

Econometrics of network models

Áureo de Paula

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP06/16

Econometrics of Network Models *

Áureo de Paula[†]

University College London, São Paulo School of Economics,
CeMMAP, IFS and CEPR

This Version: January 2016

Abstract

In this article I provide a (selective) review of the recent econometric literature on networks. I start with a discussion of developments in the econometrics of group interactions. I subsequently provide a description of statistical and econometric models for network formation and approaches for the joint determination of networks and interactions mediated through those networks. Finally, I give a very brief discussion of measurement issues in both outcomes and networks. My focus is on identification and computational issues, but estimation aspects are also discussed.

*This article was prepared for the 2015 Econometric Society World Congress in Montreal, Canada. I am grateful to Anton Badev, Yann Bramoullé, Arun Chandrasekhar, Andreas Dzemski, Bo Honoré, Chris Julliard, Chuck Manski, Angelo Mele, Bryony Reich, Seth Richards-Shubik, Adam Rosen and Elie Tamer for comments on an earlier draft of the paper. The author gratefully acknowledges financial support from the European Research Council through Starting Grant 338187 and the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001.

[†]University College London, São Paulo School of Economics, CeMMAP, IFS and CEPR. E-mail: apaula@ucl.ac.uk

1 Introduction

Networks are “vulgar.” By that I mean: they are commonplace, ordinary. Although markets are the usual forum where economic phenomena take place, many social and economic behaviors are not mediated by prices. A great many studies have investigated the existence and quantification of spillover effects in education (e.g., Sacerdote (2010)), in the labor market (e.g., Topa (2001)) and, more recently, on non-cognitive outcomes (e.g., Neidell and Waldfogel (2010) and Lavy and Sand (2015)). Many other behaviors are mediated through prices, but in a way that it matters how agents are in contact with each other. Production and financial networks are natural examples (e.g., Atalay, Hortacsu, Roberts, and Syverson (2011) for the first and Denbee, Julliard, Li, and Yuan (2014) or Bonaldi, Hortacsu, and Kastl (2014) for the second). The main conduit in these examples is the intervening role of “connections”: who is in direct or indirect contact with whom. This structure defines (and is possibly defined by) how information, prices and quantities reverberate in a particular social or economic system. This recognition has sparked a growing literature on various aspects of networks and their role in explaining various social and economic phenomena among economic theorists, empirical researchers and, more recently, econometricians.

This article aims at providing a (selective) overview of some recent advances and outstanding challenges in the applied econometric literature on this topic. I focus on both the role of networks in aiding the measurement of outcomes determined on an underlying network structure and on the formation of such structures. I also provide a brief discussion on measurement issues related to both tasks. Given constraints in space and expertise, this article is not exhaustive. In fact, the identification and measurement of network-related phenomena has drawn increasing attention in fields as diverse as macroeconomics, industrial organization, finance, and trade, which I do not discuss in this review. There are also subclasses of network-related phenomena of empirical and econometric interest that I do not cover, such as bargaining and matching in bipartite graphs. Some of the ideas below may prove useful to developments in those areas nonetheless.

The article proceeds as follows. The following section provides a palette of basic definitions and terminology used recurrently throughout this paper. Because those are well covered elsewhere, I am deliberately succinct. Section 3 covers topics related to models where particular outcomes of interest are mediated by predetermined networks. The subsequent section focuses on econometric models for the determination of the networks themselves and also discusses the joint determination of outcomes and networks. Section 5 provides a brief

discussion of measurement issues related to networks and outcomes. The last section concludes.

2 Some Basic Terminology and Concepts

Networks are typically represented by graphs. A graph g is a pair of sets $(\mathcal{N}_g, \mathcal{E}_g)$ of *nodes* (or *vertices*) \mathcal{N}_g and *edges* (or *links* or *ties*) \mathcal{E}_g . I will denote the cardinality of these sets by $|\mathcal{N}_g|$ and $|\mathcal{E}_g|$, respectively. For our purposes, vertices are economic agents: individuals, households, firms or other entities of interest. The set of nodes is usually conceived as a finite set of elements, though in principle the node set can also be infinite (e.g., Berge (1962)). An edge represents a link or connection between two nodes in \mathcal{N}_g . A graph is undirected when \mathcal{E}_g is the set of unordered pairs with elements in \mathcal{N}_g , say $\{i, j\}$ with $i, j \in \mathcal{N}_g$. (The multiset $\{i, i\}$ with $i \in \mathcal{N}_g$ is a possibility, but I abstract away from self-links here.) This type of graph is appropriate in representing reciprocal relationships between two vertices. An example are (reciprocal) informal risk-sharing networks based on kinship or friendship (e.g., Fafchamps and Lund (2003)). To accommodate directional relationships, edges are best modeled as ordered pairs, say $(i, j) \in \mathcal{N}_g \times \mathcal{N}_g$. These graphs, known as directed graphs (or digraphs), are more adequate for handling relationships that do not require reciprocity or for which direction carries a particular meaning, as in a supplier-client relationship in a production network (e.g., Atalay, Hortacsu, Roberts, and Syverson (2011)). Further generalizations allow for weighted links, perhaps representing distances between two individuals or the intensity of a particular relationship. Such weights can be represented as a mapping from the space of pairs (unordered or ordered) into the real line. Diebold and Yilmaz (2015), for example, consider a (directed, weighted) graph obtained from the forecast-error variance decomposition for a given class of economic variables of interest. The nodes in this case would be seen as different entities, like stocks or firms, for example, and the weight of a directed link from node i to node j gives the proportion of the forecast error variance in variable of interest for node i (e.g., return or volatility if nodes represent stocks) explained by shocks to node j .¹

A common representation of a graph is through its $|\mathcal{N}_g| \times |\mathcal{N}_g|$ adjacency or incidence matrix W , where each line represents a different node. The components of W mark whether

¹They define a few measures based on this network representation to keep track of “connectedness” of a particular economic system through time. Their total connectedness measure, for example, is given by the total sum of the weights across edges divided by the number of nodes.

an edge between nodes i and j (or from i to j in a digraph) is present or not and possibly its weight (in weighted graphs). The adjacency matrix allows one to translate combinatorial operations into linear algebraic ones and can be quite useful in several settings. For an adjacency matrix W to a simple graph (i.e., no self-links and at most one link between any pair of nodes), the ij element of matrix W^k , $k \in \{1, \dots, N - 1\}$, for instance, produces the number paths of length k between i and j . Two graphs are said to be isomorphic if their adjacency matrices can be obtained from each other, through multiplication by a permutation matrix, for example. This translates into a relabeling of the vertices in the corresponding graphs.

2.1 Vertex Features

Various measures can then be defined to characterize a particular vertex in the graph, to relate two or more vertices on a graph, or to represent a global feature of the graph at hand. (Although some of the notions mentioned below apply to more general networks, in what follows I focus on simple, unweighted graphs for ease of exposition.) An important characteristic for a particular vertex i , for example, is the set of neighbors incident with that vertex in a graph g , denoted by $N_i(g)$. In an undirected graph g , this set is given by $\{j : \{i, j\} \in \mathcal{E}_g\}$, and a similar definition can be given for directed graphs. The cardinality of this set is known as the “degree” of that node, and one can then talk about the relative frequency of degrees in a given graph as a whole. (In directed graphs, one can further distinguish “in-degrees” and “out-degrees” relating to inward and outward edges from and to a given node.) A “dense” graph, for instance, is then one in which nodes display a lot of connections, and a common measure of density is the average degree divided by $|\mathcal{N}_g| - 1$, which is the maximum number of possible links available to any given node. Given two nodes i and j in an undirected graph, a sequence of nodes ($i \equiv i_1, \dots, i_{K-1}, i_K \equiv j$) defines a “walk” if every edge $\{i_k, i_{k+1}\} \in \mathcal{E}_g$. A “cycle” is a walk where $i_1 = i_K$, and a tree is a graph without cycles. A “path” is a walk where no vertex is visited more than once. (One can similarly define paths and walks on directed graphs.) It is common to define the (geodesic) distance between these two as the shortest path between those two nodes. A graph is then said to be connected if the distance between any two vertices is finite (i.e., there is at least one path between those nodes). A component of a graph is then a maximal connected subgraph, where a subgraph is defined by a subset of nodes from \mathcal{N}_g together with a subset of edges from \mathcal{E}_g between elements in the subset of nodes under consideration. (A maximal connected

subgraph is not strictly contained in any other connected subgraph of g .) A quantity encoding the connectedness of a network is given by the second smallest eigenvalue of the Laplacian matrix, defined as $L = \text{diag}(W\mathbf{1}) - W$, where $\text{diag}(W\mathbf{1})$ is a diagonal matrix with the row-sums of W along the diagonal. This value, known as the algebraic connectivity or the Fiedler value of the graph, provides a measure of how easy it is to break the graph into disconnected components by selectively eliminating a small number of edges (see Kolaczyk (2009)).

One can also define various measures to characterize the typical network structure in the vicinity of a given vertex. For brevity, I only mention a basic taxonomy of such measures as specific definitions are available in most introductory texts on the subject (see, for example, the excellent overview in Jackson (2009)). An important network aspect of particular interest in social settings is the degree of “clustering” in the system, intuitively summarized by the propensity that two neighbors to a given node are also themselves directly linked and different clustering metrics are available to quantify this feature in a network. Studying volunteer work in the civil rights movement in the United States, for example, McAdam (1986) suggests that “[a]lthough weak ties may be more effective as diffusion channels (Granovetter (1973)), strong ties embody greater potential for influencing behavior. Having a close friend engage in some behavior is likely to have more of an effect on someone than if a friend of a friend engages in that same behavior. Apparently, the above was true of the Freedom Summer project” (p.80).² Theoretically, it may be easier for clustered individuals to coordinate on certain collective actions since clustering may facilitate common knowledge (Chwe (2000)).

Another feature of potential interest in economic and social networks is the degree of “centrality” of a given vertex, and various measures of centrality are also available. Those aim at characterizing how important a given node is in comparison to the remaining nodes in g . Aside from how connected a given vertex is (degree centrality) or how far on average a vertex is from any other vertex in the network (closeness centrality), one can also compute the betweenness centrality, illustrating how crucial a given node is in connecting individuals. Another family of popular centrality measures includes those based on features of the adjacency matrix aimed at summarizing a node’s centrality in reference to its neighbors centrality (more on this later). The simplest of these measures is the eigenvector centrality (a.k.a. Gould’s index of accessibility), corresponding to the dominant eigenvector of the adjacency matrix (Gould (1967), Bonacich (1972)).³ Among the most popular metrics in this

²To properly parse the effect of clustering, one would of course like to account for homophily among those who would already be prone to activism. If those individuals tend to associate with other like-minded individuals, this effect of “strong ties” will be confounded.

