q,\ldots,2p-q\}]$ will not include any edges between $\{1,\ldots,p-q-1\}$, the vertices in R alone, and $\{p+1,\ldots,2p-q\}$, the vertices in S alone, while $G_N[\{1,\ldots,2p-q\}]$ may. By exchangeability the right-hand-side of (87) is the same for any vertex sets $\mathbf{i}_p = \{i_1, i_2, \ldots, i_p\}$ and $\mathbf{j}_p = \{j_1, j_2, \ldots, j_p\}$ sharing, as is implicitly assumed in what follows, q vertices in common.

To check whether $R \cong G_N[\mathbf{i}_p]$ and $S \cong G_N[\mathbf{j}_p]$ we therefore check whether $G_N[\mathbf{i}_p] \cup G_N[\mathbf{j}_p]$ coincides with a particular (labeled) graphlet stitching of R and S. Doing so, in turn, requires us to check for the presence *or* absence of only $p(p-1) - \binom{q}{2}$ potential edges. The presence or absence of the $(p-q)^2$ possible edges from the vertices unique to R to those unique to S is immaterial. Equation (86) gives neither an induced or partial subgraph frequency, but what I will call a *graphlet stitching frequency*.

Calculating graphlet stitching frequencies

To understand how to calculate graphlet stitching frequencies in practice it is helpful to work through a few examples. Figure 5 shows all the elements of $\mathcal{W}_{1,\wedge}$ on vertex set $\{1,2,3,4,5\}$, with vertex 1 being the vertex in common. The top row shows all isomorphisms of $\wedge$ on vertices $\{1,4,5\}$, while the left-most column shows all such isomorphisms on vertices $\{1,2,3\}$. The nine figures in the corresponding grid show all the associated graphlet stitchings.

A more complicated example is provide by $\mathcal{W}_{2,\wedge}$, which is shown in Figure 6. The format of the figure is the same as that of Figure 5. The two vertices in common are 1 and 2. An interesting feature of this example is that not all graphlet stitchings are feasible.

In evaluating $\Xi(\mathcal{W}_{q,S,R})$ it is helpful to observe that $\mathcal{W}_{q,S,R}$ may include multiple isomorphisms of the same graph. Since the probabilities $\Pr(W = G_N[\{1,\ldots,p\}] \cup G_N[\{p-q,\ldots,2p-q\}])$ and $\Pr(W' = G_N[\{1,\ldots,p\}] \cup G_N[\{p-q,\ldots,2p-q\}])$ coincide when W and W' are isomorphic to one another, we can also "represent" $\mathcal{W}_{q,S,R}$ as a multi-set, with one (arbitrary) labelling of each of the non-isomorphic graphlet stitchings retained as elements, but with multiplicities equal to the number of isomorphic appearances. For example, the cardinality of $\mathcal{W}_{1,\wedge}$ is $|\mathrm{iso}(\wedge)| \times |\mathrm{iso}(\wedge)| = 9$, but with only three non-isomorphic elements.

Inspecting Figure 5 we define the multi-set:

$$\mathcal{W}^{\mathrm{m}}_{1,\wedge} = \left(\left\{\times, \underline{\times}, \overline{\underline{\times}}\right\}, \left\{(\times, 1), (\underline{\times}, 4), (\overline{\underline{\times}}, 4)\right\}\right).$$

Let $\nu_{q,R,S}(W)$ denote the multiplicity of W in $\mathcal{W}^{\mathrm{m}}_{q,R,S}$; for example the multiplicity of $\underline{\times}$ in $\mathcal{W}^{\mathrm{m}}_{1,\wedge}$ is $\nu_{1,\wedge}(\underline{\times}) = 4$.

We then have that, using equation (87), the equality $\Xi\left(\mathcal{W}^{\mathrm{m}}_{1,\wedge}\right) = \Xi\left(\mathcal{W}_{1,\wedge}\right)$. Similarly, inspecting $\mathcal{W}_{2,\wedge}$ (see Figure 6) , we see that it also contains three non-isomorphic elements, yielding

$$\mathcal{W}^{\mathrm{m}}_{2,\wedge} = \left(\left\{\boxslash, \sqcap, \square\right\}, \left\{(\boxslash, 2), (\sqcap, 2), (\square, 1)\right\}\right).$$

Finally, it is easy to see that $\mathcal{W}^{\mathrm{m}}_{3,\wedge} = (\{\wedge\}, \{(\wedge, 3)\})$. The reader may verify that

$$\mathcal{W}^{\mathrm{m}}_{1,\triangle} = \left(\left\{\overline{\underline{\times}}\right\}, \left\{(\overline{\underline{\times}}, 1)\right\}\right), \ \mathcal{W}^{\mathrm{m}}_{2,\triangle} = \left(\left\{\boxslash\right\}, \left\{(\boxslash, 1)\right\}\right), \ \mathcal{W}^{\mathrm{m}}_{3,\triangle} = \left(\left\{\triangle\right\}, \left\{(\triangle, 1)\right\}\right)$$

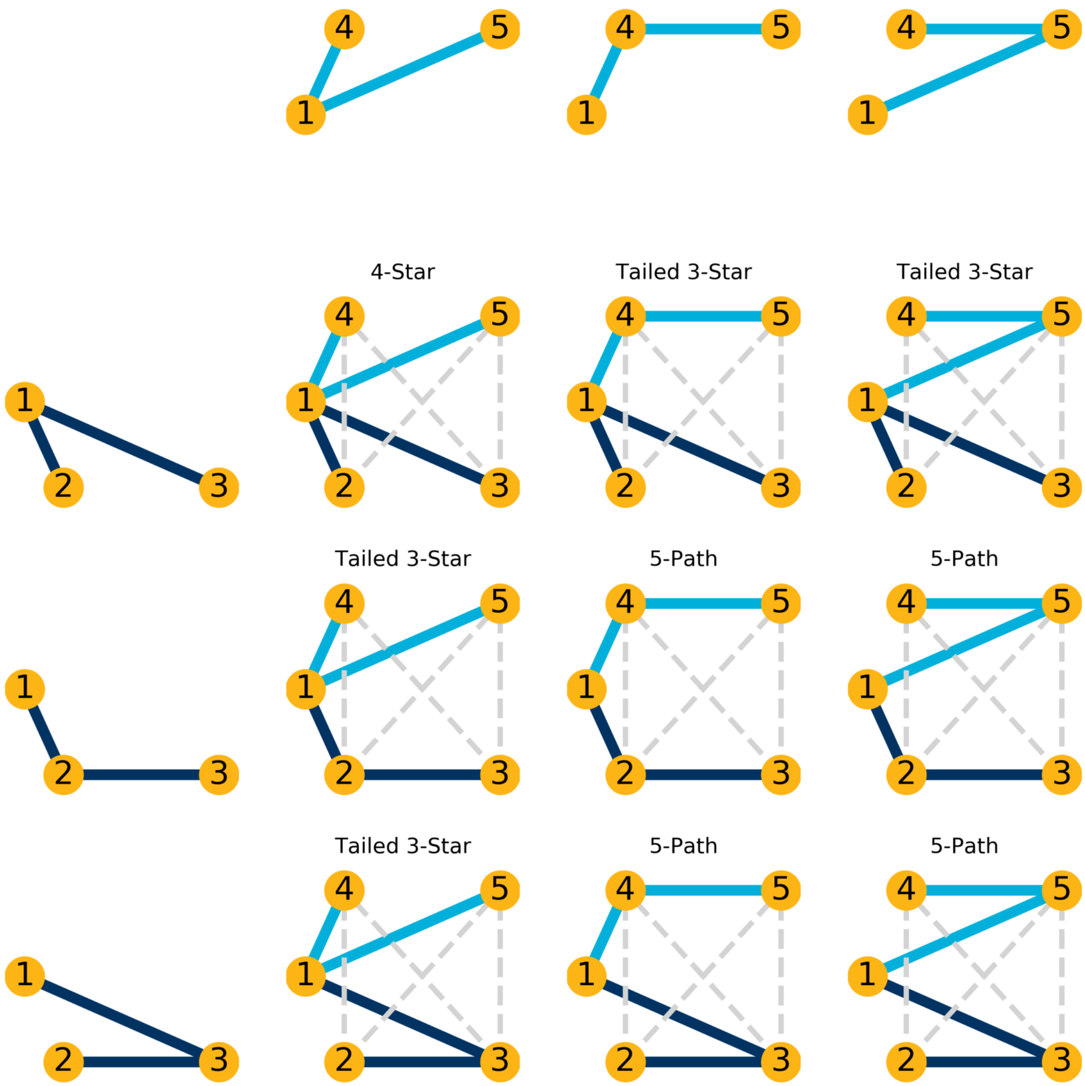
as well as that

$$\mathcal{W}^{\mathrm{m}}_{1,\wedge,\triangle} = \left(\left\{\overline{\times}, \overline{\underline{\times}}\right\}, \left\{(\overline{\times}, 1), (\overline{\underline{\times}}, 2)\right\}\right), \ \mathcal{W}^{\mathrm{m}}_{2,\wedge,\triangle} = \left(\left\{\boxslash\right\}, \left\{(\boxslash, 2)\right\}\right), \ \mathcal{W}^{\mathrm{m}}_{3,\wedge,\triangle} = \emptyset.$$

These multi-sets will be used to study the covariance of $(P_N(\wedge), P_N(\triangle))'$ as well as the variance of the transitivity index.

At the risk of overkill, the following calculations illustrate how the two stitching probability definitions, equations (86) and (87), coincide. For the two-star example, starting with

Figure 5: Stitchings of two-star graphlets with one common node



NOTES: Depiction of all possible ways to join (or "stitch") a pair of two-star (⋀) subgraphs together with one node in common. Each of the resulting subgraphs is a pentad wiring. The dashed gray edges involve pairs of nodes that are not common across the pair of two-stars. Hence the subgraph induced by the five nodes in the pentad may or may not include these edges. The set $\mathcal{W}_{1,\wedge}$ has $\left|\text{iso}\left(\wedge\right)\right| \times \left|\text{iso}\left(\wedge\right)\right| = 9$ elements.

SOURCE: Author's calculations.

equation (86), I get

$$
\begin{aligned}
\Xi\left(\mathcal{W}_{1,\wedge}\right) =& \frac{\Pr\left(\wedge \cong G_N\left[\{1,2,3\}\right] \;\&\; \wedge \cong G_N\left[\{1,4,5\}\right]\right)}{\left|\mathrm{iso}\left(\wedge\right)\right|^2} \\
=& \frac{1}{\left|\mathrm{iso}\left(\wedge\right)\right|^2}\mathbb{E}\left[\left\{D_{12}D_{13}\left(1-D_{23}\right)+D_{12}\left(1-D_{13}\right)D_{23}+\left(1-D_{12}\right)D_{13}D_{23}\right\}\right. \\
& \left.\times\left\{D_{14}D_{15}\left(1-D_{45}\right)+D_{14}\left(1-D_{15}\right)D_{45}+\left(1-D_{14}\right)D_{15}D_{45}\right\}\right] \\
=& \frac{1}{\left|\mathrm{iso}\left(\wedge\right)\right|^2}\left\{\mathbb{E}\left[D_{12}D_{13}\left(1-D_{23}\right)D_{14}D_{15}\left(1-D_{45}\right)\right]\right| \\
& + 4\mathbb{E}\left[D_{12}D_{13}\left(1-D_{23}\right)D_{14}\left(1-D_{15}\right)D_{45}\right] \\
& + 4\mathbb{E}\left[D_{12}\left(1-D_{13}\right)D_{23}D_{14}\left(1-D_{15}\right)D_{45}\right]\right\} \\
=& \frac{1}{\left|\mathrm{iso}\left(\wedge\right)\right|^2}\left[\nu_{1,\wedge}\left(\times\right)\Pr\left(\times = G_N\left[\{1,2,3\}\right]\cup G_N\left[\{1,4,5\}\right]\right)\right. \\
& + \nu_{1,\wedge}\left(\underline{\times}\right)\Pr\left(\underline{\times} = G_N\left[\{1,2,3\}\right]\cup G_N\left[\{1,4,5\}\right]\right) \\
& \left. + \nu_{1,\wedge}\left(\Sigma\right)\Pr\left(\Sigma = G_N\left[\{1,2,3\}\right]\cup G_N\left[\{1,4,5\}\right]\right)\right] \\
=& \sum_{W\in\mathcal{W}^{\mathrm{m}}_{1,\wedge}} \nu_{1,\wedge}\left(W\right)\Pr\left(W = G_N\left[\{1,2,3\}\right]\cup G_N\left[\{1,4,5\}\right]\right) \\
=& \Xi\left(\mathcal{W}^{\mathrm{m}}_{1,\wedge}\right).
\end{aligned}
\tag{88}
$$

The third equality follows from relationships like $\mathbb{E}\left[D_{12}D_{13}\left(1-D_{23}\right)D_{14}\left(1-D_{15}\right)D_{45}\right] = \mathbb{E}\left[D_{12}D_{13}\left(1-D_{23}\right)\left(1-D_{14}\right)D_{15}D_{45}\right]$, which allow for the grouping together of terms. The balance of the equalities are consequences of the definitions given above.

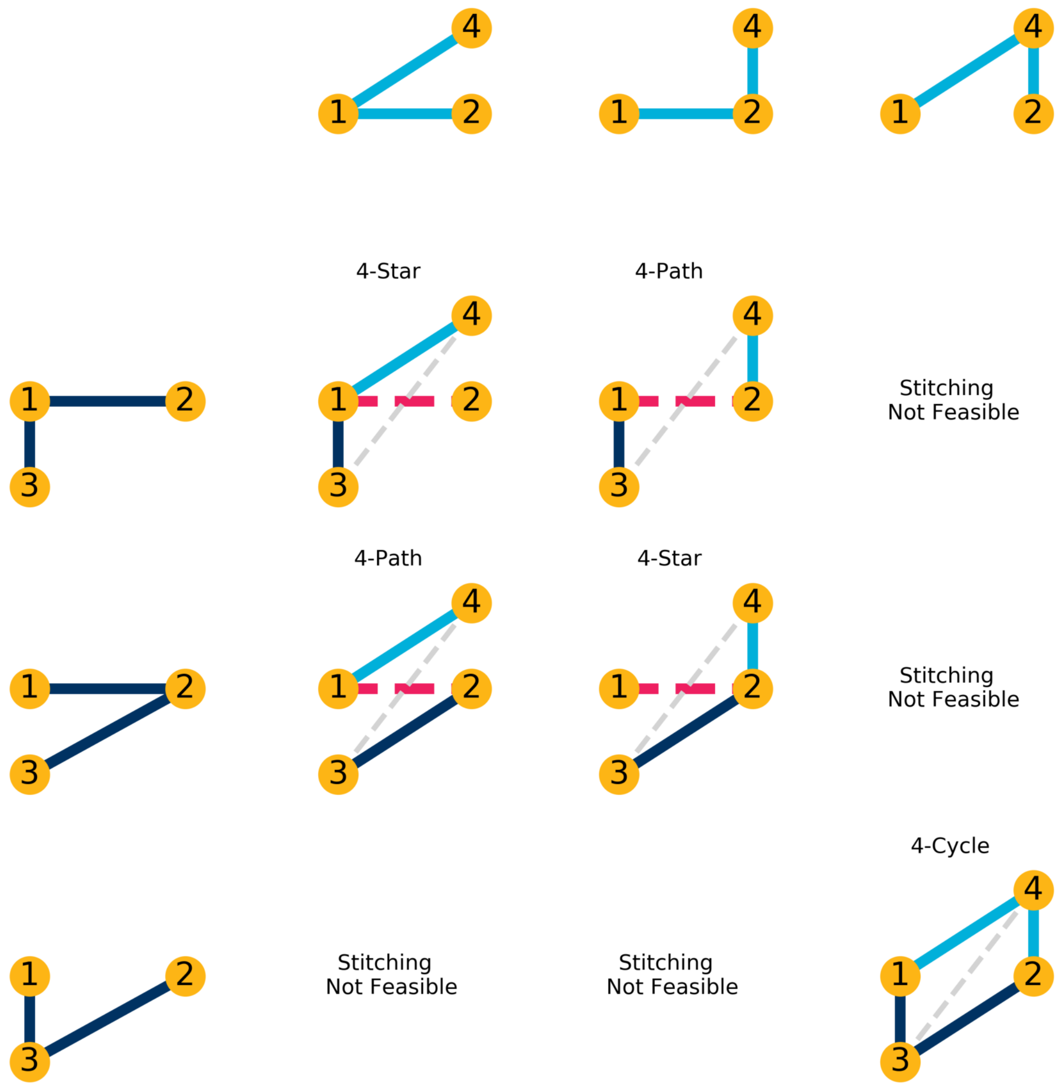
Sampling variances

With the above notations in hand, I now calculate the sampling variances of $\hat{P}(\wedge)$ and $P_N(\triangle)$ as well as their covariance. Holland & Leinhardt (1970, 1976) were the first to derive variance expressions for subgraph counts. The specific development presented here follows Bhattacharya & Bickel (2015). A Hoeffding (1948) variance-composition gives

$$
\mathbb{V}\left(\begin{pmatrix} P_N\left(\triangle\right) \\ P_N\left(\wedge\right)\end{pmatrix}\right) = \binom{N}{3}^{-2}\sum_{q=0}^{3}\binom{N}{3}\binom{3}{q}\binom{N-3}{3-q}\begin{pmatrix}\Sigma_q\left(\triangle\right) & \Sigma_q\left(\triangle,\wedge\right) \\ \Sigma_q\left(\triangle,\wedge\right) & \Sigma_q\left(\wedge\right)\end{pmatrix}
$$

with $\Sigma_q\left(\triangle\right)$, $\Sigma_q\left(\wedge\right)$ and $\Sigma_q\left(\triangle,\wedge\right)$ as defined by (85) above (using the shorthand $\Sigma_q\left(S,S\right) = \Sigma_q\left(S\right)$ etc). Using the fact that each of these variances and covariances is