³The most profitable of these measures is perhaps Google’s PageRank index (Brin and Page (1998)).

family were those proposed by Katz (1953) and Bonacich (1987). The Katz centrality of a node i can be motivated by ascribing a value of $\tilde{\beta}^k > 0$ to each connection reached by a walk of length k . Since the (i, j) entry in W^k counts the number of walks between i and j , if one adds up the weights for each individual, one has a centrality measure for each individual given by the components of the vector $\tilde{\beta}W\mathbf{1} + \tilde{\beta}^2W^2\mathbf{1} + \tilde{\beta}^3W^3\mathbf{1} + \dots$. If $\tilde{\beta}$ is below the reciprocal of W 's largest eigenvalue, we can write the above as $\tilde{\beta}(\mathbf{I} - \tilde{\beta}W)^{-1}W\mathbf{1}$, where $\tilde{\beta}$ is a small enough positive number. The Bonacich centrality generalizes this formula to a two-parameter index defined by the vector $\alpha(\mathbf{I} - \tilde{\beta}W)^{-1}W\mathbf{1}$. Recently, Banerjee, Chandrasekhar, Duflo, and Jackson (2014) introduced another centrality measure, which they named diffusion centrality and which subsumes the degree, eigenvector, and Katz-Bonacich centralities as special cases. Such measures turn out to play an important role in the analysis of games and dissemination on networks (e.g., Ballester, Calvó-Armengol, and Zenou (2006) and the survey by Zenou (2015)).

2.2 Random Graphs

Having characterized the objects of interest here, one is then well-positioned to discuss data-generating processes giving rise to potentially observable social and economic networks and their sampling. Letting \mathcal{G} be a particular set of graphs, one can define a probability space $(\mathcal{G}, \sigma(\mathcal{G}), \mathbb{P})$, where $\sigma(\mathcal{G})$ is a σ -algebra of events in the sample space \mathcal{G} and \mathbb{P} is a probability measure on the measurable space $(\mathcal{G}, \sigma(\mathcal{G}))$. These models can and usually are indexed by features common to the graphs in \mathcal{G} , like the number of vertices and/or other features. One of the early models, for example, imposes a uniform probability on the class of graphs with a given number of nodes, $n = |\mathcal{N}_g|$, and a particular number of edges, $e = |\mathcal{E}_g|$, for $g \in \mathcal{G}$ (see Erdős and Rényi (1959) and Erdős and Rényi (1960)). Another basic, canonical random graph model is one in which the edges between any two nodes follow an independent Bernoulli distribution with equal probability, say p . For a large enough number of nodes and sufficiently small probability of link formation p , the degree distribution approaches a Poisson distribution, and the model is consequently known as the Poisson random-graph model. This class of models appears in Gilbert (1959) and Erdős and Rényi (1960) and has since been studied extensively. Nevertheless, they fail to reproduce important dependencies observed in social and economic networks. One category of models that aims at a better representation of the regularities usually encountered in social systems involves models where nodes are incorporated into the graph sequentially and form ties more or less randomly. These

models are able to reproduce features that the simple framework above is incapable of. For example, whereas the models above deliver degree distributions with exponential tails, many datasets appear to feature polynomial, Pareto tails. Models like the “preferential attachment” model (Barabási and Albert (1999)), whereby the establishment of new links is more likely for higher-degree existing nodes, produce Pareto tails (as well as other regularities usually observed; see the presentation in Jackson (2009) or Kolaczyk (2009) for a more thorough exposition).⁴

Another alternative is to rely on more general (static) random graph models that explicitly acknowledge the probabilistic dependencies in link formation. Such dependencies can be approached using probabilistic graphical models as in Frank and Strauss (1986). Probabilistic graphical models provide a (non-random) graphical representation of the probabilistic dependencies among a set of random vectors (e.g., Koller and Friedman (2009)). In the context of random (undirected) graphs with $|\mathcal{N}_g| = n$ vertices, those random variables are the $n(n - 1)/2$ (random) edges potentially formed between the n nodes. To represent a given stochastic dependence structure in the formation of edges, Frank and Strauss rely on a (non-random) dependency graph having as vertices all the potential edges between the elements of an original set of nodes of interest. In this dependency graph, links between two nodes (i.e., (random) edges in the random graph of interest) are present if these two random variables are conditionally dependent given the remaining random variables (i.e., the remaining (random) edges in the original random graph of interest). For example, since edges form independently in the Poisson graph model, its dependency graph is an empty one, with no links among its nodes (which are the random edges in the original Poisson network of interest). On the other hand, if the probability that an edge between i and j depends on the existence of a link between i and k given the remaining edges in the graph, the dependency graph will feature a link between nodes that represent $\{i, j\}$ and $\{i, k\}$. Applying results previously employed in the spatial statistics literature (i.e., the Hammersley-Clifford Theorem, see Besag (1974)), one has

$$\mathbb{P}(G = g) \propto \exp\left(\sum_{C \subset g} \alpha_C\right),$$

where $\alpha_C \in \mathbb{R}$ and C indexes all completely connected subgraphs (i.e., “cliques”) of the

⁴It should be noted that, while polynomial tails characterize the degree distribution for many networks, the evidence in favor of Pareto tails is less consensual that it might appear from a casual reading of the literature. This is highlighted, for example, in Pennock, Flake, Lawrence, Glove, and Giles (2002), Jackson and Rogers (2007) and Clauset, Shalizi, and Newman (2009), who find that many networks appear to be better characterized by non-polynomial tails.

(non-random) dependency graph representing the random graph model of interest.⁵

In general, the task of enumerating the set of cliques can be computationally complex. This expression can nevertheless be simplified for many interesting specific dependency structures. Frank and Strauss (1986), for example, focus on (pairwise Markov) random graphs where two (random) edges that do not share a vertex are conditionally independent given the other remaining (random) edges (and hence are not linked in the dependency graph). It reflects the intuition that ties are not independent of each other, but their dependency arises only through those who are directly involved in the connections in question. This, and a homogeneity assumption (i.e., that all graphs that are the same up to a permutation of vertices have the same probability), delivers that

$$\mathbb{P}(G = g) \propto \exp \left(\alpha_0 t + \sum_{k=1} \alpha_k s_k \right),$$

where t is the number of triangles (completely connected triples of vertices) and s_k is the number of k -stars (tuples of $k + 1$ vertices where one of the vertices has degree k and the remaining ones have degree one). (Notice that the Poisson model is a specific case of the above model, where $\alpha_0 = \alpha_2 = \dots = \alpha_k = 0$.) This structure suggests a class of probabilistic models that reproduce the exponential functional form above even in cases where the Markov property used by Frank and Strauss does not hold. Those models are such that $\mathbb{P}(G = g) \propto \exp \left(\sum_{k=1}^p \alpha_k S_k(g) \right)$, where $S_k(g), k = 1, \dots, p$ enumerate certain features of the graph g . These would be characteristics like the number of edges, the number of triangles and possibly many others. These models are known as exponential random graph models (or p^* models in the social sciences literature, see Robins, Pattison, Kalish, and Lusher (2007)) and can be extended beyond undirected random graphs. The models above constitute an exponential family of distribution over (random) graphs and exponential distributions (e.g., Bernoulli, Poisson) have well-known probabilistic and statistical properties. For example, the vector $(S_1(g), \dots, S_p(g))$ constitute a p -dimensional sufficient statistic for the parameters $(\alpha_1, \dots, \alpha_p)$. All the models above (and many others) are presented in detail elsewhere (e.g., Bollobás (2001), Jackson (2009), Kolaczyk (2009)) and I will selectively discuss features and difficulties as they articulate with the literature reviewed here.

⁵This representation is sometimes expressed in terms of *maximal* cliques. Any representation based on nonmaximal cliques can be converted into one based on maximal cliques by redefining α_c for a maximal clique as the sum of the α s on the subsets of that clique (Jordan and Wainwright (2008)).

3 Outcomes on Networks

As pointed out in the introduction, many social and economic outcomes are mediated by interactions among the entities involved (individuals, households, firms). In fact, the interaction structure can be instrumental in shaping the outcomes in various social and economic settings. Although the very determination of the social links on which those outcomes are resolved is plausibly informed by those outcomes or expectations about those outcomes in many cases, we start by assuming here that the peer structure, i.e., the ties among the various individuals involved, is determined independently.

3.1 Linear Models

The canonical representation for the joint determination of outcomes mediated by social interactions builds on the linear specification presented in Manski (1993). This representation postulates that the individual outcome variable for individual $i \in \{1, \dots, N\}$, y_i , is determined according to

$$y_i = \alpha + \beta \sum_{j=1}^N W_{ij} y_j + \eta x_i + \gamma \sum_{j=1}^N W_{ij} x_j + \epsilon_i, \quad \mathbb{E}(\epsilon_i | \mathbf{x}, W) = 0 \quad (1)$$

where $j \in \{1, \dots, N\}$, x_i represents a covariate observed by the researcher (with $\mathbf{x} = [x_1 \dots x_N]^\top$), ϵ_i represents a latent variable unobserved by the researcher, and W_{ij} are entries in the adjacency matrix that register the social network structure. For example, if y_i is affected by the average of all other individuals' outcomes and covariates, $W_{ij} = (N - 1)^{-1}$ and $W_{ii} = 0$. (This model is a spatial auto-regressive model in spatial statistics.) I assume here that x is scalar though the arguments hold more generally. Stacking the individual equations above, one then obtains

$$\mathbf{y}_{N \times 1} = \alpha \mathbf{1}_{N \times 1} + \beta W_{N \times N} \mathbf{y}_{N \times 1} + \eta \mathbf{x}_{N \times 1} + \gamma W_{N \times N} \mathbf{x}_{N \times 1} + \epsilon_{N \times 1}, \quad \mathbb{E}(\epsilon | \mathbf{x}, W) = \mathbf{0} \quad (2)$$

with $\mathbf{1}$ as a vectors of 1s. Whereas I take this as the point of departure for my presentation of (linear) social interaction models, I note that the specification above can be obtained from more primitive foundations (e.g., Blume, Brock, Durlauf, and Jayaraman (2015)). Supposing that $1/\beta$ is not an eigenvalue of W , the equations above produce the following reduced form

system:

$$\mathbf{y} = \alpha(\mathbf{I} - \beta W)^{-1}\mathbf{1} + (\mathbf{I} - \beta W)^{-1}(\eta\mathbf{I} + \gamma W)\mathbf{x} + (\mathbf{I} - \beta W)^{-1}\boldsymbol{\epsilon}, \quad (3)$$

where \mathbf{I} is an identity matrix of order N .

In his celebrated article, Manski examined the identification of the various parameters above (although using a different representation, see below). In doing so, he distinguished social influences between endogenous and exogenous (or contextual) effects. The latter represents any influence encoded by peer (observable) characteristics. The former translates into the influence of peers' outcomes on one's own outcomes. Although related, those two have different repercussions: endogenous effects act as conduits for the reverberation of shocks, leading to a multiplier effect, which is absent if contextual effects are the main driving mechanism for social influences. The separation of these two parameters is made more difficult by the possibility of correlation in unobservables, which Manski terms correlated effects. In the model above, this is reflected in the potential for $\mathbb{E}(\epsilon_i\epsilon_j|\mathbf{x}, W) \neq 0$ for $i \neq j$. (Notice that the covariance structure for the error vector is left unrestricted above.) As illustrated in that paper, if y represents school achievement, endogenous effects arise if one's achievement tends to vary with the average achievement in that person's reference group. If achievement is affected by the reference group's socio-economic background, there is an exogenous or contextual effect. Correlated effects may arise because pupils are exposed to the same teacher or have similar features that are relevant for achievement but are not observed by the researcher. Because all of these may explain similarities in outcomes, it gives rise to what Manski calls the "reflection problem."

The reduced form in (3) arises rather naturally in interaction models for maximizing agents endowed with quadratic payoffs. Blume, Brock, Durlauf, and Jayaraman (2015), for instance, suggest that observed outcomes can be construed as Bayes-Nash equilibria of a game with incomplete information. For a model of strategic complementarities in production, payoffs are given by

$$U_i(\mathbf{y}; W) = \left(\alpha + \eta x_i + \gamma \sum_{j \neq i} W_{ij} x_j + z_i \right) y_i + \beta \sum_{j \neq i} W_{ij} y_i y_j - \frac{1}{2} y_i^2, \quad (4)$$

where z_i is private information to individual i . The first two terms reflect a production function mapping effort into an outcome of interest where the second term reflects complementarity among individuals. The reduced form in (3) then corresponds to the equilibrium

profile of the game where the unobservable error is a function of the private information z_i .⁶

Calvó-Armengol, Patacchini, and Zenou (2009) study a slightly different model of educational achievement. There, a pupil’s educational achievement y_i is the sum of two effort choices, an idiosyncratic effort level e_i that is unaffected by peers’ choices, and a “peer effect” effort level ϵ_i , which is potentially complemented by other individuals’ efforts. The payoff function for a student is given by

$$U_i(e_i, \epsilon; W) = \left(\eta x_i + \gamma \sum_{j \neq i} W_{ij} x_j \right) e_i - \frac{1}{2} e_i^2 + (\alpha W_i \mathbf{1} + \nu_i) \epsilon_i - \frac{1}{2} \epsilon_i^2 + \tilde{\beta} \sum_{j=1}^N W_{ij} \epsilon_j,$$

where W_i is the i th row of W so that its multiplication by $\mathbf{1}$ produces the degree for individual i and ν_i is an idiosyncratic taste shock. The Nash equilibrium of this model leads to a reduced form econometric model given by⁷

$$\begin{aligned} y_i &= \eta x_i + \gamma \sum_{j=1}^N W_{ij} x_j + \epsilon_i \\ \epsilon_i &= \alpha W_i \mathbf{1} + \tilde{\beta} \sum_{j=1}^N W_{ij} \epsilon_j + \nu_i \end{aligned} \Rightarrow \mathbf{y} = \frac{\alpha}{\tilde{\beta}} (\mathbf{I} - \tilde{\beta} W)^{-1} \tilde{\beta} W \mathbf{1} + (\eta \mathbf{I} + \gamma W) \mathbf{x} + (\mathbf{I} - \tilde{\beta} W)^{-1} \boldsymbol{\nu}. \quad (5)$$

Notice that, since both e_i and ϵ_i are choice variables, endogenous effects are now encoded into $\tilde{\beta}$. It is also noteworthy that the simultaneity is in the determination of the unobservable error ϵ instead of the observable outcome variable \mathbf{y} . Finally, I should also point out that this special structure delivers a direct dependence of the outcome variable \mathbf{y} on the Katz-Bonacich centrality index for each individual at $\tilde{\beta}$, listed in the vector $(\mathbf{I} - \tilde{\beta} W)^{-1} \tilde{\beta} W \mathbf{1}$ if the “peer effect” effort depends on the degree of the individual (i.e., $\alpha \neq 0$).⁸ In their study of bank liquidity holdings in the United Kingdom, Denbee, Julliard, Li, and Yuan (2014) use a variation of this model and its connection with the Katz-Bonacich centrality to define a network impulse-response function of total outcome. In their analysis, W_{ij} is the (predetermined) borrowing by bank i from bank j . The variable e_i is interpreted as a bank’s

⁶Blume, Brock, Durlauf, and Jayaraman (2015) focus their analysis on a different payoff structure corresponding to a narrative where individuals have preference for conformity. As they point out, the models are observationally equivalent.

⁷To guarantee uniqueness and interiority of the equilibrium, the authors impose the restriction that $\tilde{\beta}$ is less than the reciprocal of the largest eigenvalue of W . Whereas the estimates for the vast majority of the networks analyzed in the study satisfy this condition, 9% (=18/199) of the networks do not. It would be interesting to extend the analysis to incorporate the possibility of multiplicity and/or corner equilibria.

⁸If the rows of W add up to one, the intercept in the reduced form (3) is also given by a multiple of the Katz-Bonacich centrality index. Cohen-Cole, Kirilenko, and Patacchini (2014) estimate the model (2) on trading networks in financial futures markets and analyze the centrality indices for the traders in their sample.

liquidity holdings when it is isolated, and the variable ϵ_i gives the liquidity holdings in a banking network.

In spite of the apparent differences between the preceding model and the model represented by (2)-(3), much of the identification and estimation analysis of this model follows along the same lines, and I will focus on the model given by (2)-(3) (unless explicitly stated).⁹

Earlier analyses of the model (2) focused on a peer structure given by a complete network where $\sum_{j=1}^N W_{ij} = 1$, $W_{ij} = W_{ik}$ for $j \neq k$ and it is customary to assume that $W_{ii} = 0$ (e.g., Moffitt (2001)).¹⁰ It is also commonplace to suppose that $|\beta| < 1$, which together with row-sum normalization guarantees that $\mathbf{I} - \beta W$ is invertible and a well-defined reduced form exists. Under this specification and no further restrictions, it can be formally demonstrated that the structure represented by $(\alpha, \beta, \eta, \gamma)$ is not point-identified:

Proposition 1. *If $|\beta| < 1$, $\eta\beta + \gamma \neq 0$, $W_{ij} = (N - 1)^{-1}$ if $i \neq j$ and $W_{ii} = 0$, $(\alpha, \beta, \eta, \gamma)$ is not point-identified.*

This result is originally indicated in Manski (1993) and demonstrated, for instance, as a corollary to Proposition 1 in Bramoullé, Djebbari, and Fortin (2009). This negative result is also examined, for example, by Kelejian, Prucha, and Yuzevovich (2006) in an estimation context. The outlook on identification improves if one imposes further restrictions on the model and/or the available data. To illustrate this, I focus on the related representation originally considered in Manski (1993). Instead of specification (1), Manski studies a model akin to

$$y_i = \alpha + \beta \mathbb{E}(y_j|w) + \eta x_i + \gamma \mathbb{E}(x_j|w) + \epsilon_i, \quad \mathbb{E}(\epsilon_i|\mathbf{x}, w) = \delta w,$$

where w stands for a (scalar) identifier of the group, and expected values are taken so that the model is equilibrated and corresponds to a “social equilibrium.”¹¹ The coefficients above

⁹Calvó-Armengol, Patacchini, and Zenou (2009) use a variation of Proposition 4 below. If there are no correlated effects (i.e., correlation in ν_i), it might be possible to establish identification, as in Proposition 2 below.

¹⁰This weighting scheme amounts to collecting the endogenous and contextual covariates as the mean of one’s peer group (sometimes termed the “exclusive mean”). As pointed out in Guryan, Kory, and Notowidigdo (2009), this mechanically generates a correlation between x_i and $\sum_{j \neq i} x_j / (N - 1)$ within the group even when x_i and x_j are independent for all pairs, and a regression of the first on the second will lead to a biased estimator. This can be seen by noting that the usual requirement for OLS unbiasedness (strict exogeneity) is not satisfied, even though the covariance $\mathbb{C}(x_i, \sum_{j \neq i} x_j / (N - 1)) = 0$ and the Best Linear Projection slope coefficient is zero. This happens since x_i shows up either as regressand or as regressor (as part of the “exclusive mean”) in all observations within the group. Whereas the regression will produce a biased estimator, the OLS estimators are still consistent (since contemporaneous exogeneity is preserved) and the problem is attenuated for larger groups as pointed out by those authors.

¹¹In the literature, this structure is sometimes emulated by assuming that $\sum_{j=1}^N W_{ij} y_j$ is an “inclusive

retain the same interpretation as before, and $\delta \neq 0$ when there are correlated effects. Using this model, for example, a corollary to Proposition 2 in Manski is that, when $\delta = \gamma = 0$, the remaining parameters (α, β and η) are point identified if $1, \mathbb{E}(x_j|w)$ and x_i are “linearly independent in the population.” Allowing for $\gamma \neq 0$, but (although not explicitly stated) otherwise under the same conditions (i.e., $\delta = 0$ and the variation previously implied by the linear independence condition), Angrist (2014) (see also Acemoglu and Angrist (2001) and Boozer and Cacciola (2001)) demonstrates an analogous result that β is point identified, drawing an interesting connection of this parameter to the population counterparts of the regression coefficient of y_i on x_i and a regression of group averages of y_i on group averages of x_i , which can be interpreted as the 2SLS estimator using group dummies as instruments for x_i . As pointed out by Manski (1993) and using my notation, “the ability to infer the presence of social effects depends critically on the manner in which x varies with w ” (p.535). The non-identification result in Proposition 1 does not use this variation, whereas these positive identification results explore the between-group variation of the regressor x , without which the linear independence condition stated above fails and the variance $\mathbb{V}(\mathbb{E}(x_j|w)) = 0$, jeopardizing the results.