Figure 6: Stitchings of two-star graphlets with two common nodes



NOTES: Depiction of all possible ways to join (or "stitch") a pair of two-star (⋀) subgraphs together with two nodes in common. Each of the resulting subgraphs is a tetrad wiring. The dashed gray edges involve the pair of nodes that is not common across the pair of two-stars. Hence the subgraph induced by the four nodes in the tetrad may or may not include this edge.
SOURCE: Author's calculations.

zero when $q=0$ and reorganizing terms gives

$$\mathbb{V}\left(\left(\begin{array}{c} P_N(\triangle) \\ P_N(\wedge) \end{array}\right)\right) = \binom{N}{3}^{-2}\sum_{q=1}^{3}\binom{N}{3}\binom{3}{q}\binom{N-3}{3-q}\left[\begin{array}{cc} \Xi\left(\mathcal{W}_{q,\triangle}\right) & \Xi\left(\mathcal{W}_{q,\triangle,\wedge}\right) \\ \Xi\left(\mathcal{W}_{q,\triangle,\wedge}\right) & \Xi\left(\mathcal{W}_{q,\wedge}\right) \end{array}\right]$$
$$-\left[1-\frac{(N-3)!^2}{N!\,(N-6)!}\right]\left[\begin{array}{cc} P(\triangle)^2 & P(\triangle)P(\wedge) \\ P(\triangle)P(\wedge) & P(\wedge)^2 \end{array}\right].$$

In what follows I assume that the network generating process is such that, for each N, $\Sigma_q(\triangle)$ and $\Sigma_q(\wedge)$ are not identically equal to zero for $q \geq 1$. This prevents $P_N(\triangle)$ and $P_N(\wedge)$ from exhibiting degenerate U-Statistic-like attributes (c.f., Graham, 2017, Theorem 1). The restriction is a real one, ruling out the Erdös-Renyi case. Separate results for this special case are presented below.

As introduced earlier, in order to accommodate sequences of networks with varying degrees of sparsity, we can index the underlying population graphon by N, setting $h_N(u,v) = \rho_N w(u,v)$ with $w(u,v) = f_{U_i,U_j|D_{ij}}(u,v|\, D_{ij}=1)$ and allowing $\rho_N \to 0$ as $N \to \infty$. Under such a sequence of GGPs $P(\triangle)$ and $P(\wedge)$ will tend to zero. In order to understand the properties of $P_N(\triangle)$ vis-a-vis $P(\triangle)$ we must normalize. It is natural to normalize according to the number edges in the subgraph under consideration.

Let $\tilde{P}(\triangle) = P(\triangle)/\rho_N^3$, $\tilde{P}(\wedge) = P(\wedge)/\rho_N^2$, $\tilde{P}_N(\triangle) = P_N(\triangle)/\rho_N^3$ and so on. So normalizing I get

$$\mathbb{V}\left(\left(\begin{array}{c} \tilde{P}_N(\triangle) \\ \tilde{P}_N(\wedge) \end{array}\right)\right) = \binom{N}{3}^{-2}\sum_{q=1}^{3}\binom{N}{3}\binom{3}{q}\binom{N-3}{3-q}\left[\begin{array}{cc} \rho_N^{-6}\Xi\left(\mathcal{W}_{q,\triangle}\right) & \rho_N^{-5}\Xi\left(\mathcal{W}_{q,\triangle,\wedge}\right) \\ \rho_N^{-5}\Xi\left(\mathcal{W}_{q,\triangle,\wedge}\right) & \rho_N^{-4}\Xi\left(\mathcal{W}_{q,\wedge}\right) \end{array}\right]$$
$$-\left[1-\frac{(N-3)!^2}{N!\,(N-6)!}\right]\left[\begin{array}{cc} \tilde{P}(\triangle)^2 & \tilde{P}(\triangle)\tilde{P}(\wedge) \\ \tilde{P}(\triangle)\tilde{P}(\wedge) & \tilde{P}(\wedge)^2 \end{array}\right]. \quad (89)$$

Expression (89) agrees with the corresponding expression of Bhattacharya & Bickel (2015) for injective homomorphism frequencies (Equation (3.8), p. 2395).[31] The main difference is the analog of $\Xi\left(\mathcal{W}_{q,\triangle}\right)$ in their expression is itself an injective homomorphism density, whereas here $\Xi\left(\mathcal{W}_{q,\triangle}\right)$ is neither an injective homomorphism nor an induced subgraph density and instead involves checking for particular patterns of *both* adjacency and non-adjacency as described above.

[31] See also Green & Shalizi (2017, Lemma 1).

Rates of convergence

To understand the rate of convergence in mean square of, for example, $\tilde{P}_N(\triangle)$ toward $\tilde{P}(\triangle)$, we need to determine the order of each of the terms in (89). Let $e(R) \stackrel{def}{\equiv} |\mathcal{E}(R)|$ and $e(S) \stackrel{def}{\equiv} |\mathcal{E}(S)|$ denote the number of edges in graphlets R and S. Next observe that $\binom{N}{p}^{-1}\binom{p}{q}\binom{N-p}{p-q} = O(N^{-q})$. We therefore have that the terms in the summation indexed by q in (89) are $O\left(N^{-q}\rho_N^{-e(R)}\rho_N^{-e(S)}\right)O(\Xi(\mathcal{W}_{q,R,S}))$ for $q = 1, \ldots, p$. I divide these terms, closely following Bhattacharya & Bickel (2015), into three cases:

<u>**Case 1 ($q = 1$):**</u> when $q = 1$ the number of edges in all elements of $\mathcal{W}_{q,R,S}$ equals $e(R) + e(S)$ for any subgraphs R and S. Hence $O(\Xi(\mathcal{W}_{1,R,S})) = O\left(\rho_N^{e(R)}\rho_N^{e(S)}\right)$, yielding

$$O\left(N^{-1}\rho_N^{-e(R)}\rho_N^{-e(S)}\right)O(\Xi(\mathcal{W}_{1,R,S})) = O\left(N^{-1}\right).$$

The $q = 1$ summand in (89) is of order N^{-1}. In general, from the theory of U-statistics, one would expect this to be the leading variance term; however, the present situation is more complicated.

<u>**Case 2 ($q = 3$ or $q = p$):**</u> In this case the order of $\Xi\left(\mathcal{W}_{3,\triangle}\right)$ is $O(\rho_N^3)$, $\Xi\left(\mathcal{W}_{3,\wedge}\right)$ is $O(\rho_N^2)$ and $\mathcal{W}_{3,\triangle,\wedge}$ is empty so that $\Xi\left(\mathcal{W}_{3,\triangle,\wedge}\right) = 0$. Therefore, recalling that $\lambda_N = (N-1)\rho_N$ equals average degree,

$$O\left(N^{-3}\rho_N^{-2e(\triangle)}\right)O\left(\Xi\left(\mathcal{W}_{3,\triangle}\right)\right) = O\left(\lambda_N^{-3}\right)$$
$$O\left(N^{-3}\rho_N^{-2e(\wedge)}\right)O\left(\Xi\left(\mathcal{W}_{3,\wedge}\right)\right) = O\left(N^{-1}\lambda_N^{-2}\right)$$
$$O\left(N^{-3}\rho_N^{-e(\triangle)}\rho_N^{-e(\wedge)}\right)O\left(\Xi\left(\mathcal{W}_{3,\triangle,\wedge}\right)\right) = o(1).$$

<u>**Case 3 ($q = 2$ or ($2 \le q \le p-1$)):**</u> Here the order of $\Xi\left(\mathcal{W}_{2,\triangle}\right)$ equals $O\left(\rho_N^{2e(\triangle)-(q-1)}\right) = O(\rho_N^5)$, $\Xi\left(\mathcal{W}_{2,\wedge}\right)$ equals $O\left(\rho_N^{2e(\wedge)-(q-1)}\right) = O(\rho_N^3)$ and that of

$\Xi\left(\mathcal{W}_{2,\triangle,\wedge}\right)$ equals $O\left(\rho_N^{e(\triangle)+e(\wedge)-(q-1)}\right) = O\left(\rho_N^4\right)$. Therefore

$$O\left(N^{-2}\rho_N^{-2e(\triangle)}\right)O\left(\Xi\left(\mathcal{W}_{2,\triangle}\right)\right) = O\left(N^{-1}\lambda_N^{-1}\right)$$
$$O\left(N^{-2}\rho_N^{-2e(\wedge)}\right)O\left(\Xi\left(\mathcal{W}_{2,\wedge}\right)\right) = O\left(N^{-1}\lambda_N^{-1}\right)$$
$$O\left(N^{-2}\rho_N^{-e(\triangle)}\rho_N^{-e(\wedge)}\right)O\left(\Xi\left(\mathcal{W}_{2,\triangle,\wedge}\right)\right) = O\left(N^{-1}\lambda_N^{-1}\right).$$

For the two variance terms we have

$$\mathbb{V}\left(\tilde{P}_N(\triangle)\right) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{N\lambda_N}\right) + O\left(\frac{1}{\lambda_N^3}\right)$$
$$\mathbb{V}\left(\tilde{P}_N(\wedge)\right) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{N\lambda_N}\right) + O\left(\frac{1}{N\lambda_N^2}\right)$$

indicating that the rate at which, for example, $\tilde{P}_N(\triangle)$ converges in mean square toward $\tilde{P}(\triangle)$, depends on the behavior of average degree as the network grows large. This reflects the fact that, depending on a combination of the nature of the graphlet of interest and the rate at which λ_N does, or does not, grows with N, several of the terms in (89) may be of equal order.

For any increasing sequence of average degree we have

$$\mathbb{V}\left(\tilde{P}_N(\triangle)\right) = O\left(\max\left(\frac{1}{N}, \frac{1}{\lambda_N^3}\right)\right)$$
$$\mathbb{V}\left(\tilde{P}_N(\wedge)\right) = O\left(\frac{1}{N}\right)$$
$$\mathbb{C}\left(\tilde{P}_N(\triangle), \tilde{P}_N(\wedge)\right) = O\left(\frac{1}{N}\right).$$

If $\lambda_N \geq CN^{1/3}$, then the rate of convergence is $\sqrt{N}$ for both $\tilde{P}_N(\wedge)$ and $\tilde{P}_N(\triangle)$. In the sparse case, with $\lambda_N \to \lambda$, $\tilde{P}(\wedge)$, due to the acyclic structure of the two-star graphlet, remains estimable at the $\sqrt{N}$ rate. However in this case all three of its variance terms are of equal order. In contrast $\tilde{P}(\triangle)$ is (evidently) not consistently estimable in the sparse case.

Asymptotic normality

When average degree is $\lambda_N > CN^{1/3}$, such that both $\tilde{P}_N(\wedge)$ and $\tilde{P}_N(\triangle)$ converge at the $\sqrt{N}$ rate, an application of Theorem 1.c of Bickel et al. (2011) establishes that, under some regularity conditions,

$$\sqrt{N}\left(\begin{array}{c}\tilde{P}_N\left(\triangle\right)-\tilde{P}\left(\triangle\right)\\ \tilde{P}_N\left(\wedge\right)-\tilde{P}\left(\wedge\right)\end{array}\right)\stackrel{D}{\rightarrow}\mathcal{N}\left(\left(\begin{array}{c}0\\0\end{array}\right),9\left(\begin{array}{cc}\tilde{\Sigma}_1\left(\triangle\right) & \tilde{\Sigma}_1\left(\triangle,\wedge\right)\\ \tilde{\Sigma}_1\left(\triangle,\wedge\right) & \tilde{\Sigma}_1\left(\wedge\right)\end{array}\right)\right), \tag{90}$$

where $\tilde{\Sigma}_1\left(\triangle\right)=\rho_N^{-6}\Sigma_1\left(\triangle\right)$, $\tilde{\Sigma}_1\left(\triangle,\wedge\right)=\rho_N^{-5}\Sigma_1\left(\triangle,\wedge\right)$ etc. Proving (90) is relatively straightforward. I do not sketch the argument here, but note that the main tools needed were already introduced in the analysis of dyadic regression appearing in Section 4 above.

As noted previously, if $\lambda_N\rightarrow\lambda$ as $N\rightarrow\infty$ such that the network is sparse in the limit, then a general result on $\sqrt{N}\left(\tilde{P}_N\left(\triangle\right)-\tilde{P}\left(\triangle\right)\right)$ is unavailable. In contrast, part (b) of Theorem 1 in Bickel et al. (2011) implies that not only does $\tilde{P}_N\left(\wedge\right)$ remain $\sqrt{N}$ consistent for $\tilde{P}\left(\wedge\right)$ in this case, but also that $\sqrt{N}\left(\tilde{P}_N\left(\wedge\right)-\tilde{P}\left(\wedge\right)\right)$ remains asymptotically normal. The limiting variance in this case differs from the one given in (90); all terms in $\mathbb{V}\left(\tilde{P}_N\left(\wedge\right)\right)$ are of equal order (and hence should be retained).

More generally the sampling properties of induced subgraph frequencies under sparse graph limits remains relatively unexplored. The sensitivity of rates of convergence and distributional properties to assumptions about λ_N raises concerns about uniformity of inference procedures. A similar concern is suggested by the properties of these statistics when the graphon is constant. This last case is considered next.

Two-star and triangle counts in Erdös-Renyi networks

The analysis above assumes that the graphon is such that $\mathbb{C}\left(\mathbf{1}\left(R\cong G_N\left[\mathbf{i}_p\right]\right),\mathbf{1}\left(S\cong G_N\left[\mathbf{j}_p\right]\right)\right)\neq 0$ when $\mathbf{i}_p$ and $\mathbf{j}_p$ share exactly one index in common (such that $\Xi\left(\mathcal{W}_{1,S}\right)-P\left(S\right)^2>0$). This condition will generally hold for graphons which vary in u and v (such that the events $D_{12}=1$ and $D_{13}=1$ are not independent), but it does rule out the Erdös-Renyi case (where links form independently with constant probability ρ).[32] This graph generation process has been extensively studied by probabilists for over sixty years (e.g., Janson et al., 2000).

In statistics, Janson & Nowicki (1991) and Nowicki (1991) studied the sampling properties of induced and partial subgraph frequencies when the network is an Erdös-Renyi one. They demonstrated asymptotic normality of such frequencies with a $\sqrt{\binom{N}{2}}$ rate of convergence. These earlier results, at first glance, appear to be in tension with the more general results of Bickel et al. (2011), who showed asymptotic normality with a $\sqrt{N}$ rate of convergence under general graphons. It turns out, however, that the leading (i.e., $q=1$) term in (89) is

[32]See Menzel (2017) for more examples of degenerate graphons.

identically equal to zero under the Erdös-Renyi GPP. The Erdös-Renyi GPP is a "degenerate" special case.