Alternative restrictions on the model (2) also allow us to achieve identification using higher moments. If there are no correlated effects, for example, and the conditional variance $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I}$, we have

$$\mathbb{V}(\mathbf{y}|\mathbf{x}) = \sigma^2(\mathbf{I} - \beta W)^{-2}.$$

This is enough to identify β and, consequently, the remaining parameters, even when $W_{ij} = (N - 1)^{-1}$ if $i \neq j$ and $W_{ii} = 0$. In the peer effects literature, this result is indicated in Moffitt (2001) but is actually reminiscent of earlier results on covariance restrictions and identification of simultaneous equation models (see Fisher (1966), Bekker and Pollock (1986), Hausman, Newey, and Taylor (1987)). Below, I state it for the general case of N individuals, and a direct demonstration is available in the appendix.¹²

Proposition 2. *If $|\beta| < 1$, $W_{ij} = (N - 1)^{-1}$ if $i \neq j$, $W_{ii} = 0$, and $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I}$ then $(\alpha, \beta, \eta, \gamma)$ is point-identified.*

mean,” where $W_{ii} \neq 0$ and $W_{ij} = N^{-1}$ for any i and j . This “expectational” equation can also be explained from more foundational models of behavior, as illustrated, for instance, in Blume, Brock, Durlauf, and Jayaraman (2015). As indicated there, whereas the appearance of $\mathbb{E}(y_j|w)$ or $\sum_{j=1}^N W_{ij}y_j$ may reflect different informational assumptions on the microeconomic model under consideration, they are in fact econometrically isomorphic.

¹²Covariance restrictions alone are not enough to identify the model without additional coefficient restrictions. The coefficient restrictions in the present model are different from those dealt with in the earlier works, which appear to focus on exclusion restrictions across equations.

Interestingly, the covariance restrictions above also imply a lower bound on the correlation among observable outcomes, which is strictly greater than the lower bound for the pairwise correlation of a collection of equi-correlated random variables when $N \geq 3$. The reasoning for this is as follows: If a person i 's outcome is increased and β is negative, this has a downward direct influence on a given peer j . If a third individual k is also in the group, that person's outcome will also be negatively affected by the increase in i 's outcome. This negative influence in k will, on the other hand, put upward pressure on j 's outcome and the effect of the original increase in i 's outcome will tend to be attenuated. (Of course, this indirect effect is not present if $N = 2$ and, accordingly, the lower bound there is exactly -1 . The bound is non-trivial when $N > 2$.)¹³ Although the restrictions contemplated here are strong (no correlated effects and equal variance across individuals), Proposition 2 suggests that covariance restrictions may not only be useful in identifying the parameters of interest, but also in providing testable implications. This result is summarised below.

Proposition 3. *If $|\beta| < 1$, $W_{ij} = (N - 1)^{-1}$ if $i \neq j$, $W_{ii} = 0$, and $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I}$ then*

$$\frac{\mathbb{C}(y_i, y_j|\mathbf{x})}{\mathbb{V}(y_i|\mathbf{x})} > \frac{4 - 3N}{4N^2 - 11N + 8}.$$

Since the presence of an additive common shock will tend to increase the correlation between two observable variables, I conjecture that a similar lower bound on correlations as in Proposition 3 is possible in that case. When correlated effects manifest themselves through an additive group effect (i.e, for a group $l = 1, \dots, L$, the intercept is a random, possibly covariate dependent α_l), Davezies, d'Haultfoeuille, and Fougère (2009) show that the covariance restriction $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I}$ still provides identification if there are *at least* two groups of different sizes (see their Proposition 3.2). Recently, Rose (2015) examines identifiability using second moments under the (weaker) assumption that $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I} + \sigma_{\epsilon\epsilon}(W + W^\top)$. There, identification is established under conditions on W that are reminiscent of (though stronger than) the linear independence assumptions in Bramoullé, Djebbari, and Fortin (2009) (see below).

In fact, the use of restrictions on unobservables and higher moments for identification has been explored elsewhere in the literature for the identification and estimation of variations of the peer effects model presented in (2) (see Glaeser, Sacerdote, and Scheinkman (1996)

¹³When $N = 3$, for example, the correlation implied by Proposition 3 is $-0.\overline{45}$. For three equi-correlated random variables, positive definiteness of the variance-covariance matrix implies a smaller lower bound of -0.50 .

for an early example). Graham (2008), for instance, studies identification when outcomes within a group $l = 1, \dots, L$ are defined by

$$\mathbf{y}_l \mathbf{1}_{N_l \times 1} = \tilde{\gamma} W_l \mathbf{1}_{N_l \times N_l} \epsilon_l \mathbf{1}_{N_l \times 1} + \alpha_l \mathbf{1}_{N_l \times 1} + \epsilon_l \mathbf{1}_{N_l \times 1},$$

where N_l is the number of individuals in group l , $W_{ij,l} = (N_l - 1)^{-1}$ if $i \neq j$ and $W_{ii,l} = 0$, and the group-specific intercept α_l is allowed to vary across groups.¹⁴ (A similar model is also contemplated in Glaeser, Sacerdote, and Scheinkman (2003).) The unobservables are separated into three components: an individual idiosyncratic component $\epsilon_{i,l}$, the average of that variable among a person's peers $\sum_{j \neq i} \epsilon_{j,l} / (N_l - 1)$, and a group-specific shock α_l . The main identification target is $\tilde{\gamma}$, which is interpreted as a contextual effect parameter on (unobservable) group characteristics.¹⁵ Since the contextual effects here are unobserved, the difficulty lies in separating this unobservable component from the group-wide error α_l , which stand for the usual correlated effects.

Graham (2008) shows that $\tilde{\gamma}$ is identified if (i) two groups are available ($L \geq 2$), (ii) there is random assignment ($\Rightarrow \mathbb{E}(\epsilon_{i,l} \epsilon_{j,l}) = \mathbb{E}(\alpha_l \epsilon_{j,l}) = 0$ for any i, j, l ; i.e., there is no sorting or matching in group formation); (iii) the variance of α_l does not differ across groups; and (iv) there is difference in within-group variance of outcomes. Recently, Blume, Brock, Durlauf, and Jayaraman (2015) demonstrate how this result can be generalized to allow for identification of the model in (2) through higher moments (when there are no correlated effects) (see their Theorem 5).

An interesting avenue for identification appears when the (observed) social network graph is not complete in a way that introduces enough exclusion restrictions into the equation system (2) to reestablish the (necessary and sufficient) rank condition for point-identification. This insight is formalized in Bramoullé, Djebbari, and Fortin (2009):

Proposition 4 (Bramoullé, Djebbari, and Fortin (2009)). *If $\eta\beta + \gamma \neq 0$ and \mathbf{I}, W, W^2 are linearly independent, $(\alpha, \beta, \eta, \gamma)$ is point-identified.*

If $W_{ij} = (N - 1)^{-1}$ if $i \neq j$ and $W_{ii} = 0$, $W^2 = (N - 1)^{-1} \mathbf{I} + (N - 2) / (N - 1) W$ and the linear independence condition fails. One way in which that condition is satisfied is if W is block-diagonal with at least two blocks of different order. Suppose, for example, that the social network is comprised of two complete subgraphs of size N_1 and N_2 such that

¹⁴Exogenous covariates \mathbf{x}_l and (observed) contextual effects on those can be accommodated (see Blume, Brock, Durlauf, and Jayaraman (2015)), but are omitted as in Graham (2008).

¹⁵Alternatively, it can be seen as an amalgam of contextual and endogenous effect parameters.

$N_1 + N_2 = N$. In this case, $W^2 = \lambda_0 \mathbf{I} + \lambda_1 W$. Whereas direct computation shows that W^2 has N_1 diagonal elements equal to $(N_1 - 1)^{-1}$ and N_2 diagonal elements equal to $(N_2 - 1)^{-1}$, the diagonal elements of $\lambda_0 \mathbf{I} + \lambda_1 W$ all equal λ_0 . This produces $\lambda_0 = (N_1 - 1)^{-1} = (N_2 - 1)^{-1}$ if we focus on diagonal elements of W^2 . Another way to see how identification comes about in this case is to notice that the reduced form equation for individual i in group $l = 1, 2$ becomes

$$y_i = \frac{\alpha}{1 - \beta} + \left[\eta + \frac{\beta(\eta\beta + \gamma)}{(1 - \beta)(N_l - 1 + \beta)} \right] x_i + \frac{\eta\beta + \gamma}{(1 - \beta)(1 + \frac{\beta}{N_l - 1})} \bar{x}_i + \nu_i,$$

where \bar{x}_i is the average covariate in group l , and ν_i is the corresponding reduced-form error. The variation of the reduced-form coefficients across groups of different sizes then allows one to identify the parameters of interest. Hence, if $N_1 \neq N_2$, the linear independence condition is satisfied and the model is identified. The use of groups with different sizes to obtain identification is also employed in Lee (2007) and Davezies, d’Haultfoeuille, and Fougère (2009).¹⁶

Identification is also made possible if there are vertices whose peers are linked to nodes that are not themselves directly connected to the original nodes. This allows one to use indirect peers to generate instrumental variables for the endogenous outcomes in the right-hand side of equation (2). These are naturally encoded in the requirement that \mathbf{I} , W , and W^2 be linearly independent. Suppose, for instance, that nodes $1, \dots, N$ are placed on a circle, and links are directed from i to $i + 1$ (and N to 1) as represented in Figure 1.

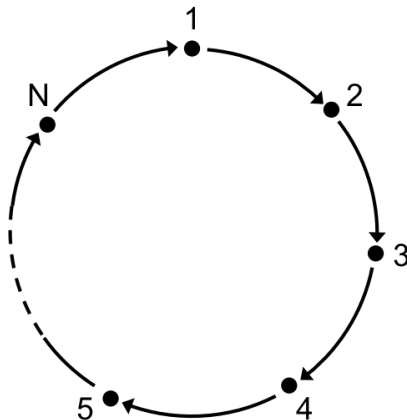
In this case, $W_{i,i+1} = W_{N,1} = 1, i = 1, \dots, N - 1$, and all other entries are zero. Each node i is affected directly by $i - 1$ and indirectly by every other vertex. The matrix W^2 is such that $(W^2)_{N-1,1} = (W^2)_{N,2} = (W^2)_{i,i+2} = 1$ and records nodes that can be reached through a direct link. Since none of those is directly connected to the original vertices, they provide leverage for the model to identify the structure. This insight is used empirically in De Giorgi, Pellizari, and Redaelli (2010) and Bramoullé, Djebbari, and Fortin (2009). The model above is essentially a spatial autoregressive model and estimation can be pursued using spatial statistics methods (e.g., Kelejian and Prucha (2010), Lin and Lee (2010), Lee and Yu (2010), or Lee, Liu, and Lin (2010)).¹⁷

A maintained assumption in the above identification result is that covariates are econo-

¹⁶Boucher, Bramoullé, Djebbari, and Fortin (2014) apply the estimator proposed in Lee (2007) to a study of peer effects in educational achievement.

¹⁷The command `spreg` in Stata, for example, implements maximum likelihood and GS2SLS estimators (see Drukker, Egger, and Prucha (2013)).

Figure 1: Directed Circle Network



metrically exogenous, which may be problematic as well. Bramoullé, Djebbari, and Fortin (2009) also extend the theoretical analysis above to the possibility of network-specific “fixed effects,” α_l , potentially correlated with the covariates.¹⁸ Of course, if excluded instruments are available, one may also employ them to identify the relevant parameters. It should be noted that the econometric exogeneity of the network structure itself is usually a maintained hypothesis. (Qu and Lee (2015) provide an instrumental variable estimator that allows for endogeneity of W under certain conditions.)

Recently, Blume, Brock, Durlauf, and Jayaraman (2015) pointed out that, in fact, the identifying assumption on \mathbf{I} , W , and W^2 is the norm, rather than the exception! Indeed, when W is such that $\sum_{j=1}^N W_{ij} = 1$ and $W_{ii} = 0$ for any $i \in \{1, \dots, N\}$, the condition essentially only fails for the case when the social network is made up of equally-sized components with equal nonzero entries.¹⁹ This is indicated in Theorem A2 in that paper:

Proposition 5 (Blume, Brock, Durlauf, and Jayaraman (2015)). *If W is such that $\sum_{j=1}^N W_{ij} = 1$ and $W_{ii} = 0$ for any $i \in \{1, \dots, N\}$ and \mathbf{I} , W , and W^2 are linearly dependent, then W is block diagonal with blocks of the same size, say $N_l (\leq N)$, and any nonzero entry is given by $(N_l - 1)^{-1}$.*

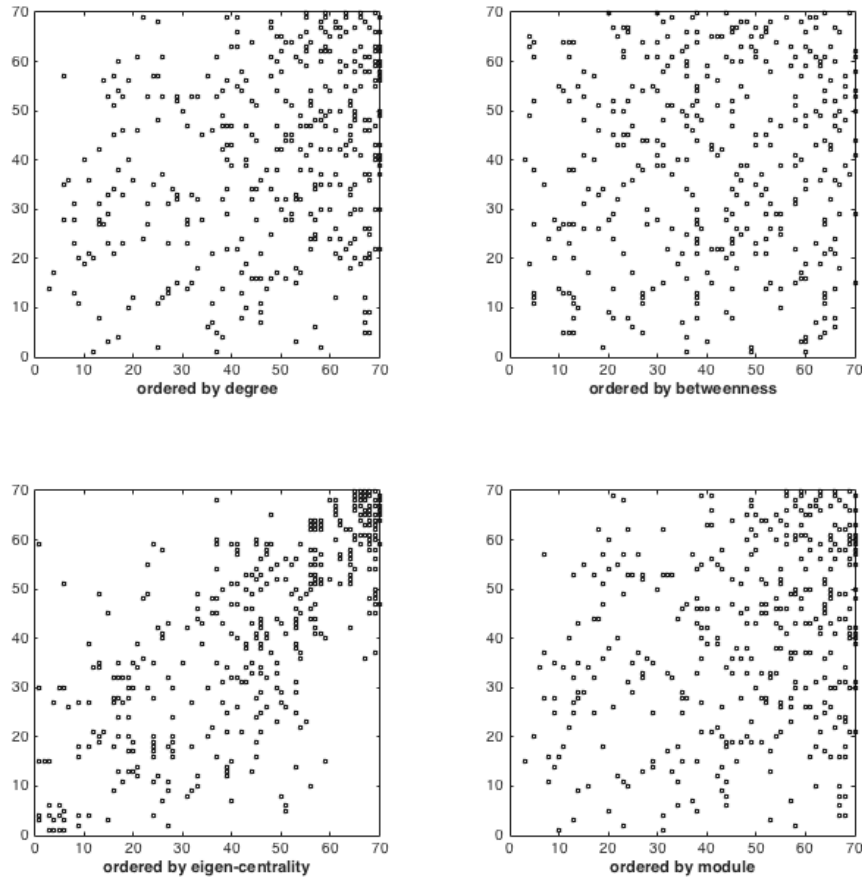
In fact, many social graphs appear to display a connection structure that greatly departs

¹⁸An uncorrelated “random effect” α_l would still allow for identification under the conditions above. Bramoullé, Djebbari, and Fortin (2009) difference away the “fixed effect,” which requires the linear independence condition to be strengthened to the linear independence of \mathbf{I} , W , W^2 , and W^3 .

¹⁹The row-sum normalization corresponds to what Liu, Patacchini, and Zenou (2014) call *local average* interactions, as opposed to, say, *local aggregate* interactions, where entries in W would correspond to 0 or 1.

from the block diagonal and complete within each block. Figure (2), a pictorial representation for the adjacency matrix of a network of friendships among teenagers, displays such an instance, with patterns that are much distinct from those obtained from a partition of individuals into completely connected subnetworks leading block diagonality in the subplots from Figure (2).

Figure 2: Adjacency Matrix: High School Friendships



Note: In 1957 and 1958, boys in a small high school in Illinois were asked the following: “What fellows here in school do you go around with most often?” The data aggregates information from both years and appears in Coleman (1964). The panels display nodes ordered by degree centrality, betweenness centrality, eigenvector centrality, and modularity. The latter is a measure used to detect community structure in a graph.

The setup described up to this point presumes that researchers observe the social structure represented by W . Whereas connections are sometimes elicited in survey instruments (e.g., the National Longitudinal Study of Adolescent to Adult Health, known as AddHealth),

“[i]f researchers do not know how individuals form reference groups and perceive reference-group outcomes, then it is reasonable to ask whether observed behavior can be used to infer these unknowns” (Manski (1993), p.536). Although this is not possible with a complete social graph, because observed outcomes are informative about the underlying social structure acting as conduit, one may still hope to retain identification under plausible additional restrictions. Blume, Brock, Durlauf, and Jayaraman (2015), for example, demonstrate that partial knowledge of W can be used to identify the relevant parameters. In particular, they focus on a variation of the setup above, where the social structures mediating endogenous and contextual effects may differ (i.e., $y_i = \alpha + \beta \sum_{j=1}^N W_{i,j,y} y_j + \eta x_i + \gamma \sum_{j=1}^N W_{i,j,x} x_j + \epsilon_i$ with possibly distinct W_x and W_y). They show that when W^x is known and two (known) nodes are also known to not be connected, the parameters of the model can be identified (Theorem 6 in that paper). An analogous result demonstrates that when there are enough unconnected nodes for each of the graphs represented by W_x and W_y , and the identity of those nodes is known, identification is also (generically) possible (Theorem 7 in that paper).

Observed outcomes can possibly offer further possibilities when W is not directly observed. As Manski (1993) suggests, “[i]f researchers do not know how individuals form reference groups and perceive reference-group outcomes, then it is reasonable to ask whether observed behavior can be used to infer these unknowns” (p.536). Such possibilities are investigated in de Paula, Rasul, and Souza (2015).²⁰ Letting Π denote the matrix of reduced-form coefficients in the system (3), one has

$$\Pi = (\mathbf{I} - \beta W)^{-1}(\eta \mathbf{I} + \gamma W).$$

Under the assumption that $|\beta| < 1$, one obtains that W and Π in fact share eigenvectors and Π 's eigenvalues are functions of W 's eigenvalues and the parameters of interest. Among other things, this indicates that the eigenvector centrality of W —corresponding to the dominant eigenvector—can also be directly obtained from Π . Additional restrictions often employed in the literature, such as $\eta\beta + \gamma \neq 0$, linear independence of \mathbf{I}, W and W^2 and row-sum normalization (which implies that the largest eigenvalue for W and W^2 is one) allow one to provide a tight characterisation for the set of observationally equivalent parameters. This and other similar results are demonstrated in detail by de Paula, Rasul, and Souza (2015).

If Π can be estimated, an estimator for (at one element in the set of identified) parameters

²⁰Blume, Brock, Durlauf, and Ioannides (2011) show local identification when there is a partial order on individuals and W displays weights decaying exponentially in distance. Souza (2014) suggests a probabilistic model for W and an integrated likelihood method for the estimation of the (identified set of) parameters in model (2).

of interest can be obtained (say, via indirect least squares). Since the number of parameters (reduced or structural) is $O(N^2)$ though, to estimate those one would in practice need obtain at least as many observations of (\mathbf{y}, \mathbf{x}) for a given social system (i.e., $TN > N^2$). Whereas this is empirically conceivable when N is small, it is less plausible for even moderately sized networks.²¹ Estimation can nonetheless be possible with further, empirically credible restrictions on the system. Many social and economic networks (though not all) tend to be sparse, for instance. The density of the production networks examined in Atalay, Hortacsu, Roberts, and Syverson (2011) for the United States, for example, amounts to less than 1% of possible links.²² Also relying on United States data, Carvalho (2014) finds an edge density of about 3%.²³ If one defines an undirected network from reciprocal friendship nominations in the AddHealth dataset, which elicits teenage friendships, the density is about 2%.

This, potentially coupled with additional restrictions, opens the possibility of application of penalization methods well-suited to handle sparse models, like the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani (1996), see Belloni, Chernozhukov, and Hansen (2013) for a recent review focused on econometric applications), the Smoothly Clipped Absolute Deviation (SCAD) penalty (Fan and Li (2001)), the Elastic Net (Zou and Hastie (2005)), or the Minimax Concave Penalty (MCP) (Zhang (2010)). If T is the number of observed instances of (\mathbf{y}, \mathbf{x}) , applying those methods directly to the reduced form would entail an estimator defined as

$$\hat{\pi}_i = \operatorname{argmin}_{\pi_i} \frac{1}{T} \sum_t (y_{it} - \pi_i^\top \mathbf{x}_t)^2 + \lambda \sum_j p_T(\pi_{ij}),$$

for each $i \in \{1, \dots, N\}$, where π_i is a column-vector corresponding to the i th row from Π , and $p_T(\cdot)$ is a sample-size-related penalty function that depends on the particular penalization method used. (The adequate notion of sparsity here is that T be sufficiently large compared to nonzero entries in the adjacency matrix.) This reduced form estimator (using the Elastic Net penalty function) is applied by Bonaldi, Hortacsu, and Kastl (2014), for example, to

²¹There are, nevertheless, data environments, like financial systems, where information on outcomes is collected frequently and reduced-form parameters can potentially be estimated without additional restrictions.