To see this, evaluate the stitching probabilities (87) under the Erdös-Renyi GPP to get

$$\Xi\left(\mathcal{W}_{1,\wedge}\right)=\rho^{4}\left(1-\rho\right)^{2},\ \Xi\left(\mathcal{W}_{2,\wedge}\right)=\frac{4}{9}\rho^{3}\left(1-\rho\right)^{2}+\frac{1}{9}\rho^{4}\left(1-\rho\right),\ \Xi\left(\mathcal{W}_{3,\wedge}\right)=\frac{1}{3}\rho^{2}\left(1-\rho\right)$$

and

$$\Xi\left(\mathcal{W}_{1,\triangle}\right)=\rho^{6},\ \Xi\left(\mathcal{W}_{2,\triangle}\right)=\rho^{5},\ \Xi\left(\mathcal{W}_{3,\triangle}\right)=\rho^{3}$$

and

$$\Xi\left(\mathcal{W}_{1,\wedge,\triangle}\right)=\rho^{5}\left(1-\rho\right),\ \Xi\left(\mathcal{W}_{2,\wedge,\triangle}\right)=\frac{2}{3}\rho^{4}\left(1-\rho\right),\ \Xi\left(\mathcal{W}_{3,\wedge,\triangle}\right)=0.$$

Under these graphlet stitching probabilities the $q=1$ variance term, which is generally the leading variance term in Bickel et al. (2011, Theorem 1), instead equals

$$\begin{pmatrix}\Sigma_{1}\left(\triangle\right) & \Sigma_{1}\left(\triangle,\wedge\right)\\ \Sigma_{1}\left(\triangle,\wedge\right) & \Sigma_{1}\left(\wedge\right)\end{pmatrix}=\mathbf{0}_{2}\mathbf{0}_{2}'.$$

Hence, under the (dense) Erdös-Renyi GPP, the leading variance term is instead the $q=2$ one, yielding for $0<\rho<1$ but $\rho\neq 2/3$,

$$\sqrt{\binom{N}{2}}\begin{pmatrix}P_{N}\left(\triangle\right)-P\left(\triangle\right)\\ P_{N}\left(\wedge\right)-P\left(\wedge\right)\end{pmatrix}\xrightarrow{D}\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},9\begin{pmatrix}\Sigma_{2}\left(\triangle\right) & \Sigma_{2}\left(\triangle,\wedge\right)\\ \Sigma_{2}\left(\triangle,\wedge\right) & \Sigma_{2}\left(\wedge\right)\end{pmatrix}\right) \tag{91}$$

where

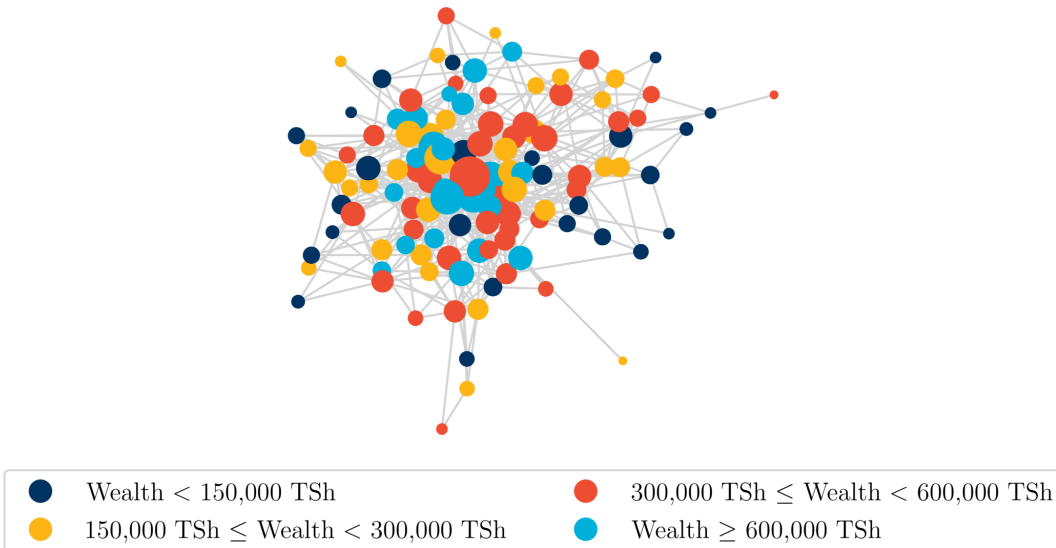
$$\begin{pmatrix}\Sigma_{2}\left(\triangle\right) & \Sigma_{2}\left(\triangle,\wedge\right)\\ \Sigma_{2}\left(\triangle,\wedge\right) & \Sigma_{2}\left(\wedge\right)\end{pmatrix}=\rho^{3}\left(1-\rho\right)\begin{pmatrix}\rho^{2} & \frac{1}{3}\rho\left(2-3\rho\right)\\ \frac{1}{3}\rho\left(2-3\rho\right) & \frac{1}{9}\left(2-3\rho\right)^{2}\end{pmatrix}.$$

See Corollaries 2 and 4 of Nowicki (1991) for additional context and references to the primary literature.

Uniform Inference

The analysis of the previous two subsections showed how the limiting distributions of two-star and triangle frequencies are sensitive to the form of the graphon, $h_{N}\left(u,v\right)=\rho_{N}w\left(u,v\right)$. If ρ_{N} approaches zero too quickly, or $w\left(u,v\right)$ is a constant, the rate of convergence of the

Figure 7: Nyakatoke risk-sharing network



Sources: De Weerdt (2004) and authors' calculations. 'TSh' is an abbreviation for Tanzanian Shillings.

estimator changes. This raises concerns about how to conduct inference in settings where the limiting graph is 'close to sparse' and/or the graphon is 'nearly' constant, or equivalently, dependence across dyads sharing agents in common is weak. In such settings an approach to inference based on (90), may have poor properties when N is finite. This is because the $q = 1$, $q = 2$ and $q = 3$ terms in the variance expression (89) may all be of similar order. For this reason, it seems advisable to keep all terms when calculating variances for test statistics. Clearly, there are open questions on how best to undertake testing in this setting.

Application of results to inference on transitivity in Nyakatoke

De Weerdt (2004) collected information of risk-sharing relationships across 119 houses in Nyakatoke, a small village in Tanzania (see Figure 8). The density of this network is 0.0698, while its transitivity index is 0.1884, nearly three times a large. A natural question is whether the high transitivity index simple reflects "chance" or is a real feature of Nyakatoke. To assess this I construct a confidence interval for the transitivity index using the delta method and the results outlined above. Other than the empirical illustration included in Bhattacharya & Bickel (2015), I am aware of no other published examples of large network inference on the transitivity index.

The natural analog estimates of $\Xi\left(\mathcal{W}_{1,\triangle}\right)$, $\Xi\left(\mathcal{W}_{1,\wedge}\right)$, and $\Xi\left(\mathcal{W}_{1,\triangle,\wedge}\right)$ involve summations over all $\binom{N}{3}\binom{3}{1}\binom{N-3}{3-1} = 30 \times \binom{N}{5}$ *pairs* of triads sharing exactly one common agent. This requires evaluating the configuration of all $\binom{N}{5}$ *pentads* in the network; a computationally non-trivial task even for medium-sized networks.[33] It is for this reason that Bhattacharya & Bickel (2015) suggest a subsampling approach to variance estimation.

For the Nyakatoke network we have a total of $\binom{119}{3} = 273,819$ triad configurations to count and a total of $\binom{119}{5} = 182,637,273$ pentads that need to be inspected in order to calculate variances. These are large numbers, but nevertheless small enough enough for a desktop computer to handle in a few minutes. Direct calculation gives

$$P_N\left(\triangle\right) = \begin{array}{c} 0.00115 \\ (0.00030) \end{array}, \; P_N\left(\wedge\right) = \begin{array}{c} 0.00496 \\ (0.00100) \end{array}$$

These standard errors include estimates of both the first *and* last terms in (89) above, although the second of these is asymptotically negligible as long as average degree grows fast enough (which is assumed for the asymptotic normality result).

Applying the delta method I get an estimated standard error for the transitivity index of 0.011; this suggests that transitivity is significantly greater than what we would expect to observe under the Erdös-Renyi random graph null.

7.2 Moments of the degree distribution

Networks are complex objects, making their analysis both conceptually and technically challenging. One approach to simplification involves looking only at the number of links each agent has, that is their degree, $D_{i+} = \sum_{j \neq i} D_{ij}$, ignoring all other architectural features of the network. Indeed, a substantial empirical literature focuses on the degree sequence of a network as its primary object of interest (Barabási & Albert, 1999; Barabási, 2016).

Most real world networks exhibit substantial degree heterogeneity, making the degree sequence an interesting statistic to study and model. A network's degree sequence is also straightforward to measure. A researcher need only ask about the number of friends, suppliers, or partners each agent has, not their identity. Many general purpose datasets collect such information. For example General Social Survey (GSS) sometimes collects information on the number of close confidants (cf., Marsden, 1987; McPherson et al., 2006), while demographers routinely collect information on the number of lifetime and/or concurrent sexual partners. Simplicity and data availability both drive the substantial focus on degree distributions in empirical work.

[33] For each pentad we look at the thirty pairs of triads that can be constructed from it, such that the two triads share exactly one agent in common.

It is possible for two graphs with the same degree sequence to be topologically different; their diameters and transitivity indices, for example, may differ substantially. At the same time a network's degree sequence is an important summary statistic, constraining other features of it, even local ones. This is shown by Theorem 3, which I believe is an original result (albeit perhaps folk wisdom; cf., Snijders et al., 2006).

Theorem 3. (Degree Sequence Moments) *Let G be an exchangeable random graph of order N. The m^{th} moment of $D_{i+} = \sum_{j \neq i} D_{ij}$ equals*

$$\mathbb{E}\left[D_{i+}^{m}\right] = \sum_{k=1}^{m} C_{k,m} \times \mathbb{E}\left[D_{ij_1} \times \cdots \times D_{ij_k}\right]$$

for $m = 1, 2, \ldots, N-1$ and $C_{k,m} = \binom{N-1}{k} \left(\sum_{\mathbf{p} \in \mathcal{P}_{k,m}} \frac{m!}{p_1! \times \cdots \times p_k!}\right)$ with

$$\mathcal{P}_{k,m} = \left\{ (p_1, \ldots, p_k) \,:\, \sum_{j=1}^{k} p_j = m,\ p_j \in \mathbb{N} \text{ for } j = 1, \ldots, k \right\}$$

and $\mathbb{N}$ the set of positive integers.

Proof. See Appendix A. □

Theorem 3 implies that the first four uncentered moments of D_{i+} equal

$$\mathbb{E}\left[D_{i+}\right] = (N-1)\, P\left(\relbar\right) \quad (92)$$

$$\mathbb{E}\left[D_{i+}^{2}\right] = (N-1)\, P\left(\relbar\right) + (N-1)(N-2)\, Q\left(\wedge\right) \quad (93)$$

$$\begin{aligned} \mathbb{E}\left[D_{i+}^{3}\right] = &(N-1)\, P\left(\relbar\right) + 3(N-1)(N-2)\, Q\left(\wedge\right) \\ &+ (N-1)(N-2)(N-3)\, Q\left(\nwarrow\right) \end{aligned} \quad (94)$$

$$\begin{aligned} \mathbb{E}\left[D_{i+}^{4}\right] = &(N-1)\, P\left(\relbar\right) + 7(N-1)(N-2)\, Q\left(\wedge\right) \\ &+ 6(N-1)(N-2)(N-3)\, Q\left(\nwarrow\right) \\ &+ (N-1)(N-2)(N-3)(N-4)\, Q\left(\times\right). \end{aligned} \quad (95)$$

In dense networks it is natural to divide D_{i+} by $N-1$. With degrees so normalized, all terms in equations (92) to (95) are asymptotically dominated by the last one as $N \to \infty$. Hence, in the limit, the k^{th} moment of normalized degree equals the injective homomorphism density of k-stars in the limiting graphon (cf., Diaconis et al., 2008, Lemma 4.1). In the dense case we have, for example, that

$$\lim_{N\to\infty} \mathbb{V}\left(\frac{D_{i+}}{N-1}\right) = Q\left(\wedge\right) - P\left(\text{—}\right)P\left(\text{—}\right). \tag{96}$$

When the network is not dense, the natural normalization is instead by average degree, $\lambda_N = (N-1)\,\rho_N$, which may no longer be proportional to N in the limit. In the sparse case, $\lambda_N \to \lambda$, and all terms in equations (92) to (95) are of equal order. For example, the fourth uncentered moment of D_{i+}/λ_N equals, in a sparse limit,

$$\lim_{N\to\infty} \mathbb{E}\left[\left(\frac{D_{i+}}{\lambda_N}\right)^4\right] = \frac{\tilde{P}\left(\text{—}\right)}{\lambda^3} + \frac{7\tilde{Q}\left(\wedge\right)}{\lambda^2} + \frac{6\tilde{Q}\left(\nwarrow\right)}{\lambda} + \tilde{Q}\left(\times\right),$$

where a tilde above a subgraph/homomorphism density, as earlier, denotes the density divided by ρ_N raised to the power of the number of edges in the subgraph under consideration (e.g., $\tilde{Q}\left(\wedge\right) = Q\left(\wedge\right)/\rho_N^2$).

From (92) and (93) have that, in the sparse case,

$$\lim_{N\to\infty} \mathbb{V}\left(\frac{D_{i+}}{\lambda_N}\right) = \left[\tilde{Q}\left(\wedge\right) - \tilde{P}\left(\text{—}\right)\tilde{P}\left(\text{—}\right)\right] + \frac{\tilde{P}\left(\text{—}\right)}{\lambda}. \tag{97}$$

There are several peculiarities in these expressions. Returning to the dense case, when the graph is an Erdös-Renyi homogenous random graph, $Q\left(\wedge\right) = P\left(\text{—}\right)P\left(\text{—}\right)$, and (96) indicates that the distribution of $D_{i+}/\left(N-1\right)$ is degenerate in the limit. In that case normalizing D_{i+} by $\sqrt{N-1}$ results in a random variable with a non-degenerate variance in the limit since

$$\mathbb{V}\left(D_{i+}\right) = (N-1)(N-2)\left[Q\left(\wedge\right) - P\left(\text{—}\right)P\left(\text{—}\right)\right] + (N-1)\,P\left(\text{—}\right)\left(1 - P\left(\text{—}\right)\right).$$

Observations such as these suggest that, as with subgraph frequencies, it may be desirable to retain all terms – including nominally asymptotically dominated ones – when calculating the variance and other moments of the degree distribution.

Atalay et al. (2011) construct a theoretical model of supply chain formation. They informally assess the plausibility of a calibrated version of their model by comparing their model-predicted degree sequence with the one observed in the US Buyer-Supplier network (see their Figure 1). A formal minimum χ^2 type specification test of their model could be constructed on the basis of Theorem 3.

7.3 Further reading and open questions

Subgraph frequencies are, in many ways, analogous to moments of a distribution. Relatedly methods of estimation and inference for subgraph frequencies have many applications, from attaching a measure of uncertainty to statistics like the transitivity index, to facilitating specification testing and model estimation. As the discussion here shows, the large network properties of empirical subgraph frequencies depend on the nature and magnitude of dependence across links induced by the graphon as well as properties like sparsity. Formulating methods of inference for subgraph frequencies that are adaptive to these features of the GGP would be useful. Menzel (2017) makes some progress in this direction, but substantial work remains.

Volfosky & Airoldi (2016), extending results due to Diaconis & Freedman (1980), present results relating finitely and infinitely exchangeable arrays. Results of this type could be useful for understanding how best to proceed when the network in hand corresponds to the equilibrium of an N-player game where the conditional independence structure associated Aldous-Hoover type GGPs may not formally hold, but where – for N large enough – it should hold approximately.

Bhattacharya & Bickel (2015), Green & Shalizi (2017) and Menzel (2017) discuss subsampling and bootstrap methods for exchangeable random arrays. Adapting ideas introduced by, for example, Menzel (2017) to accommodate sparse networks would be theoretically interesting and practically useful.

I have emphasized more recent work on subgraph frequencies, but the earlier papers, beginning with the seminal one by Holland & Leinhardt (1976) are rewarding to read (or re-read) in the light of contemporary developments. The survey paper by Jackson et al. (2017) presents many real work examples of degree distributions and other network statistics. This paper also relates these measures to theoretical ideas in the economic literature on network formation and network games.

A rather different approach to asymptotic analysis of network statistics builds off the probability literature on random geometric graphs (Penrose, 2003). These models posit a strong form of homophily such that agents which are far apart from one another (in some, perhaps latent, space) link infrequently (or not at all). The (latent) spatial structure renders agents non-exchangeable. This mechanism generates sufficient independence among distant units such that LLNs and CLTs can be proven. Leung (2019), Kuersteiner (2019), and Leung & Moon (2019) develop these ideas to prove LLNs and CLTs for network statistics where the observed network is assumed to be a strategic network equilibrium configuration. A challenge of this approach is that valid inference appears to require information on agents' positions (so that HAC type variance estimators can be used). Unfortunately it is often most natural to view such positions as latent (e.g., Hoff et al., 2002; Krivitsky et al., 2009). Nevertheless, this approach, by building on insights from the literature of random geometric graphs, as well as spatial statistics, seems well calibrated to some network applications. For example, sparseness seems to be easier to handle in this framework (cf., Graham, 2016).

Understanding the connections between approaches to large network inference based upon random geometric graphs versus exchangeable random graphs remains, to my knowledge, largely unexplored.

8 Strategic models of network formation

The models of network formation introduced in Sections 4, 5 and 6 are externality free: the utility two agents create by forming a link is invariant to the presence or absence of links elsewhere in the network. In contrast, the theoretical literature on network formation, beginning with the seminal paper by Jackson & Wolinsky (1996), is decidedly focused on the study of models where agents' preferences are interdependent. That is, the utility dyad $\{i, j\}$ generates by forming an edge may vary with the presence or absence of additional edges elsewhere in the network. For example, if i and j share many neighbors ("friends") in common, they may reap utility gains from 'triadic closure' when linking; incidentally also forming many triangles (Simmel, 1908; Coleman, 1988; Jackson et al., 2012).

Models with interdependencies in preferences are typically called *strategic* models of network formation. The use of the word strategic here stems from connections, both historical and substantive, between recent theoretical research on networks and game theory. I will comment on this nomenclature after first introducing the standard notion of equilibrium used by theoretical network researchers in this area: *pairwise stability*. Pairwise stability is the equilibrium concept introduced by Jackson & Wolinsky (1996). Here I introduce the definition which excludes the possibility of transfers between agents; the transferable utility case was introduced in Section 3.