²²Atalay, Hortacsu, Roberts, and Syverson (2011) used data from Compustat from 1979 to 2007 to study supply networks. The average number of suppliers (indegrees) reported in the study was 3.67 over the sample period. The number of firms in the sample varied between 631 (in 1979) and 1848 (in 2002). The total number of possible links in the directed graph is $N(N - 1)$, and the density of links, defined as the ratio of observed links to potential links, is then given by the average (in)degree divided by $N - 1$. Assuming that the average indegree is constant across years, the density is then between 0.2% (for $N = 1848$) and 0.6% (for $N = 631$).

²³Carvalho (2014) uses input-output tables from the Bureau of Economic Analysis in 2002, defining 417 sectors as nodes in the network.

study the evolution and interconnection of banks' cost-of-funding inferred from bank bids in the main refinancing operation (MRO) auctions by the European Central Bank. There, y_{it} gives bank i 's cost of funding, and covariates \mathbf{x}_t are *lagged* cost-of-funding measures for all banks in the system. The authors use the estimated parameters to construct centrality indices for the banks in their sample. This estimator is also pursued by Manresa (2013) in a version of the model in (2) without endogenous effects ($\beta = 0$), in which case $\Pi = \eta\mathbf{I} + \gamma W$ (using my notation), allowing on the other hand for time- and individual-fixed effects.²⁴

The estimation strategy above relies on sparsity of the reduced-form coefficients. Since $\Pi = \eta\mathbf{I} + \gamma W$ when $\beta = 0$ (as in Manresa (2013)), row-sum normalization of W (as required previously) is unnecessary for identification (one can normalize $\gamma = 1$, and the entries in each row can be heterogeneous). In this case, given that Π is a linear function of η and W , sparsity of W is clearly transferred to sparsity in Π . Nevertheless, because

$$\Pi = (\mathbf{I} - \beta W)^{-1} (\eta\mathbf{I} + \gamma W) \Leftrightarrow W = (\Pi - \eta\mathbf{I}) (\beta\Pi + \gamma\mathbf{I})^{-1},$$

it is not immediate that sparsity of W translates into sparsity of the reduced-form coefficient matrix Π . This will be the case when β is small (in which case $\Pi \approx \eta\mathbf{I} + \gamma W$ and sparsity in \mathbf{I} and W carries over). Take, for example, the directed circle analyzed earlier, where $W_{i,i+1} = W_{N,1} = 1, i = 1, \dots, N - 1$, and all other entries are zero. Since there are N links out of possibly $N(N - 1)$ directed connections, the density of edges is given by $1/(N - 1)$. A directed circle with 100 nodes hence has an edge density of (approximately) 1%. Assume then that $\gamma = \eta = 1$ for simplicity. Figure 3 plots the proportion of zeros in Π as a function of β for a directed circle with $N = 100$.

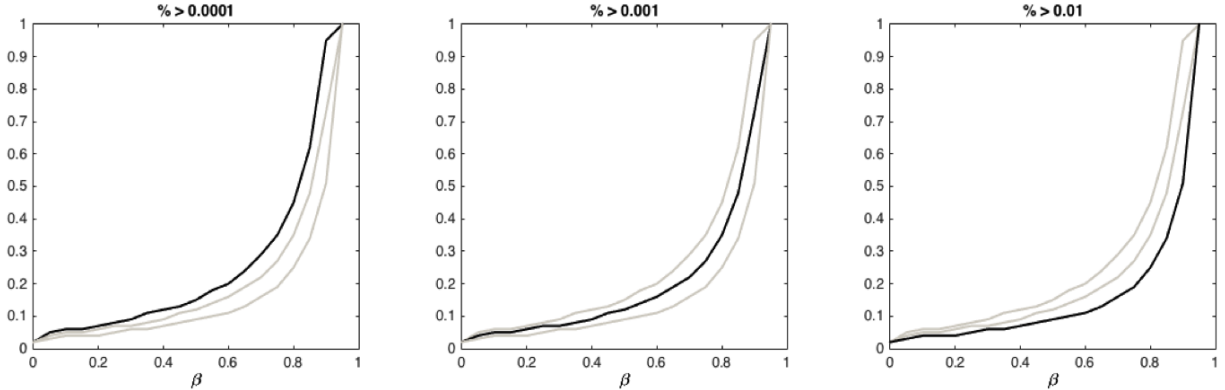
Alternatively, note that (for $|\beta| < 1$) we can expand the inverse using a Neumann series and obtain

$$\Pi = \eta\mathbf{I} + (\beta\eta + \gamma) \sum_{k=1}^{\infty} \beta^{k-1} W^k.$$

Since $\beta\eta + \gamma \neq 0$, $\pi_{ij} = 0$ if, and only if, there are no paths between i and j in W . The (i, j) entry in W^k is non-zero whenever there is a path of length k between i and j . If a pair is not connected at any length, that entry is zero for every k . Therefore, Π is sparse if there is a large number of (i, j) unconnected pairs in W . When that is the case, de Paula, Rasul, and

²⁴Manresa suggests a computational procedure that alternates between the individual LASSO estimator for individual specific parameters (individual fixed effects and individual row in W), penalizing the L^1 norm of W_i , given the remaining parameters and a pooled ordinary least squares estimator for the remaining parameters given individual specific parameters until convergence.

Figure 3: Sparsity of Π for Directed Circle



Note: The dark lines in the first, second, and third panels show the proportion of entries in $(\mathbf{I} - \beta W)^{-1}(\mathbf{I} + W)$ that are larger than 0.0001, 0.001 and 0.01, respectively, as a function of β . The matrix W is a 100×100 matrix ($N = 100$), such that $W_{ii+1} = W_{100,1} = 1$ for $i = 1, \dots, 100$ and zero, otherwise. The density (proportion nonzero entries) of W is 1%, and the density of $\mathbf{I} + W$, corresponding to $\beta = 0$, is 2%.

Souza (2015) say that W is “sparsely connected.” (Note that the circular graph used above is sparse, but not sparsely connected.) A sparsely connected network, where many pairs are not linked, will exhibit many components. A notion of *approximate* sparse connectedness can then be envisioned where the (i, j) entry of W^k is nonzero, but small. Because the number of components in a network equals the multiplicity of the zero eigenvalue of the Laplacian matrix (i.e., $\text{diag}(W\mathbf{1}) - W$) (see Kolaczyk (2009)), the spectrum of this matrix can be used as a measure of sparse connectedness; the more eigenvalues there are in the vicinity of zero, the closer the matrix is to a sparsely connected matrix. The directed circle above is not approximately sparsely connected either.

I should note that there is a built-in tension between sparse connectedness and the linear independence assumption used for the identification of the model. For a (row-)stochastic matrix, for example, the n th smallest eigenvalue for the Laplacian matrix $\text{diag}(W\mathbf{1}) - W$ corresponds to one minus the n th largest eigenvalue of W .²⁵ Then, if the spectrum of the Laplacian matrix is close to zero, the spectrum of the matrix W is close to a constant vector (of ones) and \mathbf{I} , W , and W^2 are linearly dependent. An alternative is to introduce the penalization directly on W , which may be more naturally expected to be sparse. In this case, it should be pointed out that the identification condition that W be non-negative, row-sum normalized to one implies that the L_1 -norm of W equals N . Hence, within this class of models, penalty schemes used in the LASSO or the Elastic Net might find difficulties discriminant

²⁵This is because $(\text{diag}(W\mathbf{1}) - W)\mathbf{v} = (\mathbf{I} - W)\mathbf{v} = \lambda\mathbf{v}$ implies that $W\mathbf{v} = (1 - \lambda)\mathbf{v}$.

among adjacency matrices. Other penalization strategies using non-convex penalty functions that also allow for sparse estimates, like the Smoothly Clipped Absolute Deviation (SCAD) penalty (Fan and Li (2001)) or the Minimax Concave Penalty (Zhang (2010)), will not be constant within the class of unit row-normalized, non-negative matrices and might be employed. On the other hand, I should note that sparsity of W itself can potentially provide identification power *per se*, as it provokes a relatively large number of entries in W to be zero. In this sense, unrestricted estimation is a possibility using the LASSO or Elastic Net. (The relaxation of positivity and row-sum normalization could be regarded as allowing for individual heterogeneity in the magnitude and sign of β .)

The penalization of W is pursued, for example, by Lam and Souza (2013), after dispensing with the positivity and normalization assumption on W . They estimate the parameters of the model by minimizing an objective function written directly in terms of the structural system. In terms of our specific model and notation,

$$\min_{(W, \beta, \delta, \gamma)} \frac{1}{T} \sum_t \|\mathbf{y}_t - \alpha - \beta W \mathbf{y}_t - \eta \mathbf{x}_t - \gamma W \mathbf{x}_t\|_2^2 + \lambda \sum_{i \neq j} p_T(W_{ij}), \quad (6)$$

where $\|\cdot\|_2$ is the Euclidean norm and the penalty term depends on the L_1 norm of W . Since y is endogenous, it is expected that additional assumptions need to be imposed, and Lam and Souza (2013) suppose that the variance of the structural errors ϵ_{it} vanishes asymptotically (see their Assumption A2). (The asymptotics are in N and T .) Intuitively, this assumption can be seen as a version of the Proximity Theorem: “[I]f the variance-covariance matrix of the regressors is bounded away from singularity, the least-squares estimator approaches consistency either as the variance of the disturbance approaches zero or as the probability limits of the correlations between the disturbance and regressors approach zero” (see Theorem 3.9.1 in Fisher (1966)). The estimator, then, may be best suited for circumstances when error variances are relatively small, but less suited when the variance of the structural errors is comparatively large (see, e.g., Table 5 in Lam and Souza (2013) for Monte Carlo results when error variances are comparatively large).²⁶

In lieu of the above strategy, we can instead opt to minimize the following objective

²⁶If the errors are normally distributed with a diagonal variance-covariance matrix, a full-information maximum likelihood estimator based on the model 2 would involve a term on the (log)- L^2 norm of W (corresponding to the logarithm of the Jacobian term), and a penalized (quasi-)maximum likelihood estimator could be employed.

function:

$$\begin{aligned} \min_{(W, \beta, \delta, \gamma)} & \frac{1}{T} \sum_t \|\mathbf{y}_t - \Pi \mathbf{x}_t\|_2^2 + \lambda \sum_{i \neq j} p_T(W_{ij}) \\ \text{s.t.} & (\mathbf{I} - \beta W)\Pi - (\eta \mathbf{I} + \gamma W) = 0 \end{aligned}$$

Absent the penalization term, this can be seen to correspond to an indirect least squares estimator. This estimator nevertheless penalizes the adjacency matrix W directly (as opposed to the previous estimator focusing on the reduced form). A potential complication is that the objective function will no longer be convex since $\Pi = (\mathbf{I} - \beta W)^{-1}(\eta \mathbf{I} + \gamma W)$.²⁷ Of course, if the row-sum normalization condition is imposed also, one would also likely have to resort to a non-convex penalization scheme (like SCAD or MCP). The analysis and performance of this estimator are subjects of ongoing research.²⁸

Aside from its role in estimation, sparsity of W can also be useful in identification as it essentially imposes exclusion restrictions on the different structural equations. Supposing enough sparsity, for example, Rose (2015) obtains identification results under additional conditions on the reduced-form coefficient matrix Π . These conditions are rank restrictions on sub-matrices of Π (whose verification is nevertheless computationally demanding). Intuitively, given two observationally equivalent systems, sparsity guarantees the existence of pairs that are not connected in either. Since observationally equivalent systems are linked via the reduced-form coefficient matrix, this pair allows one to identify certain parameters in the model and, having identified those, one can then proceed to identify other aspects of the structure. (This is related to the ideas in Theorem 6 of Blume, Brock, Durlauf, and Jayaraman (2015) (see discussion above).)

3.2 Nonlinearities and Multiple Equilibria

One can enumerate various empirical circumstances where a linear model may not be ideal (see, e.g., Kline and Tamer (2011)). Nonlinearities can occur through two possible, *non mutually exclusive*, avenues: by nonlinearities in the “link” function through which the (possibly weighted) average of peer outcomes determine an individual’s outcome (i.e., $y_i = f\left(\sum_{j=1}^N W_{ij}y_j, x_i, \sum_{j=1}^N W_{ij}x_j, \epsilon_i\right)$) or through nonlinear aggregation of peer outcomes (e.g.,

²⁷Another possibility which may attenuate the computational issues is to minimize the above objective function with respect to both Π and W . (I thank Lars Nesheim for this suggestion.)

²⁸If instrumental variables are available for each of the endogenous variables, (penalized) IV estimators could also provide an estimation avenue (see Gautier and Tsybakov (2014), Lam and Souza (2014)).

$\min_{j:W_{ij}\neq 0} y_j$ instead of $\sum_{j=1}^N W_{ij}y_j$).

As noted by Manski (1993), even under nonlinearities, nonparametric versions of the social interactions system with global interactions analyzed earlier (i.e., $W_{ij} = 1/(N - 1), i \neq j$ and $W_{ii} = 0, i, j = 1, \dots, N$) are not identified. Brock and Durlauf (2001) show that this is not the case in a (parametric) binary choice model with social interactions (without correlated effects), and Brock and Durlauf (2007) extend the analysis to nonparametric binary choice models showing various point- and partial-identification results.²⁹ Blume, Brock, Durlauf, and Ioannides (2011) provide a comprehensive survey of social interactions in nonlinear models covering, for example, multiple discrete choice (Brock and Durlauf (2006)), duration models (Sirakaya (2006), de Paula (2009), Honoré and de Paula (2010)). Recently, Bramoullé, Kranton, and D’Amours (2014) investigated a class of game theoretic models on general networks and suggested a Tobit-type social interactions econometric model (on this, see Xu and Lee (forthcoming)).

Manski (1993) also points out that “social effects might be transmitted by distributional features other than the mean” (p.534). Tao and Lee (2014) consider, for example, peer effects defined by the minimum outcome among one’s peers. Also looking at a model where individuals are affected by nonlinear functionals of their peers’ outcome distribution, Tincani (2015) obtains testable implications and finds evidence of nonlinearities in education peer effects in Chile using a recent earthquake and its differential impact on students at different distances from its rupture. Whereas these models use a global interactions network structure, more general network structures could prove useful in many respects.

One interesting aspect in such nonlinear models is the possibility for multiple equilibria. In their study of social effects on fertility choices, for example, Manski and Mayshar (2003) explore a utility-maximizing model where nonlinear child allowance schemes may lead to multiple social equilibria. Whereas multiplicity can, at times, pose issues for identification, this is not necessarily always the case. Once again, Manski (1993), for instance, suggested that “[t]he prospects for identification may improve if [the model] is non-linear in a manner that generates multiple social equilibria” (p.539). Indeed, identification is facilitated in certain cases, as suggested by the analysis of a binary choice model (without correlated effects) in de Paula and Tang (2012) (see de Paula (2013) for a general survey on econometric analysis of interaction models with multiplicity).

Recently, Manski (2013) contemplated the analysis of potential outcomes in a social in-

²⁹Bisin, Moro, and Topa (2011) and Menzel (2015) investigate alternatives for estimation of these models (and others) for interaction systems with a large number of players.

teractions context. Letting $x_i = \{0, 1\}$ denote a particular (binary) treatment for individual i and \mathbf{x} be a vector collecting treatments for the group as a whole, the potential outcome for individual i when the group is faced with the treatment profile \mathbf{x} is denoted by $y_i(\mathbf{x})$. Manski enumerates a few circumstances where individual treatments may spill over to other network members. The epidemiology of infectious diseases provides a salient example, where vaccination of an individual presumably affects that person’s likelihood of infection, which is additionally influenced by the infection status of others in the community (there is a well-established epidemiology literature on network diffusion of infectious diseases). Using the conditional cash transfer program Progresa in Mexico and using consumption as an outcome, Angelucci and De Giorgi (2009) found evidence of spillovers operating mostly through insurance and credit relationships. Dieye, Djebbari, and Barrera-Osorio (2014) investigate the effects of a scholarship program in Colombia with possible spillovers. Using our notation, we can represent the setup envisioned by Manski as

$$y_i(\mathbf{x}) = f(W_i, \mathbf{y}_{-i}(\mathbf{x}), \mathbf{x}, \epsilon_i),$$

where $y_i(\mathbf{x})$ is the potential outcome for individual i when the network receives treatment \mathbf{x} .³⁰ Manski provides characterizations on the set of identified distributions for the potential outcomes $\mathbf{y}(\mathbf{x}), \mathbf{x} \in \{0, 1\}^N$ when the above system has a single solution. As he points out, a nonlinear system may lead to multiple solutions (or maybe no solution at all), complicating the characterization of the identified set of potential outcome distributions. These ideas are further investigated (theoretically and empirically) by Lazzati (2015), who further imposes monotonicity restrictions on the treatment effects.

3.3 Other Considerations

Spillovers mediated through network structures are, of course, present in a myriad of economic and social circumstances. Demand externalities arise naturally in goods and services whose values depend on the topology and volume of the network of consumers and producers involved. There is a enough literature in theoretical and applied industrial organization on the topic to fill a completely separate survey article.

³⁰Manski also mentions the possibility that the treatment vector may affect the reference group for each individual, in which case $W_i = W_i(\mathbf{x})$. In fact, Comola and Prina (2014), studying the introduction of a savings product in Nepal, found that insurance-motivated connections are likely to be rewired after the intervention.

More traditional industries are nevertheless also interconnected via input-output directed relationships (see Kranton and Minehart (2001) for a theoretical analysis of buyer-seller networks). Carvalho, Nirei, and Saito (2014), for instance, examined supply chain disruptions following the 2011 earthquake in Japan. Also using Japanese data, Bernard, Moxnes, and Saito (2014) studied buyer-seller networks and firm performance using infra-structure developments (high-speed trains) for variation in travelling costs, leading to the creation of new buyer-seller linkages. In fact, supplier-client networks present many of the same features described earlier (e.g., clustering, sparsity, intransitivities) (e.g., Atalay, Hortacsu, Roberts, and Syverson (2011) and Carvalho (2014)). It is *conceivable* that some of the ideas highlighted in previous subsections (e.g., the use indirect peers' outcomes or the output of suppliers of one's suppliers for identification) could be used in the identification of these relationships and primitives of interest, like production or value-added functions (e.g., Gandhi, Navarro, and Rivers (2013) for a recent article with a good overview on the econometric literature about production and value-added functions). Production networks and their composition have obvious connections to the macroeconomy (e.g., Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012)) and trade (e.g., Antras and Chor (2013)), and I refer the reader to Carvalho (2014) and Acemoglu, Ozdaglar, and Tahbaz-Salehi (forthcoming)) for recent reviews on the propagation of microeconomic shocks in production and financial systems.

Buyer-seller networks have repercussions in other dimensions as well. One interesting dimension is in taxation chains. The collection of value-added taxes is usually done via a credit mechanism, for example, whereby firms remit taxes on their revenues and claim tax credits on their inputs. Theoretically, this binds compliance through the network and is another channel through which individual firm decisions may reverberate through the network (e.g., de Paula and Scheinkman (2010) and Pomeranz (forthcoming)).³¹ Again, the insights described previously could prove useful here.

4 Network Formation

As seen above, the connection structure among networked agents can assist with the identification and estimation of models describing the resolution of economic and social outcomes of interest. It is nevertheless apparent that, whereas in some cases the peer structure can be taken as (econometrically) exogenous or predetermined, many times the very formation of a

³¹Network spillovers can also be present at the individual compliance level (e.g., Fortin, Lacroix, and Villeval (2007), Galbiati and Zanella (2012) and Kopczuk, Alstadsaeter, and Telle (2014)).

connection arises in response to incentives that may or may not articulate with the outcomes to be determined using the networks as a conduit. Models of network formation are, then, of interest on their own as well as in conjunction with the simultaneous equation models covered in the previous section.