Let $\nu_i : \mathbb{D}_N \to \mathbb{R}$ be a utility function for agent i, which maps adjacency matrices into utils. In order to define pairwise stability I need a definition of *marginal utility*. As earlier, the marginal utility for agent i associated with (possible) edge (i, j) is

$$MU_{ij}(\mathbf{D}) = \begin{cases} \nu_i(\mathbf{D}) - \nu_i(\mathbf{D} - ij) & \text{if } D_{ij} = 1 \\ \nu_i(\mathbf{D} + ij) - \nu_i(\mathbf{D}) & \text{if } D_{ij} = 0 \end{cases} \tag{98}$$

recalling that $\mathbf{D} - ij$ is the adjacency matrix associated with the network obtained after deleting edge (i,j) and $\mathbf{D}+ij$ the one obtained via link addition.

Definition 6. (Pairwise stability without Transfers) The network G is pairwise stable if (i) no agent wishes to dissolve a link

$$\forall (i,j) \in \mathcal{E}(G),\ MU_{ij}(\mathbf{D}) \geq 0 \,\text{and}\, MU_{ji}(\mathbf{D}) \geq 0 \tag{99}$$

and (ii) no pair of agents wishes to form a link

$$\forall (i,j) \notin \mathcal{E}(G),\ MU_{ij}(\mathbf{D}) > 0 \ \Rightarrow\ MU_{ji}(\mathbf{D}) < 0. \tag{100}$$

Two features of Definition 6 merit emphasis. First, an implication of the definition is that utility is *nontransferable* across agents. This differs from some of the models introduced earlier. Second, the strategic moniker aside, pairwise stability is a really non-strategic/myopic notion of equilibrium. This point is elegantly made by Ostrovsky (2008) in a related context. Pairwise stability does not require agents to engage in any "what if" or forward-looking introspection. Specifically it does not require agents to imagine what might happen to the rest of the network were they to add or delete a link, rather it simply requires them to behave optimally given the actions of all other agents in the network. The key feature of so-called strategic models relative to those in Sections 4, 5 and 6 is not behavioral, but in their different assumptions about the nature of preferences. Here utility is interdependent; this is the interesting complication.[34]

For any profile of preferences $\{\nu_i\}_{i=1}^{M}$ there many be many network configurations satisfying Definition 6.[35] The potentially high cardinality of the set of pairwise stable network configurations makes the direct application of econometric methods designed for the analysis of games computationally prohibitive (c.f., Bajari et al., 2010b, 2013). Nevertheless insights from research in this area is valuable for analyzing empirical models of network formation.

8.1 A fixed point approach with increasing preferences

One example of this claim is provided by an elegant and interesting paper by Miyauchi (2016). This paper draws on insights from the theory of supermodular games (e.g., Topkis, 1998) and their empirical analysis (Jia, 2008; Uetake & Watanabe, 2013) to formulate a tractable estimation strategy for a class of strategic network formation models. Following

[34] It might also be interesting to consider estimation and inference under different refinements of the pairwise stability concept; such refinements might posit more sophisticated play by agents.

[35] For results on the existence and uniqueness of pairwise stable networks see Jackson & Watts (2001) and Hellmann (2013).

Miyauchi (2016) consider the mapping $\varphi(\mathbf{D}) : \mathbb{D}_N \to \mathbb{I}_{\binom{N}{2}}$:

$$\varphi(\mathbf{D}) \equiv \begin{bmatrix} \mathbf{1}(MU_{12}(\mathbf{D}) \geq 0)\,\mathbf{1}(MU_{21}(\mathbf{D}) \geq 0) \\ \mathbf{1}(MU_{13}(\mathbf{D}) \geq 0)\,\mathbf{1}(MU_{31}(\mathbf{D}) \geq 0) \\ \vdots \\ \mathbf{1}(MU_{N-1N}(\mathbf{D}) \geq 0)\,\mathbf{1}(MU_{NN-1}(\mathbf{D}) \geq 0) \end{bmatrix}. \tag{101}$$

Observe that $\mathbf{1}(MU_{ij}(\mathbf{D}) \geq 0)\,\mathbf{1}(MU_{ji}(\mathbf{D}) \geq 0)$ equals 1 if condition (99) of pairwise stability holds (which implies edge (i,j) is present) and zero otherwise (which implies condition (100) and hence the absence of edge (i,j)). Under the maintained assumption that the observed network is pairwise stable, its adjacency matrix is therefore the fixed point

$$\mathbf{D} = \mathrm{vech}^{-1}\left[\varphi(\mathbf{D})\right]. \tag{102}$$

Here vech$(\cdot)$ vectorizes the $\binom{N}{2}$ elements in the lower triangle of an $N \times N$ matrix and I define its inverse operator as creating a symmetric matrix with a zero diagonal. There may, of course, be many $\mathbf{d} \in \mathbb{D}_N$ such that $\mathbf{d} = \mathrm{vech}^{-1}[\varphi(\mathbf{d})]$. Miyauchi (2016) notes, however, that if the preference profile $\{\nu_i\}_{i=1}^N$ satisfies what he calls a *non-negative externality* condition, namely that the marginal utilities $MU_{ij}(\mathbf{d})$ are weakly increasing in $\mathbf{d}$ for all i and j, then one can characterize the set of pairwise stable networks with Tarski's (1955) fixed point theorem (Miyauchi, 2016, Proposition 1). The invocation of Tarski (1955) implies that the set of pairwise stable networks corresponds to a complete lattice with a *maximum* and *minimum equilibrium*. Furthermore any pairwise stable network is a partial subgraph, defined on nodes $\{1, \ldots, N\}$ of the maximum equilibrium. And the minimum equilibrium is always a partial subgraph, again defined on nodes $\{1, \ldots, N\}$ of any pairwise stable equilibrium. This has many useful implications. Trivially, the set of equilibrium networks can be sorted according to density; less trivially their degree sequences can also be ordered.

Of course, the non-negative externality requirement is restrictive; there are many settings where diminishing marginal utility in links might be plausible (e.g., capacity constraints). At the same time, many extant empirical models of network formation do satisfy the restriction, so exploring estimation maintaining it is reasonable. Miyauchi (2016, Section 3.3) provides additional discussion.

Again borrowing results from the theory of supermodular games, Miyauchi (2016) shows that the minimum equilibrium, say $\underline{\mathbf{d}}$, can be computed by fixed point iteration of (101) starting from the empty adjacency matrix, while the maximum equilibrium, say $\overline{\mathbf{d}}$, may be computed by fixed point iteration starting from the adjacency matrix associated with the complete graph K_N. A similar computational insight, albeit in non-network settings, features in Jia (2008) and Uetake & Watanabe (2013).

At this stage, to show how the above insights can be used concretely, it is helpful to parameterize the utility function, introducing both explicit heterogeneity and a parameter vector. Adopting the random utility approach pioneered by McFadden (1974), assume, for example, that

$$\nu_i\left(\mathbf{d},\mathbf{U};\theta_0\right)=\sum_j d_{ij}\left[\alpha_0+\beta_0\left[\sum_k d_{ik}d_{jk}\right]-U_{ij}\right], \tag{103}$$

with $\mathbf{U}=\left[U_{ij}\right]_{i,j\in\{1,\ldots,N\},i\neq j}$, $\theta=(\alpha,\beta)'$ and the change in notation for the utility function emphasizing that the econometrician does not observe the matrix of random utility shifters $\mathbf{U}$. In practice the elements of $\mathbf{U}$, as is common in discrete choice analysis, are assumed to be i.i.d random draws from some known distribution (e.g, the standard Normal or Logistic distribution).

Equation (103) implies that the marginal utility agent i gets from a link with j is

$$MU_{ij}\left(\mathbf{d},\mathbf{U};\theta_0\right)=\alpha_0+\beta_0\left[\sum_k d_{ik}d_{jk}\right]-U_{ij} \tag{104}$$

This marginal utility is increasing in the number of links i and j have in common, embodying a structural taste for transitive closure (here I assume that $\beta_0>0$). Clearly (104) is weakly increasing in $\mathbf{d}\in\mathbb{D}_N$ and hence Tarski's (1955) theorem applies. For a given draw of $\mathbf{U}$ and value of θ we can compute minimum and maximum equilibria, respectively $\underline{\mathbf{d}}\left(\mathbf{U};\theta\right)$ and $\overline{\mathbf{d}}\left(\mathbf{U};\theta\right)$, by fixed point iteration. Let $\underline{G}_N\left(\mathbf{U};\theta\right)$ and $\overline{G}_N\left(\mathbf{U};\theta\right)$ be the graphs corresponding to these adjacency matrices. Using these graphs we can compute, for example, the injective homomorphism frequencies $t_{\text{inj}}\left(S,\underline{G}_N\left(\mathbf{U};\theta\right)\right)$ and $t_{\text{inj}}\left(S,\overline{G}_N\left(\mathbf{U};\theta\right)\right)$ for $S=\wedge,\triangle$ etc. These homomorphism frequencies correspond to model predictions associated with specific draws of $\mathbf{U}$ and values of θ. Using simulation to integrate out the former, yields the vectors

$$\underline{\pi}\left(\theta\right)=\frac{1}{B}\sum_{b=1}^{N}\begin{pmatrix} t_{\text{inj}}\left(S_1,\underline{G}_N\left(\mathbf{U}^{(b)};\theta\right)\right)\\ \vdots \\ t_{\text{inj}}\left(S_J,\underline{G}_N\left(\mathbf{U}^{(b)};\theta\right)\right)\end{pmatrix},\quad \overline{\pi}\left(\theta\right)=\frac{1}{B}\sum_{b=1}^{N}\begin{pmatrix} t_{\text{inj}}\left(S_1,\overline{G}_N\left(\mathbf{U}^{(b)};\theta\right)\right)\\ \vdots \\ t_{\text{inj}}\left(S_J,\overline{G}_N\left(\mathbf{U}^{(b)};\theta\right)\right)\end{pmatrix}$$

for $\mathbf{U}^{(1)},\mathbf{U}^{(2)}\ldots,\mathbf{U}^{(B)}$ a sequence of independent simulated random utility shifter profiles and $S_1,\ldots,S_J$ a set of J identifying motifs of interest.

Miyauchi (2016) works under the assumption that the econometrician observes of $c=1,\ldots,C$ independent networks, with, in a slight change relative to earlier notation, G_c denoting the c^{th} network/graph. Let $\pi(G_c)$ be the vector of $S_1, \ldots, S_J$ injective homomorphism frequencies as observed in the c^{th} network and let $\underline{\pi}_c(\theta)$ and $\overline{\pi}^c(\theta)$ be the corresponding expected frequencies at the minimum and maximum pairwise stable equilibria for that network at parameter θ. These frequencies are computed using simulation as described above. Under preferences (103) the only reason these frequencies might vary with c is if the networks observed by the econometrician vary in the number of agents within them.[36]

Miyauchi (2016) focuses on assumptions which may only partially identify θ, but to begin with consider adding to the set-up the assumption that agents select the maximum equilibrium (cf., Jia, 2008). In that case

$$\mathbb{E}\left[\overline{\pi}_c(\theta_0) - \pi(G_c)\right] = 0, \tag{105}$$

is a valid moment condition. If the set of chosen motifs is sufficiently rich so as to point identify θ, then consistent estimation of θ_0 by the method of simulated moments is straightforward (McFadden, 1989; Pakes & Pollard, 1989; Gourieroux et al., 1993).

Because the asymptotic approximation involves $C \to \infty$, this approach hinges upon the availability of a large number of independent networks (each described by the same θ_0). If, instead, only a single large network is observed, then econometrician might use the methods outlined in Section 7 to form an estimate of the variance of $\pi(G_N)$, say $\hat{\Omega}_\pi$. An estimate of θ could then be formed by minimizing the simulated minimum distance (SMD) criterion:

$$\hat{\theta}_{\mathrm{SMD}} = \left(\overline{\pi}(\theta) - \pi(G_N)\right)' \hat{\Omega}_\pi^{-1} \left(\overline{\pi}(\theta) - \pi(G_N)\right).$$

Note that $\hat{\Omega}_\pi$ is constructed under an Aldous-Hoover dependence/independence structure; such a structure may not characterize the finite N structural model. An additional approximation argument is involved; understanding and formalizing this argument is required to (rigorously) derive the law of $\hat{\theta}_{\mathrm{SMD}}$ (suitably scaled and centered).

When analyzing incomplete models, researchers are often reluctant to make assumptions about equilibrium selection (which complete the model). Miyauchi (2016) shows that if the chosen vector of moments satisfies a certain monotonicity property (see his Property 1), then inference can be based upon the pair of moment inequalities

$$\begin{aligned} \mathbb{E}\left[\overline{\pi}_c(\theta_0) - \pi(G_c)\right] &\geq 0 \\ \mathbb{E}\left[\underline{\pi}_c(\theta_0) - \pi(G_c)\right] &\leq 0. \end{aligned} \tag{106}$$

[36]In more complicated models, with covariates, these minimums and maximums will vary with c due to differences in the distribution of covariates across networks.

Confidence intervals which asymptotically cover θ_0 with probability at least $1-\alpha$ can be constructed using the approach outlined by, for example, Andrews & Soares (2010). Injective homomorphism frequencies appear to satisfy the needed property, although some induced subgraph frequencies may not.

8.2 Directed links with private information

Leung (2015) studies a model of simultaneous *directed* network formation where agents have private information. In a directed network agent i may send a link to agent j such that $D_{ij}=1$; agent j may or may not decide to reciprocate and send a link back to i. The adjacency matrix is no longer symmetric, although it retains a diagonal of structural zeros. The i^{th} row of the adjacency matrix records the set of links agent i chooses to send to other agents, while the i^{th} column records the set of links that other agents choose to send to i.

To describe Leung's (2015) approach let $\mathbf{D}_{[-i,\cdot]}$ be the sub-adjacency matrix constructed by deleting the i^{th} row from $\mathbf{D}$. The marginal utility agent i receives when she directs a link to agent j is

$$\mathrm{MU}_{ij}\left(\mathbf{D}_{[-i,\cdot]},\mathbf{X};\theta_0\right)-U_{ij}. \tag{107}$$

An important implication of (107) is that while i's gain from sending a link to j may vary with the presence or absence of links elsewhere in the network, it *does not* vary with the presence or absence of other links which i herself may or may not direct. This restriction rules out interesting preference structures (see below), but simplifies the analysis substantially. Ridder & Sheng (2017) develop an approach to relaxing this feature of Leung's (2015) setup. To describe the main ideas I work with a special case of (107) :

$$\mathrm{MU}_{ij}\left(\mathbf{D}_{[-i,\cdot]},\mathbf{X};\theta_0\right)=\alpha_0+\beta_0 D_{ji}+\gamma_0\sum_{k\neq i,j}D_{ki}D_{kj}+t\left(X_i,X_j\right)'\delta_0, \tag{108}$$

for $\theta_0=(\alpha_0,\beta_0,\gamma_0,\delta_0')'$ and $t\left(X_i,X_j\right)$ a vector of possibly non-symmetric functions of exogenous agent attributes. The parameter β_0 indexes the utility gain associated with *reciprocity* in links, while γ_0 captures the utility gain arising when a link is *supported*. A directed edge from i to j is supported by agent k, if k directs links to both i and j (this allows, for example, k to "referee" transactions between i and j).

Support, although related, differs from transitivity (cf., Jackson et al., 2012); replacing $\sum_k D_{ki}D_{kj}$ with $\sum_k D_{ik}D_{jk}$ in (108) means that γ_0 would instead index a structural taste for *transitivity* (i.e., that a link to a "friend of one of my friends" generates more utility). However a transitivity term of this type is ruled out by the restriction that the marginal utility of an i to j link does not vary with the presence or absence of other links which i may or may not send (cf., Ridder & Sheng, 2017).

Leung (2015) assumes that $\mathbf{U}_i = (U_{i1}, \ldots, U_{ii-1}, U_{ii+1}, \ldots, U_{iN})'$, the idiosyncratic components of link utilities, are private information to agent i, while all other features of the game are common knowledge to all agents. Let P_{ij} denote the common prior held by all players other than i regarding the probability that she directs a link to j. Let $\mathbf{P}$ denote the $N(N-1) \times 1$ vector of such common priors. In a Bayes-Nash equilibrium agent i will best respond to the common prior by choosing to direct an edge toward j according to

$$D_{ij} = \mathbf{1}\left(\alpha_0 + \beta_0 P_{ji} + \gamma_0 \sum_{k \neq i,j} P_{ki} P_{kj} + t(X_i, X_j)' \delta_0 - U_{ij} \geq 0\right);$$

that is i forms the directed edge only if the expected marginal utility from doing so is positive. Assuming, for example, that the $\{U_{ij}\}_{i,j=1}^{N}$ are i.i.d. standard normals. Let

$$\varphi_{ij}(\mathbf{P}, \mathbf{X}; \theta_0) = \Phi\left(\alpha_0 + \beta_0 P_{ji} + \gamma_0 \sum_{k \neq i,j} P_{ki} P_{kj} + t(X_i, X_j)' \delta_0\right)$$

with $\Phi(\cdot)$ the standard normal CDF. A Bayesian-Nash equilibrium requires self-consistency of beliefs such that $\mathbf{P}$ corresponds to a fixed point of the mapping.

$$\varphi(\mathbf{P}, \mathbf{X}; \theta_0) = \begin{bmatrix} \varphi_{12}(\mathbf{P}, \mathbf{X}; \theta_0) \\ \vdots \\ \varphi_{1N}(\mathbf{P}, \mathbf{X}; \theta) \\ \vdots \\ \varphi_{N1}(\mathbf{P}, \mathbf{X}; \theta) \\ \vdots \\ \varphi_{NN-1}(\mathbf{P}, \mathbf{X}; \theta_0) \end{bmatrix}. \tag{109}$$

One approach would be to apply ideas analogous to those developed in Miyauchi (2016) using (109). Leung (2015), instead creatively adapts the two-step approach familiar from the wider econometrics literature on incomplete information games (Bajari et al., 2010a, 2013). Let $\hat{\mathbf{P}}$ be a nonparametric estimate of the belief vector $\mathbf{P}$. With this estimate in hand, θ_0, may be estimated by finding the maximum of the criterion

$$\max_{\theta} \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} D_{ij} \ln \varphi_{ij}\left(\hat{\mathbf{P}}, \mathbf{X}; \theta\right) + (1 - D_{ij}) \ln\left[1 - \varphi_{ij}\left(\hat{\mathbf{P}}, \mathbf{X}; \theta\right)\right], \tag{110}$$

using a standard Probit MLE program.

The challenge with this approach is that it is not obvious how one can consistently estimate $\mathbf{P}$. Unlike in, for example, the literature on entry games, where the same player is observed playing independent replications of a game across different markets, in the present set-up there is only a single game. Leung's (2015) key insight is to note that under an exchangeability assumption and a focus on symmetric equilibria, estimation of $\mathbf{P}$ is possible because it implies that (ordered dyads) with identical covariate configurations have identical *ex ante* linking probabilities. For the case of discretely-valued X_i Leung (2015, Proposition 1) shows, under assumptions, that for $\hat{P}_{ij} = \left[\sum_{k,l} \mathbf{1}\left(t\left(X_k, X_l\right) = t\left(X_i, X_j\right)\right)\right]^{-1} \times \left[\sum_{k,l} D_{kl}\mathbf{1}\left(t\left(X_k, X_l\right) = t\left(X_i, X_j\right)\right)\right]$

$$\sup_{i,j \in \mathcal{V}(G_N)} \left|\hat{P}_{ij} - P_{ij}\right| \xrightarrow{p} 0$$

at rate $N^{1/2}$. Using these estimates in (110) results in a consistent and asymptotically normal estimate of θ_0 under regularity conditions. The interesting features of these results involve the need to account for first step estimation error as well as for dependencies across dyads sharing one agent in common. Leung (2015) also presents a variance estimator and an empirical illustration based on the network data collected by Banerjee et al. (2013).

8.3 Bounded degree and restricted heterogeneity

Like Miyauchi (2016), de Paula et al. (2018) study a simultaneous-move complete information model of network formation. They place three key restrictions on the graph generating process. First, they assume that agents only wish to maintain a small number of links. Second, that utility only varies with the addition or deletion of links within a finite radius. For example an agent may care about the friends of her friends, but not the friends of the friends of her friends. Third, there are only a finite number of agent types and, crucially, agents are indifferent among links of the same type. There is some nuance to the last restriction since indirect connections may matter. Consider two Black individuals, each with a Black and White friend, the restriction is that any third agent is indifferent between forming a link with either of these two individuals. Similar restrictions feature prominently in one-to-one transferable utility matching models (e.g., Choo & Siow, 2006; Graham, 2013; Galichon & Salanie, 2017). The first and last of these assumptions are "non-standard", but de Paula et al. (2018) show how they make identification analysis tractable.

They begin by noting that, under their assumptions, any *rooted network* – a configuration of links within a fixed distance about a focal "root" node – will take one of a finite number of configurations. Identification of preference parameters comes from comparing model predictions about the frequency of these configurations with their empirical counterparts.

The operationalization of this intuition into a workable method of inference is the main contribution of their paper. To describe this contribution assume that the utility agent i gets from network configuration $\mathbf{D} = \mathbf{d}$ is, for example,

$$\nu_i\left(\mathbf{d}, \mathbf{U}; \theta_0\right) = \sum_j d_{ij} \left[\alpha_0' R_{ij} + \beta_0 \left[\sum_k d_{ik} d_{jk}\right] + U_i\left(X_j\right)\right] - \infty \cdot \mathbf{1}\left(\sum_j d_{ij} > L\right).$$

Here $R_{ij} = r\left(X_i, X_j\right)$ is a vector of known symmetric functions of X_i and X_j, L is the maximum number links an agent might desire (known by the econometrician), and $U_i\left(x\right)$ is an unobserved utility-shifter with known distribution. This utility shifter varies with i, but only depends on j via the covariate X_j. The expression above suggests that associated with each agent are just $|\mathbb{X}|$ shocks, one for each type of agent. de Paula et al. (2018) actually attach $L \times |\mathbb{X}|$ shocks to each agent, but their main ideas can be conveyed under the more restrictive set-up.

Let $U_i = \left(U_i\left(x_1\right), \ldots, U_i\left(x_{|\mathbb{X}|}\right)\right)'$ denote the vector of taste shocks associated with agent i. Since agents maintain no more than L links, and preferences are only affected by network structure within a certain radius, the number of logically observable rooted network configurations is finite. For each of these configurations we can ask, for a given value of θ and draw of U_i, whether an agent will unilaterally reject it (e.g., given the configuration's structure she may prefer to unilaterally dissolve some links). de Paula et al. (2018) call the set of acceptable rooted networks a *preference class*. Since the distribution of U_i is known, the ex ante probability that any individual falls into a particular preference class when θ takes a particular value can be computed (typically via simulation).

A network can be generated by choosing the frequency with which agents of (i) a particular type, and (ii) belonging to a particular preference class, are assigned to specific rooted network configurations. Theorem 1 of de Paula et al. (2018) shows how to construct these frequencies in a way which satisfies pairwise stability. A parameter value belongs to the identified set, if there exists a feasible vector of allocation probabilities such that the predicted frequencies with which the various rooted network configurations occur match their corresponding empirical frequencies. Theorem 2 of de Paula et al. (2018) shows how this question may be answered by solving a particular quadratic program.

The identification analysis assumes there are continuum of agents. Since their graph is sparse, the object they work with is not a graphon, but its sparse graph analog, called a *graphing* in the literature (Lovász, 2012). For inference they assume that the econometrician observes a random sample of rooted networks, perhaps collected via snowball sampling.

8.4 Many agent approximations

Menzel (2016) studies a class of large network formation models with exchangeable agents. He characterizes the limiting network as $N \to \infty$ and investigates how to use this limit to approximate the finite network in hand. One example of the family of preference structures accommodated by his set-up is

$$MU_{ij}\left(\mathbf{D},\mathbf{U};\theta_0\right) = \underbrace{\alpha_0' R_{ij} + \beta_0 \left[\min\left(\sum_k D_{ik}D_{jk}, 1\right)\right] + \sigma U_{ij}}_{\text{marginal benefit}} - \underbrace{\left(\ln J + \sigma U_{i0}\right)}_{\text{marginal cost}}. \tag{111}$$

The scale and location parameters, σ and J, vary with N in a particular way in order to obtain useful limits.

Menzel (2016) observes that a network is pairwise stable if and only if $D_{ij} = 1$ when $MU_{ij} \geq 0$ and $D_{ij} = 0$ when $MU_{ij} < 0$ for all $j \in \mathcal{W}_i$. The set $\mathcal{W}_i$ includes all agents j who are willing to form a link with agent i or, equivalently, who would not veto such a link:

$$\mathcal{W}_i = \left\{j \in \mathcal{V}\left(G_N\right) \setminus \{i\} \, : \, MU_{ji} \geq 0\right\}.$$

When J grows with N at the appropriate rate, the number of links accepted by agent i, among those available in $\mathcal{W}_i$, is stochastically bounded (ensuring that the limiting network is sparse). Furthermore, using extreme value theory, Menzel (2016) shows that the effect of the endogenous choice set, $\mathcal{W}_i$, on the probability of forming a particular link is completely summarized by a conditional logit type inclusive value.

Let

$$Z_{ij} = \left(X_i, X_j, T_{ij}\right)'$$

for $T_{ij} = \min\left(\sum_k D_{ik}D_{jk}, 1\right)$. Further, for purposes of illustration, let $X_i \in \{0,1\}$ be a binary indicator for, say, gender. In this case Z_{ij} takes values within the finite set $\mathbb{Z}$. For example woman i may link with woman j, with whom she shares at least one friend in common, such that $Z_{ij} = (1,1,1)'$. Menzel (2016) demonstrates that the probability that agent i's highest utility link is of type $Z_{ij} = z$, among all those available to her, takes a logit form. Furthermore, when N is large, the inclusive value in this probability depends only on agent i's exogenous attribute X_i (for preferences structure different than (111) the argument may be a bit more complicated).

In setting up the sequence of network formation games appropriately, and also in characterizing the resulting limit, Menzel (2016) demonstrates considerable ingenuity and technical skill. Stepping back, the underlying intuition is quite simple. Under exchangeability of agents, the link formation process for observationally identically agents should be similar when N is large.

Next consider the link frequency "distribution"

$$F_N(z) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{1}\left(D_{ij}=1, X_i \leq x_1, X_j \leq x_2, T_{ij} \leq t_{12}\right)$$

for $z = (x_1, x_2, t_{12})'$. This is not a proper measure, it integrates to average degree, not one. Nevertheless these frequencies have well-defined limits which Menzel (2016) is able to relate to the limiting choice probabilities associated with the infinite agent network formation game. Note that $F_N(z)$ is closely related to a network moment, as introduced in (3.5) earlier. As in the other papers surveyed in this section, the identification/estimation approach relates empirical frequencies with model-implied counterparts. Characterizing these model-implied counterparts (in the limit) is non-trivial. Menzel (2016) shows that depending on the preference structure considered, as well as researcher assumptions about equilibrium selection, preference parameters may be point or set identified. For point identified models Menzel (2016) suggests a constrained maximum likelihood estimator based on the form of the limiting model.

8.5 Models with (unobserved) sequential meeting processes

Miyauchi (2016), Leung (2015), de Paula et al. (2018) and Menzel (2016) all model network formation as a static game. Any underlying dynamics governing link formation are left unmodeled. This is in keeping with the agnosticism regarding equilibrium selection maintained by these researchers. Mele (2017) and Mele & Zhu (2017), in contrast, present models of network formation which make explicit assumptions about how agents meet, form, dissolve and maintain links.

In their model pairs of agents meet sequentially. Upon meeting a dyad decides to either form, maintain, or dissolve a link. Although the utility attached to any given link may depend on current network structure, agents are not forward looking. Rather agents myopically add, maintain, or subtract links in order to raise current utility without anticipating the effects of their actions on the future decisions of other agents in the network.

To discuss their results I work with the preference specification featured in Mele & Zhu

(2017). Let $\mathbf{d}_t$ be a particular undirected network configuration in period t. The utility agent i gets from such a configuration is given by

$$\nu_i\left(\mathbf{d}_t, \mathbf{X}, \mathbf{U}_t; \theta_0\right) = \sum_j d_{ijt}\left[\alpha_0' R_{ij}^* + \frac{\beta_0}{N}\left[\sum_k d_{jkt}\right] - U_{ijt}\right] \tag{112}$$

with $\theta = (\alpha', \beta)'$. Here $R_{ij}^* = r^*\left(X_i, X_j\right)$ is a vector of known functions of X_i and X_j. This term indexes, for example, the utility gains from homophilous sorting. The second term in (112) captures the benefits associated with indirect connections; that is, the return agent i receives from linking with j may, in part, depend on the number of links j already has. If $\beta_0 > 0$ ($\beta_0 < 0$), then there exist utility gains from linking with more (less) popular agents. It is also possible to incorporate a transitivity, or mutual friends, term into (112).

The preference shock U_{ijt} is a Type I extreme value random variable; independently distributed across dyads and over time. I will return to the implications of these assumptions for the interpretation and identification of the model shortly.

Under (112) the marginal utility agent i gets from a link with j is

$$MU_{ij}\left(\mathbf{d}_t, \mathbf{X}, \mathbf{U}_t; \theta_0\right) = \alpha_0' R_{ij}^* + \frac{\beta_0}{N}\left[\sum_k d_{jkt}\right] - U_{ijt}. \tag{113}$$

Mele & Zhu (2017) assume utility is transferable. This implies that if i and j meet in period t they will form (or maintain) a link if the net surplus from doing so is positive (cf., Bloch & Jackson, 2007):

$$\begin{aligned} MU_{ij}\left(\mathbf{d}_t, \mathbf{X}, \mathbf{U}_t; \theta_0\right) + MU_{ji}\left(\mathbf{d}_t, \mathbf{X}, \mathbf{U}_t; \theta_0\right) \geq 0 \Longleftrightarrow & R_{ij}'\alpha_0 + \frac{\beta_0}{N}\left[\sum_k \left(d_{ikt} + d_{jkt}\right)\right] \\ & \geq U_{ijt} + U_{jit}, \end{aligned} \tag{114}$$

where $R_{ij} = R_{ij}^* + R_{ji}^*$. The R_{ij} term is analogous to the vector of regressors appearing in the dyadic regression model discussed in Section 4. Observe, in keeping with the undirected nature of the network, that (114) is invariant to permutations of the agents' indices.

Dyads meet one at a time (i.e., sequentially). In each period the probability that a particular ij dyad is chosen, say ρ_{ij}, is greater than zero. Let $Z_t = ij$ if dyad $\{i, j\}$ is chosen to meet in period t. This meeting variable equals one of the $\binom{N}{2}$ possible dyad index pairs each period. Conditional on i and j meeting, as well as the beginning-of-period-t network structure, the probability that they form (or maintain) a link is logistic:

$$\Pr\left(\mathbf{D}_{t+1}=\mathbf{D}_t+ij\middle|\,\mathbf{D}_t,\mathbf{X},Z_t=ij;\theta_0\right)=\frac{\exp\left(R_{ij}'\alpha_0+\frac{\beta_0}{N}\left[\sum_k\left(D_{ikt}+D_{jkt}\right)\right]\right)}{1+\exp\left(R_{ij}'\alpha_0+\frac{\beta_0}{N}\left[\sum_k\left(D_{ikt}+D_{jkt}\right)\right]\right)}.$$

This link probability function augments the simple dyadic logistic regression model introduced earlier with terms, in this case a popularity effect, which arise due to interdependencies in preferences.

Under these assumptions the sequence of adjacency matrices $\mathbf{D}_0, \mathbf{D}_1, \ldots$ is a Markov chain with transition probabilities depending on the exact specification of the meeting process and the logistic probabilities specified above. This chain is irreducible and aperiodic. Therefore the ergodic theorem implies that, in the limit, realized networks will correspond to draws from a unique stationary distribution. Mele & Zhu (2017, Theorem 2.1) show that this stationary distribution equals (cf., Blume, 1993)

$$\pi_N\left(\mathbf{d};\mathbf{X},\theta_0\right)=\frac{\exp\left(Q_N\left(\mathbf{d};\mathbf{X},\theta_0\right)\right)}{\sum_{\mathbf{v}\in\mathbb{D}_N}\exp\left(Q_N\left(\mathbf{v};\mathbf{X},\theta_0\right)\right)} \tag{115}$$

for $Q_N\left(\mathbf{d};\mathbf{X},\theta\right)=\sum_{i=1}^{N}\nu_i\left(\mathbf{d},\mathbf{X},\mathbf{0};\theta_0\right)$. See also Mele (2017, Theorem 1). Equation (115) corresponds to what network researchers call an *exponential random graph model* (ERGM). Robins et al. (2007a) and Robins et al. (2007b) provide an overview of ERGMs for social network analysis. de Paula (2017, Section 4.1) provides an interesting overview from the vantage point of an econometrician. The results of Mele (2017) and Mele & Zhu (2017) provide a microeconomic foundation for (certain forms of) ERGMs. This is interesting, especially in light of peculiarities of the ERGM modeling framework emphasized by others (e.g., Shalizi & Rinaldo, 2013).

It turns out that $Q_N\left(\mathbf{d};\mathbf{X},\theta\right)$ is also the potential function, in the sense of Monderer & Shapley (1996), associated with a particular network formation game. Consider preference structure (112), but with all the pair-specific preference shocks set identically equal to zero. The set of networks which (locally) maximize $Q_N\left(\mathbf{d};\mathbf{X},\theta\right)$ correspond to the set of Nash equilibrium networks associated with the simultaneous move network formation game under these zero heterogeneity preferences. The stationary distribution (115) clearly has modes at these equilibria. Mele & Zhu (2017) assume that the econometrician observes a single draw from this stationary distribution. This draw, loosely, can be viewed as a random perturbation of an equilibrium network in the associated "heterogeneity free" simultaneous move static game.

Unfortunately computing the maximum likelihood estimate of θ_0 is not straightforward.

This is because the denominator in (115) involves a summation over all undirected networks of order N. It is impossible to evaluate this summation directly except for trivially small networks. Furthermore approximate computation of the MLE via, for example, Markov Chain Monte Carlo (MCMC) methods, is also difficult (e.g., Bhamidi et al., 2011).

Mele & Zhu (2017), building on ideas in Chatterjee & Diaconis (2013) and Chatterjee & Dembo (2016), propose an approximate variational estimate of θ_0 (cf., Daudin et al., 2008; Bickel et al., 2013). While their approximation does not generally coincide with the MLE, they show that the difference between the two shrinks to zero as N grows large.

At a high level they proceed as follows. First, consider the conditional edge independence model introduced in (3):

$$\Pr\left(\mathbf{D}=\mathbf{d};q\right)=\prod_{i<j}q_{ij}^{d_{ij}}\left(1-q_{ij}\right)^{1-d_{ij}}, \tag{116}$$

with q_{ij} equaling the probability that i and j link. In this context the conditional edge independence model is sometimes called the mean-field approximation. Exploiting ideas in Wainwright & Jordan (2008) and He & Zheng (2013), they observe that the (log of the) constant of integration in (115) is bounded below by

$$\frac{1}{N^2}\ln\left[\sum_{\mathbf{v}\in\mathbb{D}_N}\exp\left(Q_N\left(\mathbf{v};\theta\right)\right)\right]\geq\mathbb{E}_q\left[Q_N\left(\mathbf{D};\theta\right)\right]+\frac{1}{N^2}\mathbb{S}\left(q\right)$$

with $\mathbb{S}\left(q\right)$ denoting Shannon's Entropy and the expectation with respect to the approximating mean field model (116). Next choose the probabilities $q=\left(q_{12},q_{13},\ldots,q_{N-1N}\right)'$ to maximize the above lower bound. This is the variational problem. Clearly, the optimal approximation will vary with θ, the structural parameter of interest. The approximation will also not be exact, since conditional edge independence models represent only a restricted set of all the possible probability distributions on $\mathbb{D}_N$. The variational estimate of θ_0, say $\hat{\theta}_{\mathrm{VE}}$, is chosen to maximize (115) after replacing its denominator with the lower bound described above.

Mele & Zhu (2017), using a result in Chatterjee & Dembo (2016), show that the lower bound approximation becomes tight as $N\rightarrow\infty$. Furthermore the limit of the variational problem corresponds to finding a graphon. More precisely, they find that as $N\rightarrow\infty$,

1. the stationary distribution associated with their strategic network formation model is arbitrarily well approximated by a conditional edge independence model with some graphon $h\left(u,v\right)$, or a mixture of such models;

2. these graphons correspond to local maximizers of a limiting version of the variational problem; and
3. $\hat{\theta}_{\mathrm{VE}}$ coincides with a local maximizer of (115).

The first finding is to be expected given the Aldous-Hoover Theorem and associated discussion in Section 3. The second result is related to work by Chatterjee & Diaconis (2013). It is of interest here since it provides a connection between a structural model of strategic network formation and the exchangeable random graph theory reviewed earlier.

While Mele (2017) provides a nice microeconomic potential game interpretation of ERGMs, and Mele & Zhu (2017) make important progress on methods of estimation, major challenges in the areas of identification, estimation and inference in this class of models nevertheless remain.

Christakis et al. (2010) also model link formation as a sequential process. Their approach differs from that of Mele (2017). They assume the initial network is empty and that all $\binom{N}{2}$ dyads meet in a specific (unobserved) order. Upon meeting they myopically decide whether to form a link or not. After all pairs of agents meet once, further link revisions do not occur. In order to construct a likelihood Christakis et al. (2010) assigned a distribution to the unobserved meeting sequence and integrate it out. For computation they develop a Bayesian approach based on MCMC methods. One feature of their set-up is that the model may place positive probability on network configurations that are not pairwise stable. In contrast the ergodic distribution associated with Mele's (2017) model places most of its mass in the neighborhood of equilibrium network configurations. While this may be viewed as undesirable, from a computational standpoint the Christakis et al. (2010) method appears attractive. In principal their model could be extended to allow each pair of agents to meet multiple (but still a finite number of) times.

8.6 Further reading and open questions

With exception of the paper by Miyauchi (2016), all of the papers surveyed above base estimation and inference on a single network. To get workable LLNs and CLTs each of these authors deal with the dependence across links induced by strategic interaction in interesting ways. Leung (2015) introduces private information; in a resulting Bayes-Nash equilibrium links are conditionally independent given common information. The reduced form probability of a directed link from i to j implied by his model is quite similar to the representation result associated with X-exchangeability introduced in the context of dyadic regression in Section 4. To a first approximation this probability depends only upon X_i and

X_j (since the other sources of variation in $\sum_{k \neq i,j} P_{ki} P_{kj}$ should be rather modest when N is large enough). Therefore, relative to a simple dyadic probit model, the Leung (2015) model adds an equilibrium constraint.

In de Paula et al. (2018) the key assumption appears to be that preference heterogeneity is over *types* of links alone, with no dyad-specific component. As mentioned earlier, similar assumptions have proved to be very powerful in the literature on matching. Although Menzel (2016) works with a model which generates a sparse graph in the limit (with dependence across links vanishing), his use of exchangeability arguments does suggest connections to the Aldous-Hoover type representation results introduced earlier. Mele (2017) and Christakis et al. (2010) posit sequential meeting processes that effectively "complete" what would otherwise be an incomplete simultaneous move N-player game. Each of these approaches have pros and cons; a variety of computational and inference issues remain unsolved. At the same time the creativity and diversity of them suggests that forward progress on these types of models is possible. Better understanding the connections between different modeling assumptions would be useful.

Another approach, not surveyed here, but nevertheless promising, involves working with subnetworks. A focus on subnetworks sidesteps some of the computational challenges that arise when trying to apply methods from the econometrics of games to network formation problems (where there are typically many agents). Sheng (2014) pioneered this approach. Gualdani (ming) develops additional (related) results.

9 The bright and happy future of network econometrics

This chapter has surveyed a burgeoning literature on the econometrics of networks. This literature – combining insights from econometric research on panel data and games, new tools in applied probability and statistics, and original thinking – now provides a basic set of tools for the analysis of networks. Nevertheless substantial work remains unfinished. As noted at the start of this chapter, datasets with natural graph theoretic structure abound in economics, and increasingly feature in published research. Each dataset exhibits its own peculiarities: in some links are undirected, in others directed. The network may be bipartite or even multi-partite(e.g., Min, 2019). The size and order of available network datasets vary immensely. In some cases a network may be observed over multiple periods, in others just once. For many of these settings there exist no extant econometric modeling strategies, in all of them existing work could be improved in a number of ways.

A defining feature of the econometric approach to modeling network formation is its random utility foundation. When preferences are interdependent – where the utility two agents attach to a candidate link may vary with the presence or absence of links elsewhere in the network – multiple equilibrium network configurations are likely. The analysis of incomplete models is an important recent accomplishment of econometrics. The combinatoric complexity of large networks will require new developments in this area. The set of papers surveyed in Section 8 gives some flavor of the key issues and possible solutions.

Another defining feature of modern microeconometric research is the incorporation of unobserved heterogeneity; heterogeneity that agents observe and act upon, but which is unobserved by the researcher. In the single agent setting panel data facilitates the identification and estimation of models with rich heterogeneity structures. Networks have natural panel-like aspects. In a dense network each agent decides whether to (attempt to) form a link with all other agents. Multiple decisions per agent are observed. Leveraging this panel-like structure has been a key feature of some of the contributions surveyed in Sections 4, 5 and 6 above.

Understanding the properties of the different methods surveyed above under sequences of networks which are dense, sparse or somewhere in between, remains incomplete. Uniformity of testing procedures across these various cases would be desirable. Some preliminary work on bootstrapping methods in the networks setting now exists (e.g., Green & Shalizi, 2017; Menzel, 2017; Davezies et al., 2019), but this remains relatively unexplored. Semiparametric efficiency bounds are yet to be characterized, let alone the development of estimators attaining them. Computational advances will be important for spurring real world application.

This chapter has focused on network formation. While the question of how networks form is scientifically interesting, so is that of what they do? This latter question was a key driver of the peer effects literature which emerged after Manski (1993). Developing methods for simultaneously modeling the formation and consequences of social and economic networks remains an important open area (Auerbach, 2016; Johnsson & Moon, 2017). Finally, although more and more empirical work with a network dimensions appears each year, application of the methods outlined above in substantive empirical work is a high priority. In addition to whatever subject area insights such applications may produce, they will no doubt spur further methodological innovations.

A Appendix

Lemma 2. (U-Statistic with Estimated Parameter) *Let* $\{Z_i\}_{i=1}^N$ *be a simple random sample drawn from some population* F_Z *and* $\phi(Z_i, Z_j; \beta, \gamma)$ *be a function from* $\mathbb{Z} \times \mathbb{Z}$ *to* $\mathbb{R}^J$ *indexed by* $\beta \in \mathbb{B}$ *and* $\gamma \in \mathbb{C}$ *(with* $\mathbb{B}$ *and* $\mathbb{C}$ *compact subsets of* $\mathbb{R}^{\dim(\beta)}$ *and* $\mathbb{R}^{\dim(\gamma)}$ *respectively). Suppose that* $\phi(z_1, z_2; \beta, \gamma)$ *is twice continuously differentiable in* γ *for all* $z_1, z_2 \in \mathbb{Z} \times \mathbb{Z}$ *with*

$$\mathbb{E}\left[\|\phi(Z_1, Z_2; \beta, \gamma)\|_2\right] < \infty \tag{117}$$

$$\mathbb{E}\left[\left\|\frac{\partial \phi(Z_1, Z_2; \beta, \gamma)}{\partial \gamma'}\right\|_F\right] < \infty \tag{118}$$

$$\mathbb{E}\left[\left\|\frac{\partial}{\partial \gamma'}\left\{\frac{\partial \phi(Z_1, Z_2; \beta, \gamma)}{\partial \gamma_p}\right\}\right\|_F\right] < \infty, \quad p = 1, \ldots, \dim(\gamma). \tag{119}$$

Then, for $\hat{\gamma}$ *a* $\sqrt{N}$*-consistent estimate of* γ_0*, and defining* $\bar{\phi}_N(\beta, \gamma) \overset{def}{\equiv} \binom{N}{2}^{-1} \sum_{i=1}^N \sum_{j=i+1}^{N-1} \phi(Z_i, Z_j; \beta, \gamma)$ *and* $\Phi(\beta, \gamma) \overset{def}{\equiv} \mathbb{E}[\phi(Z_1, Z_2; \beta, \gamma)]$*, we have*

$$\sqrt{N}\left[\bar{\phi}_N(\beta, \hat{\gamma}) - \Phi(\beta, \gamma_0)\right] = \frac{2}{\sqrt{N}} \sum_{i=1}^N \psi_0(Z_1; \beta, \gamma_0) + \Gamma_{0,\beta\gamma}(\beta) \sqrt{N}(\hat{\gamma} - \gamma_0) + o_p(1) \tag{120}$$

where $\phi_1(z; \beta, \gamma) = \mathbb{E}[\phi(z, Z_1; \beta, \gamma)]$ *and*

$$\psi_0(Z_1; \beta, \gamma) = \phi_1(Z_1; \beta, \gamma) - \Phi(\beta, \gamma)$$
$$\Gamma_{0,\beta\gamma}(\beta) = \mathbb{E}\left[\frac{\partial \phi(Z_1, Z_2; \beta, \gamma_0)}{\partial \gamma'}\right].$$

Proof of Lemma 2

A Taylor expansion of $\bar{\phi}_N(\beta, \hat{\gamma})$ in $\hat{\gamma}$ about γ_0 yields, after some re-arrangement and centering,

$$\sqrt{N}\left[\bar{\phi}_N(\beta, \hat{\gamma}) - \Phi(\beta, \gamma_0)\right] = \sqrt{N}\left[\bar{\phi}_N(\beta, \gamma_0) - \Phi(\beta, \gamma_0)\right] + \Gamma_{N,\beta\gamma}(\beta, \bar{\gamma}) \sqrt{N}(\hat{\gamma} - \gamma_0), \tag{121}$$

with $\bar{\gamma}$ a mean value between $\hat{\gamma}$ and γ_0 which may vary across the rows of the Hessian $\Gamma_{N,\beta\gamma}(\beta, \gamma) \overset{def}{\equiv} \frac{\partial \bar{\phi}_N(\beta, \gamma)}{\partial \gamma'}$. Next recall the definition of the $L_{2,1}$ norm:

$$\|\mathbf{A}\|_{2,1} = \sum_{j=1}^n \left[\sum_{i=1}^m |a_{ij}|^2\right]^{1/2}. \tag{122}$$

The mean value theorem, as well as compatibility of the Frobenius matrix norm with the Euclidean vector norm, gives for any γ and γ^* both in $\mathbb{C}$,

$$\left\| \frac{\partial \bar{\phi}_N(\beta,\gamma)}{\partial \gamma'} - \frac{\partial \bar{\phi}_N(\beta,\gamma^*)}{\partial \gamma'} \right\|_{2,1} \leq \sum_{p=1}^{\dim(\gamma)} \left\| \frac{\partial}{\partial \gamma'} \left\{ \frac{\partial \bar{\phi}_N(\beta,\gamma)}{\partial \gamma_p} \right\} \right\|_F \|\gamma - \gamma_*\|_2 . \tag{123}$$

Observe that $\frac{\partial}{\partial \gamma'} \left\{ \frac{\partial \bar{\phi}_N(\beta,\gamma)}{\partial \gamma_p} \right\}$ is a matrix of U-statistics with kernels whose first moments are finite (by condition 106 above). By Serfling (1980, Theorem 5.4A) these U-statistics converge in probability and hence, from (123)

$$\left\| \frac{\partial \bar{\phi}_N(\beta,\gamma)}{\partial \gamma'} - \frac{\partial \bar{\phi}_N(\beta,\gamma^*)}{\partial \gamma'} \right\|_{2,1} \leq O_p(1) \cdot \|\gamma - \gamma_*\|_2 .$$

This condition, as well compactness of $\mathbb{C}$, continuity of $\frac{\partial \bar{\phi}_N(\beta,\gamma)}{\partial \gamma}$ in γ, and condition (118), allow for an application of Lemma 2.9 in Newey & McFadden (1994) such that $\sup_{\gamma \in \mathbb{C}} \left\| \frac{\partial \bar{\phi}_N(\beta,\gamma)}{\partial \gamma'} - \Gamma_{\beta\gamma}(\beta,\gamma) \right\|_F \xrightarrow{p} 0$ with $\Gamma_{\beta\gamma}(\beta,\gamma) = \mathbb{E}\left[\frac{\partial \phi(Z_1,Z_2;\beta,\gamma)}{\partial \gamma'} \right]$. This, along with consistency of $\hat{\gamma}$ for γ_0, is enough to ensure that $\frac{\partial \bar{\phi}_N(\beta,\bar{\gamma})}{\partial \gamma'} \xrightarrow{p} \Gamma_{0,\beta\gamma}(\beta)$. Equation (120) then follows by observing that $\bar{\phi}_N(\beta,\gamma_0) - \Phi(\beta,\gamma_0)$ is a vector of mean zero U-Statistics with Hájek projections equal to the corresponding components of the first term to the right of the equality in (120) (see, for example, Theorem 5.3.3. of Serfling (1980) and invoke condition (117) above). See Mao (2018, Lemma S1) for a related Lemma.

Order of variances and covariances for p^{th} order induced subgraph frequencies

Here I present the order of the covariance between empirical subgraph frequencies, where the subgraph is of arbitrary order. For general p^{th}-order graphlets R and S we have that

$$\begin{aligned} \mathbb{C}(P_N(R), P_N(S)) =& \binom{N}{p}^{-2} \sum_{q=1}^{p} \binom{N}{p} \binom{p}{q} \binom{N-p}{p-q} \Sigma_q(R,S) \\ =& \binom{N}{p}^{-2} \sum_{q=1}^{p} \binom{N}{p} \binom{p}{q} \binom{N-p}{p-q} \Xi(\mathcal{W}_{q,R,S}) \\ & - \left[1 - \frac{(N-p)!^2}{N!(N-2p)!} \right] P(R) P(S) . \end{aligned} \tag{124}$$

Normalizing by ρ_N raised to the number of edges in R and S, respectively $\rho_N^{e(R)}$ and $\rho_N^{e(S)}$, yields

$$\mathbb{C}\left(\tilde{P}_N(R), \tilde{P}_N(S)\right) = \left\{ \underbrace{\binom{N}{p}^{-2} \sum_{q=1}^{p-1} \binom{N}{p} \binom{p}{q} \binom{N-p}{p-q} \left[\frac{\Xi\left(\mathcal{W}_{q,R,S}\right)}{\rho_N^{e(R)} \rho_N^{e(S)}} \right]}_{O\left(N^{-q} \rho_N^{-e(R)} \rho_N^{-e(S)}\right) O\left(\Xi\left(\mathcal{W}_{q,R,S}\right)\right)} \right. \left. - \left[1 - \frac{(N-p)!^2}{N!\,(N-2p)!}\right] \tilde{P}(R)\, \tilde{P}(S) \right\}. \tag{125}$$

There are $2p-q$ vertices in each element of $\mathcal{W}_{q,R,S}$.

Case 1 ($q=1$):

If $q=1$, then $e(W) = e(R) + e(S)$ for all $W \in \mathcal{W}_{q,R,S}$. This gives

$$\begin{aligned} O\left(N^{-q} \rho_N^{-e(R)} \rho_N^{-e(S)}\right) O\left(\Xi\left(\mathcal{W}_{1,R,S}\right)\right) &= O\left(N^{-1} \rho_N^{-e(R)} \rho_N^{-e(S)}\right) O\left(\rho_N^{e(R)} \rho_N^{e(S)}\right) \\ &= O\left(N^{-1}\right). \end{aligned}$$

Case ($q=p$):

If $q=p$, then $\Xi\left(\mathcal{W}_{q,R,S}\right) = 0$ unless $R=S$. In that case, the "variance case", we have that $e(W) = p$ since $W = R = S$. This gives

$$\begin{aligned} O\left(N^{-q} \rho_N^{-2e(R)}\right) O\left(\Xi\left(\mathcal{W}_{p,R}\right)\right) &= O\left(N^{-p} \rho_N^{-2e(R)}\right) O\left(\rho_N^{e(R)}\right) \\ &= O\left(N^{-p} \rho_N^{-e(R)}\right). \end{aligned}$$

If R is a p-cycle, then $p = e(R)$, yielding the simplification $O\left(N^{-p} \rho_N^{-e(R)}\right) = O\left(\lambda_N^{-p}\right)$.

If R is a tree, then $e(R) = p-1$, yielding the simplification $O\left(N^{-p} \rho_N^{-e(R)}\right) = O\left(N^{-1} \lambda_N^{-(p-1)}\right)$.

Case $(1 < q < p)$:

For $q = 2, \ldots, p-1$ we have that $e(W) = e(R) + e(S) - (q-1)$ if R and S are both p-cycles so that

$$
\begin{aligned}
O\left(N^{-q}\rho_N^{-e(R)}\rho_N^{-e(S)}\right) O\left(Q\left(\mathcal{W}_{q,R,S}\right)\right) &= O\left(N^{-q}\rho_N^{-e(R)}\rho_N^{-e(S)}\right) O\left(\rho_N^{e(R)+e(S)-(q-1)}\right) \\
&= O\left(N^{-q}\rho_N^{-(q-1)}\right) \\
&= O\left(N^{-1}\lambda_N^{-(q-1)}\right).
\end{aligned}
$$

Whereas we have that $e(W) \geq e(R) + e(S) - (q-1)$ if R and S are both trees, or one is a tree and the other a p-cycle, so that

$$
\begin{aligned}
O\left(N^{-q}\rho_N^{-e(R)}\rho_N^{-e(S)}\right) O\left(\Xi\left(\mathcal{W}_{q,R,S}\right)\right) &\leq O\left(N^{-q}\rho_N^{-e(R)}\rho_N^{-e(S)}\right) O\left(\rho_N^{e(R)+e(S)-(q-1)}\right) \\
&= O\left(N^{-1}\lambda_N^{-(q-1)}\right).
\end{aligned}
$$

Proof of Theorem 3

Without loss of generality set $i = 1$. By the definition of degree we have that

$$
\mathbb{E}\left[D_{1+}^m\right] = \mathbb{E}\left[\left(\sum_{j=2}^{N} D_{1j}\right)^m\right],
$$

the multinomial theorem allows us to write the term inside the expectation above as

$$
D_{1+}^m = \left(\sum_{j=2}^{N} D_{1j}\right)^m = \sum_{q_2+\cdots+q_N=m} \binom{m}{q_2, q_3, \cdots, q_N} \prod_{j=2}^{N} D_{1j}^{q_j} \tag{126}
$$

where $\binom{m}{q_2,q_3,\cdots,q_N} = \frac{m!}{q_2!q_3!\cdots q_N!}$. Since D_{1j} is binary $D_{1j}^{q_j} = D_{1j}$ for all $q_j = 1, 2, \ldots, m$ and zero when $q_j = 0$. This implies that $\prod_{j=2}^{N} D_{1j}^{q_j} = D_{1j_1} \times \cdots \times D_{1j_k}$ for $D_{1j_1}, D_{1j_2}, \ldots, D_{1j_k}$ the set of $1 \leq k \leq m$ link indicators with $q_j \geq 1$. Consider agents $j_1, j_2, \ldots, j_k$, with, say, $q_{j_1} = p_1$, $q_{j_2} = p_2, \ldots, q_{j_k} = p_k$ such that $\mathbf{p} \in \mathcal{P}_{k,m}$, it follows that

$$
\prod_{j=2}^{N} D_{1j}^{q_j} = D_{1j_1}^{p_1} \times \cdots \times D_{1j_k}^{p_k}. \tag{127}
$$

By the multinomial theorem the coefficient on (127) equals $\frac{m!}{p_1! \times \cdots \times p_k!}$, but since

$$D_{1j_1}^{p_1} \times \cdots \times D_{1j_k}^{p_k} = D_{1j_1}^{p_1^*} \times \cdots \times D_{1j_k}^{p_k^*} = D_{1j_1} \times \cdots \times D_{1j_k}$$

for any $\mathbf{p}, \mathbf{p}^* \in \mathcal{P}_{k,m}$, the coefficient on $D_{1j_1} \times \cdots \times D_{1j_k}$ after combining identical terms in (126) equals $\sum_{\mathbf{p} \in \mathcal{P}_{k,m}} \frac{m!}{p_1! \times \cdots \times p_k!}$. Putting these pieces together yields

$$\mathbb{E}\left[D_{i+}^m\right] = \sum_{k=1}^{m} \left(\sum_{\mathbf{p} \in \mathcal{P}_{k,m}} \frac{m!}{p_1! \times \cdots \times p_k!} \right) \mathbb{E}\left[\sum_{j_1 < \cdots < j_k} D_{ij_1} \times \cdots \times D_{ij_k} \right]$$

The expectations of the summands in $\sum_{j_1 < \cdots < j_k} D_{ij_1} \times \cdots \times D_{ij_k}$ are all identical with cardinality $\binom{N-1}{k}$. The assertion follows.

References

Acemoglu, D., Akcigit, U., & Kerr, W. (2016a). Networks and the macroeconomy: an empirical exploration. *NBER Macroeconomics Annual*, 31(1), 273 – 335.

Acemoglu, D., Akcigit, U., & Kerr, W. R. (2016b). Innovation network. *Proceedings of the National Academy of Sciences*, 113(41), 11482 – 11488.

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

Ahern, K. R. & Harford, J. (2014). The importance of industry links in merger waves. *Journal of Finance*, 69(2), 527 – 576.

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

Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R., & Olken, B. A. (2016). Network structure and the aggregation of information: theory and evidence from indonesia. *American Economic Review*, 106(7), 1663 – 1704.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4), 581 – 598.

Ambrus, A., Mobius, M., & Szeidl, A. (2014). Consumption risk-sharing in social networks. *American Economic Review*, 104(1), 149 – 182.

An, C., O'Malley, A. J., Rockmore, D. N., & Stock, C. D. (2018). Analysis of the u.s. patient referral network. *Statistics in Medicine*, 37(5), 847 – 866.

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

Andrews, D. W. & Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1), 119 – 157.

Angrist, J. (2014). The perils of peer effects. *Labour Economics*, 30, 98–108.

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

Arellano, M. & Bover, O. (1995). Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics*, 68(1), 29 – 51.

Arellano, M. & Hahn, J. (2007). *Advances in Economics and Econometrics: Theory and Applications*, volume 3, chapter Understanding bias in nonlinear panel models: some recent developments, (pp. 381 – 309). Cambridge University Press.: Cambridge.

Arellano, M. & Honoré, B. (2001). *Handbook of Econometrics*, volume 5, chapter Panel data models: some recent developments, (pp. 3229 – 3296). North-Holland: Amsterdam.

Aronow, P. M., Samii, C., & Assenova, V. A. (2017). Cluster-robust variance estimation for dyadic data. *Political Analysis*, 23(4), 564 – 577.

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

Atalay, E., Hortaçsu, A., & Syverson, C. (2014). Vertical integration and input flows. *American Economic Review*, 104(4), 1120 – 1148.

Athey, S., Eckles, D., & Imbens, G. W. (2018). Exact p-values for network interference. *Journal of American Statistical Association*, forthcoming.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., & Meghir, C. (2012). Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, 4(2), 134 – 167.

Auerbach, E. (2016). *A matching estimator for models with endogenous network formation.* Technical report, University of California - Berkeley.

Badev, A. (2017). *Discrete games in endogenous networks: equilibria and policy.* Technical Report 1705.03137, arXiv.

Bajari, P., Hong, H., Krainer, J., & Nekipelov, D. (2010a). Estimating static models of strategic interactions. *Journal of Business and Economic Statistics*, 28(4), 469 – 482.

Bajari, P., Hong, H., & Nekipelov, D. (2013). *Advances in Economics and Econometrics, Tenth World Congress*, volume 3, chapter Game theory and econometrics: a survey of some recent research, (pp. 3 – 52). Cambridge University Press: Cambridge.

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

Baldwin, R. & Taglioni, D. (2007). Trade effects of the euro: a comparison of estimators. *Journal of Economic Integration*, 22(4), 780 – 818.

Banerjee, A., Chandrasekhar, A. G., Dulfo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144), 363 – 370.

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

Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509 – 512.

Barabási, A.-L. & Bonabau, E. (2003). Scale-free networks. *Scientific American*, (pp. 50 – 59).

Barnett, M., Landon, B., O'Malley, A., Keating, N., & Christakis, N. (2011). Mapping physician networks with self-reported and administrative data. *Health Services Research*, 46(5), 1592 – 1609.

Barrot, J.-N. & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *Quarterly Journal of Economics*, 131(3), 1543 – 1592.

Beaman, L. A. (2011). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the us. *Review of Economic Studies*, 79(1), 128 – 161.

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: the structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110(1), 44 – 91.

Bech, M. L. & Atalay, E. (2010). The topology of the federal funds market. *Physica A: Statistical Mechanics and its Applications*, 389(22), 5223 – 5246.

Bellio, R. & Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling*, 5(3), 217 – 227.

Bernard, A. B., Moxnes, A., & Saito, Y. (2018). Geography and firm performance in the japanese production network. *Journal of Political Economy*. RIETI Discussion Paper 14-E-034.

Bhamidi, S., Bresler, G., & Sly, A. (2011). Mixing time of exponential random graphs. *The Annals of Applied Probability*, 21(6), 2146 – 2170.

Bhattacharya, S. & Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics*, 43(6), 2384 – 2411.

Bhattacharya, S. & Nain, A. (2011). Horizontal acquisitions and buying power: A product market analysis. *Journal of Financial Economics*, 99(1), 97 – 115.

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

Bickel, P. J. & Chen, A. (2009). A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50), 21068 – 21073.

Bickel, P. J., Chen, A., & Levina, E. (2011). The method of moments and degree distributions for network models. *Annals of Statistics*, 39(5), 2280 – 2301.

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

Bloch, F. & Jackson, M. O. (2006). Definitions of equilibrium in network formation games. *International Journal of Game Theory*, 34(3), 305 – 318.

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

Bloom, N., Schankerman, M., & van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4), 1347 – 1393.

Blume, L. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5, 387–424.

Blume, L. E., Brock, W. A., Durlauf, S. N., & Ioannides, Y. M. (2011). Identification of social interactions. In J. Benhabib, A. Bisin, & M. O. Jackson (Eds.), *Handbook of Social Economics*, volume 1B (pp. 853 – 964). Amsterdam: North-Holland.

Blume, L. E., Brock, W. A., Durlauf, S. N., & Jayaraman, R. (2015). Linear social interaction models. *Journal of Political Economy*, 123(2), 444 – 496.

Blundell, R. & Powell, J. L. (2003). *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, chapter Endogeneity in nonparametric and semiparametric regression models, (pp. 312 – 357). Cambridge University Press.

Bobollas, B., Janson, S., & Riordan, O. (2007). The phase transition in inhomogenous random graphs. *Random Structures and Algorithms*, 31(1), 3 – 122.

Bonhomme, S. & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3), 1147 – 1184.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1), 41 – 55.

Brock, W. & Durlauf, S. (2001). Discrete choice with social interactions. *Review of Economic Studies*, 68, 235–261.

Brown, B. W. & Newey, W. K. (1998). Efficient semiparametric estimation of expectations. *Econometrica*, 66(2), 453 – 464.

Callaert, H. & Veraverbeke, N. (1981). The order of the normal approximation for a studentized u-statistic. *Annals of Statistics*, 9(1), 194 – 200.

Cameron, A. C. & Golotvina, N. (2005). *Estimation of country-pair data models controlling for clustered errors: with international trade applications*. Technical Report 06-13, University of California - Davis.

Cameron, A. C. & Miller, D. L. (2014). *Robust inference for dyadic data*. Technical report, University of California - Davis.

Cattaneo, M., Crump, R., & Jansson, M. (2014). Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory*, 30(1), 176 – 200.

Cattelan, M. & Varin, C. (2013). *Advances in Theoretical and Applied Statistics*, chapter A model for correlated paired comparison data, (pp. 167 – 176). Springer-Verlag: Heidelberg.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1), 225 – 238.

Chamberlain, G. (1984). *Handbook of Econometrics*, volume 2, chapter Panel Data, (pp. 1247 – 1318). North-Holland: Amsterdam.

Chamberlain, G. (1985). *Longitudinal Analysis of Labor Market Data*, chapter Heterogeneity, omitted variable bias, and duration dependence, (pp. 3 – 38). Cambridge University Press: Cambridge.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica*, 60(3), 567 – 596.

Chandrasekhar, A. (2015). Econometrics of network formation. In Y. Bramoullé, A. Galeotti, & B. Rogers (Eds.), *Oxford Handbook on the Economics of Networks*. Oxford University Press.

Chatterjee, S. (2017). *Large Deviations for Random Graphs*. Cham, Switzerland: Springer.

Chatterjee, S. & Dembo, A. (2016). Nonlinear large deviations. *Advances in Mathematics*, 299(20), 396 – 450.

Chatterjee, S. & Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Annals of Statistics*, 41(5), 2428 – 2461.

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

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 82(2), 347 – 361.

Choi, T. Y. & Wu, Z. (2009). Triads in supply networks: theorizing buyer-supplier-supplier relationships. *Journal of Supply Chain Management*, 45(1), 8 – 25.

Choo, E. & Siow, A. (2006). Who marries whom and why? *Journal of Political Economy*, 114(1), 175 – 201.

Christakis, N. A., Fowler, J. H., Imbens, G. W., & Kalyanaraman, K. (2010). *An empirical model for strategic network formation*. NBER Working Paper 16039, National Bureau of Economic Research.

Chu, J. S. G. & Davis, G. F. (2016). Who killed the inner circle? the decline of the american corporate interlock network. *American Journal of Sociology*, 122(3), 714 – 754.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(S), S95 – S120.

Conley, T. G. & Udry, C. R. (2010). Learning about a new technology: pineapple in ghana. *American Economic Review*, 100(1), 35 – 69.

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

Crane, H. (2018). *Probabilistic Foundations of Statistical Network Analysis*. Boca Raton: CRC Press.

Crane, H. & Towsner, H. (2018). Relatively exchangeable structures. *Journal of Symbolic Logic*, 83(2), 416 – 442.

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

Daudin, J.-J., Picard, F., & Robin, S. (2008). A mixture model for random graphs. *Statistics and Computing*, 18(2), 173 – 183.

Davezies, L., d'Haultfoeuille, X., & Guyonvarch, Y. (2019). *Empirical process results for exchangeable arrayes*. Technical report, CREST-ENSAE.

Davis, G. F. (1991). Agents without principles? the spread of the poison pill through the intercorporate network. *Administrative Sciences Quarterly*, 36(4), 583 – 613.

Davis, G. F. (1996). The significance of board interlocks for corporate governance. *Corporate Governance: An International Review*, 4(3), 154 – 159.

De Benedictis, L., Nenci, S., Santoni, G., Tajoli, L., & Vicarelli, C. (2014). Network analysis of world trade using the baci-cepii dataset. *Global Economy Journal*, 14(3-4), 287–343.

de Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale*, 4, 251 – 299.

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

de Paula, Á. (2017). *Advances in Economics and Econometrics, Eleventh World Congress*, volume 1, chapter Econometrics of network models, (pp. 268 – 323). Cambridge University Press: Cambridge.

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

De Weerdt, J. (2004). *Insurance Against Poverty*, chapter Risk-sharing and endogenous network formation, (pp. 197 – 216). Oxford University Press: Oxford.

Dhyne, E., Magerman, G., & Rubínova, S. (2015). *The Belgian production network 2002-2012*. Working Paper 288, National Bank of Belgium.

Diaconis, P. (1977). Finite forms of de finetti's theorem on exchangeability. *Synthese*, 36(2), 271 – 281.

Diaconis, P. & Freedman, D. (1980). Finite exchangeable sequences. *Annals of Probability*, 8(4), 745 – 764.

Diaconis, P., Holmes, S., & Janson, S. (2008). Threshold graph limits and random threshold graphs. *Internet Mathematics*, 5(3), 267 – 320.

Diaconis, P. & Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica*, 28(1), 33 – 61.

Dooley, P. C. (1969). The interlocking directorate. *American Economic Review*, 59(3), 314 – 323.

Ductor, L., Fafchamps, M., & Goyal, S. (2014). Social networks and research output. *Review of Economics and Statistics*, 96(5), 936 – 948.

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

Efron, B. & Stein, C. (1981). The jackknife estimate of variance. *Annals of Statistics*, 9(3), 586 – 596.

Erikson, R. S., Pinto, P. M., & Rader, K. T. (2014). Dyadic analysis in international relations: a cautionary tale. *Political Analysis*, 22(4), 457 – 463.

Fafchamps, M. & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2), 326 – 350.

Fafchamps, M. & Lund, S. (2003). Risk sharing networks in rural philippines. *Journal of Development Economics*, 71(2), 261 – 287.

Fafchamps, M. & Minten, B. (2002). Returns to social network capital among traders. *Oxford Economic Papers*, 54(2), 173 – 206.

Fee, C. E. & Thomas, S. (2004). Sources of gains in horizontal mergers: evidence from customer, supplier, and rival firms. *Journal of Financial Economics*, 74(3), 423 – 460.

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

Frank, O. (1979). Moment properties of subgraph counts in stochastic graphs. *Annals of the New York Academy of Sciences*, 319(1), 207 – 218.

Frank, O. (1980). Transitivity in stochastic graphs and digraphs. *Journal of Mathematical Sociology*, 7(2), 199 – 213.

Frank, O. (1988). Triad count statistics. *Discrete Mathematics*, 72(1-3), 141 – 149.

Frank, O. (1997). Composition and structure of social networks. *Mathematiques, Informatique, et Sciences Humaines*, 137(1), 11 – 24.

Frank, O. & Snijders, T. A. B. (1994). Estimating the size of hidden populations using snowball sampling. *Journal of Official Statistics*, 10(1), 53 – 67.

Galichon, A. & Salanie, B. (2017). The econometrics and some properties of separable matching models. *American Economic Review*, 107(5), 251 – 55.

Gao, C. & Lafferty, J. (2017). *Testing network structuring using relations between small subgraph probabilities*. arxiv 1704.06742, arXiv.

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

Gaulier, G. & Zignago, S. (2010). *BACI: International trade database at the product-level*. Working Paper 2010-23, CEPII.

Gofman, M. (2017). Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*, 124(1), 113 – 146.

Goldenberg, A., Zheng, A., Fienberg, S. E., & Airoldi, E. M. (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2), 129–333.

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

Goodman, L. (2017). The effect of the affordable care act medicaid expansion on migration. *Journal of Policy Analysis and Management*, 36(1), 211 – 238.

Gould, R. V. & Fernandez, R. M. (1989). Structures of mediation: a formal approach to brokerage in transaction networks. *Sociological Methodology*, 19, 89 – 126.

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S), S85 – S118.

Graham, B. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3), 643–660.

Graham, B. S. (2011). Efficiency bounds for missing data models with semiparametric restrictions. *Econometrica*, 79(2), 437 – 452.

Graham, B. S. (2013). Comparative static and computational methods for an empirical one-to-one transferable utility matching model. *Advances in Econometrics*, 31, 153 – 181.

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7(1), 465–485.

Graham, B. S. (2016). *Homophily and transitivity in dynamic network formation.* NBER Working Paper 22186, National Bureau of Economic Research.

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

Graham, B. S. (2018a). Dyadic regression. Lecture Notes.

Graham, B. S. (2018b). Identifying and estimating neighborhood effects. *Journal of Economic Literature*, 56(2), 450 – 500.

Graham, B. S. & de Paula, A., Eds. (2020). *The Econometric Analysis of Network Data.* Academic Press.

Graham, B. S., Imbens, G. W., & Ridder, G. (2010). *Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach.* Working Paper 16499, NBER.

Graham, B. S., Imbens, G. W., & Ridder, G. (2014). Complementarity and aggregate implications of assortative matching: a nonparametric analysis. *Quantitative Economics*, 5(1), 29 – 66.

Graham, B. S., Imbens, G. W., & Ridder, G. (2018). Identification and efficiency bounds for the average match function under conditionally exogenous matching. *Journal of Business and Economic Statistics*.

Graham, B. S., Niu, F., & Powell, J. L. (2019). *Kernel density estimation for undirected dyadic data*. Technical report, University of California - Berkeley.

Graham, B. S. & Pelican, A. (2020). *The Econometrics of Social and Economic Networks*, chapter Testing for externalities in network formation using simulation",. Elsevier: Amsterdam.

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

Granovetter, M. S. (1985). Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, 91(3), 481 – 510.

Green, A. & Shalizi, C. R. (2017). *Bootstrapping exchangeable random graphs*. Technical Report 1711.00813v1, arXiv.

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

Hagedoorn, J. (2002). Inter-firm r & d partnerships: an overview of major trends and patterns since 1960. *Research Policy*, 31(4), 477 – 492.

Hahn, J. & Newey, W. K. (2004). Jackknife and analytical bias reduction for nonlinear panel data models. *Econometrica*, 72(4), 1295 – 1319.

Hausman, J. & Taylor, W. (1981). Panel data and unobservable individual effects. *Journal of Econometrics*, 49(6), 1377 – 1398.

He, R. & Zheng, T. (2013). Estimation of exponential random graph models for large social networks via graph limits. In *Advances in Social Networks Analysis and Mining (ASONAM)*.

Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. *Review of Economic Studies*, 64(4), 487 – 535.

Heckman, J. J. & Vytlacil, E. J. (2007). *Handbook of Econometrics*, volume 6B, chapter Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation, (pp. 4779 – 4874). North-Holland: Amsterdam.

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

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: trading partners and trading volumes. *Quarterly Journal of Economics*, 123(2), 441 – 487.

Hensvik, L. & Skans, O. N. (2016). Social networks, employee selection, and labor market outcomes. *Journal of Labor Economics*, 34(4), 825 – 867.

Hilgerdt, F. (1943). The case for multilateral trade. *American Economic Review*, 33(1 (Part 2)), 393 – 407.

Ho, K. (2009). Insurer-provider networks in the medical care market. *American Economic Review*, 99(1), 393 – 430.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3), 293 – 325.

Hoff, P., Raftery, A., & Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97, 1090–1098.

Holland, P. W. & Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, 76(3), 492 – 513.

Holland, P. W. & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7, 1 – 45.

Holland, P. W. & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373), 33 – 50.

Honoré, B. E. & Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, 64(1-2), 241 – 278.

Hoover, D. N. (1979). *Relations on probability spaces and arrays of random variables*. Technical report, Institute for Advanced Study, Princeton, NJ.

Imbens, G. W. & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5 – 86.

Ioannides, Yannis, M. & Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4), 1056 – 1093.

Isakov, A., Fowler, J. H., Airoldi, E. M., & Christakis, N. A. (2019). The structure of negative social ties in rural village networks. *Sociological Science*, 6(8), 197 – 217.

Jackson, M. & Watts, A. (2001). The existence of pairwise stable networks. *Seoul Journal of Economics*, 14(3), 299 – 321.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton: Princeton University Press.

Jackson, M. O., Rodriguez-Barraquer, T., & Tan, X. (2012). Social capital and social quilts: network patterns of favor exchange. *American Economic Review*, 102(5), 1857–1897.

Jackson, M. O. & Rogers, B. W. (2007). Relating network structure to diffusion properties through stochastic dominance. *B.E. Journal of Theoretical Economics*, 7(1), (Advances) Article 6.

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

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

Jackson, M. O. & Yariv, L. (2011). *Handbook of Social Economics*, volume 1A, chapter Diffusion, strategic interaction, and social structure, (pp. 645 – 678). North-Holland: Amsterdam.

Jackson, M. O. & Zenou, Y. (2015). *Handbook of Game Theory*, volume 4, chapter Games on networks, (pp. 95 – 163). North-Holland: Amsterdam.

Jaffe, A. (1986). Technological opportunity and spillovers of r&d: evidence from firms' patents, profits, and market value. *American Economic Review*, 76(5), 984 – 1001.

Janson, S., Luczak, T., & Rucinski, A. (2000). *Random Graphs*. New York: John Wiley & Sons, Inc.

Janson, S. & Nowicki, K. (1991). The asymptotic distributions of generalized u-statistics with applications to random graphs. *Probability Theory and Related Fields*, 90(3), 341 – 375.

Janssen, P. (1994). Weighted bootstrapping of u-statistics. *Journal of Statistical Planning and Inference*, 38(1), 31–41.

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

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

Johnsson, I. & Moon, H. R. (2017). *Estimation of peer effects in endogenous social networks: control function approach*. USC-INET Research Paper No. 17-25, University of Souther California.

Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles*. New York: Springer-Verlag.

Keane, M. P. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, 62(1), 95 – 116.

Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H., & Christakis, N. A. (2015). Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet*, 386(9989), 145 – 153.

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

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

Kranton, R. & Minehart, D. (2001). A theory of buyer-seller networks. *American Economic Review*, 91(3), 485–508.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., & Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(2), 204 – 213.

Kuersteiner, G. M. (2019). *Limit theorems for data with network structure*. arXiv arXiv:1908.02375, University of Maryland.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. New York: Springer.

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

Leung, M. (2019). A weak law for moments of pairwise stable networks. *Journal of Econometrics*, 210(2), 310 – 326.

Leung, M. & Moon, H. R. (2019). *Normal approximation in large network models*. arXiv arXiv:1904.11060, University of Southern California.

Lindsey, B. G. (1988). Composite likelihood. *Contemporary Mathematics*, 80, 221 – 239.

Loury, G. C. (2002). *The Anatomy of Racial Inequality*. Cambridge, MA: Harvard University Press.

Lovász, L. (2012). *Large Networks and Graph Limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society.

Lovász, L. & Szegedy, B. (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96, 933–957.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem,. *Review of Economic Studies*, 60(3), 531 – 542.

Mao, L. (2018). On causal estimation using u-statistics. *Biometrika*, 105(1), 215 – 220.

Marsden, P. V. (1987). Core discussion networks of americans. *American Sociological Review*, 52(1), 122 – 131.

Mayda, A. M. (2010). International migration: a panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, 23(4), 1249 – 1274.

McFadden, D. (1974). *Frontiers in Econometrics*, chapter Conditional logit analysis of qualitative choice behavior, (pp. 105 – 142). Academic Press: New York.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 67(5), 995 – 1026.

McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social isolation in america: changes in core discussion networks over two decades. *American Sociological Review*, 71(3), 353 – 375.

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

Mele, A. & Zhu, L. (2017). *Approximate variational estimation for a model of network formation*. Technical report, John Hopkins University.

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

Menzel, K. (2017). *Bootstrap with clustering in two or more dimensions*. Technical Report 1703.03043v2, arXiv.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594), 824 – 827.

Min, S. (2019). *Network of loyalty programs: a sequential formation*. Technical report, University of California - Berkeley.

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

Monderer, D. & Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, 14(1), 124 – 143.

Moreno, J. L. (1934). *Who Shall Survive?* New York: Beacon House.

Munski, K. (2003). Networks in the modern economy: Mexican migrants in the us labor market. *Quarterly Journal of Economics*, 118(2), 549 – 599.

Nadler, C. (2015). Networked inequality: evidence from freelancers. University of California - Berkeley.

Newey, W. K. & McFadden, D. (1994). *Handbook of Econometrics*, volume 4, chapter Large sample estimation and hypothesis testing, (pp. 2111 – 2245). North-Holland: Amsterdam.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404 – 409.

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

Nowicki, K. (1991). Asymptotic distributions in random graphs with applications to social networks. *Statistica Neerlandica*, 45(3), 295 – 325.

Nowicki, K. & Wierman, J. C. (1988). Subgraph counts in random graphs using incomplete u-statistics methods. *Discrete Mathematics*, 72(1-3), 299 – 310.

Olhede, S. C. & Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 11(41), 14722 – 14727.

Orbanz, P. & Roy, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable randoms structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2), 437 – 461.

Ortega, F. & Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies*, 1(1), 47 – 74.

Ostrovsky, M. (2008). Stability in supply chain networks. *American Economic Review*, 98(3), 897 – 923.

Owen, A. B. (2007). The pigeonhole bootstrap. *Annals of Applied Statistics*, 1(2), 386 – 411.

Pakes, A. & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5), 1027 – 1057.

Pelican, A. & Graham, B. S. (2019). *Testing for strategic interaction in social and economic network formation*. Technical report, University of California - Berkeley.

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

Picard, F., Daudin, J. J., Koskas, M., Schbath, S., & Robin, S. (2008). Assessing the exceptionality of network motifs. *Journal of Computational Biology*, 15(1), 1 – 20.

Pržulj, N., Corneil, D. G., & Jurisica, I. (2004). Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18), 3508–3515.

Qu, X. & Lee, L.-F. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2), 209 – 232.

Ridder, G. & Sheng, S. (2017). *Estimation of large network formation games*. Technical report, University of California - Los Angeles.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007a). An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, 29(2), 173 – 191.

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007b). Recent developments in exponential random graph (p^*) models for social networks. *Social Networks*, 29(2), 192 – 215.

Rose, A. (2005). Which international institutions promote international trade? *Review of International Economics*, 13(4), 682 – 698.

Rose, A. K. (2004). Do we really know that the wto increases trade? *American Economic Review*, 94(1), 98 – 114.

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of American Statistical Association*, 102(477), 191 – 200.

Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, 6(4), 377 – 400.

Russett, B. M. & Oneal, J. R. (2001). *Triangulating Peace: Democracy, Interdependence and International Organizations*. New York: Norton.

Santos Silva, J. & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4), 641 – 658.

Santos Silva, J. & Tenreyro, S. (2010). Currency unions in prospect and retrospect. *Annual Review of Economics*, 2, 51 – 74.

Saygin, P. O., Weber, A., & Weynandt, M. A. (2014). *Coworkers, networks and job search outcomes*. Iza dp no. 8174, IZA Institute of Labor Economics.

Schwartz, A. L. & Sommers, B. D. (2014). Moving for medicaid? recent eligibility expansions did not induce migration from other states. *Health Affairs*, 33(1), 88 – 94.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc.

Serpa, J. C. & Krishnan, H. (2017). The impact of supply chains on firm-level productivity. *Management Science*.

Shalizi, C. R. (2016). *Lecture 1: Conditionally-independent dyad models*. Lecture note, Carnegie Mellon University.

Shalizi, C. R. & Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *Annals of Statistics*, 41(2), 508 – 535.

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

Simmel, G. (1908). *Soziologie. Untersuchungen über die formen der vergesellschaftung.* Leipzig: Duncker & Humblot.

Snijders, T. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1–40.

Snijders, T. A. B. & Borgatti, S. P. (1999). Non-parametric standard errors and tests for network statistics. *Connections*, 22(2), 61 – 70.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1), 99 – 153.

Tabord-Meehan, M. (2018). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business and Economic Statistics*.

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

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

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

Tomasello, M. V., Napoletano, M., Garas, A., & Schweitzer, F. (2017). The rise and fall of r&d networks. *Industrial and Corporate Change*, 26(4), 617 – 646.

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton, NJ: Princeton University Press.

Townsend, R. M. (1994). Risk and insurance in village india. *Econometrica*, 62(3), 539 – 591.

Udry, C. R. (1994). Risk and insurance in a rural credit market: an empirical investigation in northern nigeria. *Review of Economic Studies*, 61(3), 495 – 526.

Uetake, K. & Watanabe, Y. (2013). Estimating supermodular games using rationalizable strategies. *Advances in Econometrics*, 31(1), 233 – 247.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van Duijn, M. A. J., Snijders, T. A. B., & Zijlstra, B. J. H. (2004). p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2), 234 – 254.

VanderWeele, T. J. & An, W. (2013). *Handbook of Causal Analysis for Social Research*, chapter Social networks and causal inference, (pp. 353 – 374). Springer: Dordrecht.

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1), 5 – 42.

Volfosky, A. & Airoldi, E. M. (2016). Sharp total variational bounds for finitely exchangeable arrays. *Statistics and Probability Letters*, 114, 54 – 59.

Wainwright, M. J. & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2), 1 – 305.

Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wasserman, S. S. (1977). Random directed graph distributions and the triad census in social networks. *Journal of Mathematical Sociology*, 5(1), 61 – 86.

Wooldridge, J. M. (2005). *Identification and inference for econometric models*, chapter Unobserved heterogeneity and the estimation of average partial effects, (pp. 27 – 55). Number 3. Cambridge University Press: Cambridge.

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

Zijlstra, B. J. H., van Duijn, M. A. J., & Snijders, T. A. B. (2009). Mcmc estimation for the p_2 network regression model with crossed random effects. *British Journal of Mathematical and Statistical Psychology*, 62(1), 143 – 166.