4.1 Statistical Models

We can posit a data-generating process summarized by the statistical model $(\mathcal{G}, \sigma(\mathcal{G}), \mathcal{P})$, where $\sigma(\mathcal{G})$ is a σ -algebra of events in the sample space of graphs \mathcal{G} , and \mathcal{P} is a class of probability distributions on the measurable space $(\mathcal{G}, \sigma(\mathcal{G}))$. In the Erdős-Rényi model defined on N vertices, for example, \mathcal{P} would be the parametric class of models indexed by the probability $p \in (0, 1)$ that an undirected link is independently formed between any two vertices under consideration, defining a probability distribution over the set of $2^{N(N-1)/2}$ possible graphs on the set of N nodes. (I will focus here on undirected networks, though versions of many of the estimators below exist for more general graphical structures.) A great many statistical models for network formation can be seen as enrichments of this simple model (just as a probit or logit and mixture versions of those can be seen as generalizations of a Bernoulli statistical model). For the statistical models listed in this subsection, the analyst is assumed to have data on at least one network (*but not necessarily more than one*). (The estimation of a classical Erdős-Rényi model with only one network of N individuals, for instance, would essentially amount to the estimation of Bernoulli parameter on $N(N-1)/2$ observations.)

Zheng, Salganik, and Gelman (2006), for example, used a heterogeneous version of this simple random graph model to obtain estimates for the total size of hard-to-count populations. In the Erdős-Rényi model above, the expected degree for a given individual when there are N nodes equals Np , and the proportion of total links involving individuals in group k (e.g., incarcerated individuals) is given by N_k/N . The answer to the question of how many individuals of group k an individual i knows is then well-approximated by a Poisson distribution with parameter pN_k (similar to the approximation of the degree distribution by a Poisson distribution, as mentioned earlier). The authors show that a better model for the data analyzed is one where not only individual “gregariousness” (i.e., the expected degree of the individual) but also the individual propensity to know individuals in a given group k are heterogeneous. The distribution of answer counts is then given by a mixture of Poisson distributions (mixing over the heterogeneity in the relevant parameters), overdispersed rel-

ative to a (homogeneous parameter) Poisson distribution. Using this statistical model and Bayesian statistical methods, they were able to estimate the distribution of links (to any other individual) in the wider population of interest. This model is adapted by Hong and Xu (2014) to study local networks among fund managers using their portfolio allocation to local investments, where they find similar evidence of heterogeneity. Many other estimation strategies seek to characterize the degree distribution as a means to infer a well-suited probabilistic data-generating model for the network (e.g., $\ddot{\text{E}}$ rdos-Rényi, preferential attachment, etc.).

A well-known generalization of the $\ddot{\text{E}}$ rdos-Rényi model is the class of exponential random graph models introduced earlier in this article. For those models,

$$\mathbb{P}(G = g) = \exp \left(\sum_{k=1}^p \alpha_k S_k(g) - A(\alpha_1, \dots, \alpha_k) \right), \quad (7)$$

where $A(\alpha_1, \dots, \alpha_k)$ is a normalization constant ensuring that probabilities integrate to one and $S_k(g), k = 1, \dots, p$ enumerate certain features of the graph g . These would be characteristics like the number of edges, the number of triangles, and possibly many others. (If only the number of edges is considered, we obtain the $\ddot{\text{E}}$ rdos-Rényi model.) Since the probability mass function is in the exponential class, the model shares some well-understood properties. For example, the vector $(S_k(g))_{k=1}^p$ is a sufficient statistic for $(\alpha_k)_{k=1}^p$, which is called the natural parameter for the exponential family, and $A(\alpha_1, \dots, \alpha_k) = \ln \left[\sum_{g \in \mathcal{G}} \exp \left(\sum_{k=1}^p \alpha_k S_k(g) \right) \right]$ is its cumulant generating function or log partition function (i.e., the logarithm of its moment-generating function). (The distribution (7) is sometimes referred to as the Gibbs measure in the literature, given its connection to similar models in physics.) This model allows one to define probabilities using prevalent features of observed networks such as triads, clusters, and other topological characteristics.

In principle, it is also amenable to likelihood based inference just like any other exponential class distribution. The ensuing computational difficulties are nevertheless sizeable. This relates to the fact that the normalization constant (i.e., $\exp(-A(\alpha_1, \dots, \alpha_k))$) requires summing over all $2^{N(N-1)/2}$ graphs. If $N = 24$, the number of graphs (approximately 1.21×10^{83}) amounts to more than the estimated number of atoms in the observable universe. To circumvent this issue, considerable effort has been devoted to providing computational tools and approximation results that allow one to estimate this constant and carry out inference. Two main avenues are the use of variational principles and Markov chain Monte Carlo (MCMC)

methods, both rooted in statistical mechanics.

Exploring properties of the cumulant function (e.g., convexity), one can use variational methods to represent the constant as a solution to an optimization problem. Consider, for instance, an Erdős-Rényi graph on two nodes i and j . The (random) edge between these two vertices can be written as a Bernoulli random variable W_{ij} , and its probability mass function can be parameterized as $\mathbb{P}(W_{ij} = w_{ij}) = \exp(\alpha w_{ij}) / (1 + \exp(\alpha)) = \exp(\alpha w_{ij} - A(\alpha))$, $w_{ij} = 0, 1$. In this case, $A(\alpha) = \ln(1 + \exp(\alpha))$, and notice that $A''(\alpha) = \exp(\alpha) / (1 + \exp(\alpha))^2 > 0$, so $A(\alpha)$ is convex. Using results from convex analysis, one can express $A(\alpha)$ as

$$A(\alpha) = \sup_{\mu \in [0,1]} \{\alpha\mu - A^*(\mu)\}, \quad (8)$$

where $A^*(\mu) = \sup_{\alpha \in \mathbb{R}} \{\mu\alpha - A(\alpha)\} = \mu \ln \mu + (1 - \mu) \ln(1 - \mu)$ (i.e., the *convex conjugate* or Legendre-Fenchel transformation of $A(\alpha)$). Note also that $A^*(\mu)$ equals the negative of the solution to the problem: $\max_p H(p)$ subject to $\mathbb{E}_p(g_{ij}) = \mu$, where $H(p) \equiv -p \ln p - (1 - p) \ln(1 - p)$ is known as the Shannon entropy (for the Bernoulli distribution). The conjugate dual $A^*(\mu)$ can be obtained using the entropy measure more generally. Having obtained $A^*(\mu)$, if one then explicitly maximizes $\alpha\mu - A^*(\mu) = \alpha\mu - \mu \ln \mu - (1 - \mu) \ln(1 - \mu)$ over $\mu \in [0, 1]$, it can be seen that the maximum is attained at $A(\alpha)$ defined above. The key here is to use the fact that the necessary conjugate function is related to the entropy measure H and represents the sum over possible networks $A(\alpha)$ as the solution to an optimization problem. In practice, the computation of the dual function will involve calculation of the entropy measure H , and the optimization (8) is performed over a space that is not always easily characterized (in this case, it is $\{\mu : 0 \leq \mu \leq 1\}$). In high-dimensional problems, various approximations to the constraint set and the dual $A^*(\mu)$ are then pursued to provide a computationally tractable estimate of the cumulant function $A(\alpha)$ (see Section 3.3 in Jordan (2004) or Jordan and Wainwright (2008) and Braun and McAuliffe (2010) for an application to high-dimensional discrete choice models). Using variational methods, for example, Chatterjee and Diaconis (2013) provide an approximation of the cumulant function for dense graphs when the number of nodes goes to infinity³² and Chatterjee and Dembo (2014) provide error bounds for this approximation when there is a small degree of sparsity.

Perhaps a more familiar class of techniques involves MCMC methods. Such procedures have also been developed to produce maximum likelihood estimators and Bayesian posterior

³²Their arguments are based on approximations to large, dense graphs.

distributions for $(\alpha_1, \dots, \alpha_k)$ in exponential network models. Different approaches are summarized in Kolaczyk (2009).³³ It should be noted that various issues may arise in simulating ERGMs. The procedure can be very slow to converge to an invariant distribution. This is highlighted, for instance, in the discussion by Chandrasekhar and Jackson (2014) and Mele (2015) and formally demonstrated in Bhamidi, Bresler, and Sly (2011). In particular, for certain regions of the parameter space (defined as “low temperature” regions, in analogy to spin systems in physics), where the distribution (7) is multimodal, the mixing time for the MCMC procedure, i.e., the time it takes for the MCMC procedure to be within e^{-1} in total variation distance from the desired distribution, is exponential on the number of nodes (Theorem 6 in that paper). In other regions (“high temperature” ones), where the distribution (7) is unimodal, the mixing time is $O(n^2 \ln n)$ (Theorem 5 in that article).

One recurrent related issue in the application of ERGMs is what the literature terms *degeneracy* or *near degeneracy*, whereby “depending on the parameter values, the exponential random graph distribution can have a bimodal shape in the sense that most of the probability mass is distributed over two clearly separated subsets of the set of all digraphs, one subset containing only low-density and the other subset containing only high-density digraphs. The separation between these two subsets can be so extreme that (...) stochastic updating steps which change only a small number of arc variables (...), have a negligible probability of taking the Markov process from one to the other subset” (Snijders (2002), p.13). This again will lead at times to very slow convergence of a Markov chain Monte Carlo procedure to an invariant distribution. It is not uncommon either to observe abrupt changes on the class of probable graphs as parameters change, and all of these “oddities” are characterized as *degeneracy* or *near degeneracy* of the ERGM. In fact, this behavior is not at all unusual and is related to the general properties of discrete exponential distribution families as investigated in Rinaldo, Fienberg, and Zhou (2009) and Geyer (2009) (see Handcock (2003) for an earlier analysis). In such models, when the observed sufficient statistics are at (or, for all practical purposes, near) the boundary of their support, the MLE will not exist, and, even when it does, Markov chain-ML estimators will tend to not behave well. For instance, the sample average is the sufficient statistic for the natural parameter α of a Bernoulli random variable (as in the two-node Erdős-Rényi random graph above). When that sample average is one or zero, the ML estimator for the natural parameter does not exist! As indicated by Rinaldo, Fienberg, and Zhou (2009), “ERG modeling based on simple, low dimensional network statistics (...) can be rather coarse. In fact, those ERG models are invariant with respect to the relabeling of

³³The software suite `statnet` offers a package for the analysis of ERGMs in R.

the nodes and even to changes in the graph topologies, depending on the network statistics themselves. As a result, they do not specify distributions over graphs per se, but rather distributions over large classes of graphs having the same network statistics” (pp.459-460). Hence, when the model is not sufficiently rich and/or observed networks are moderately sparse, the sufficient statistics will act as a coarse classifier of networks, and one may well find herself facing the issues highlighted above. The stark “discontinuities” in the distribution of graphs generated by the model as parameters are varied are also investigated in Chatterjee and Diaconis (2013) for dense, large networks. They also discuss the (troublesome) issue that for certain regions of the parameter space—the “high temperature” ones, where the distribution (7) is unimodal, graph draws from the model are very close to those of an Erdős-Rényi model with independent link formation (see their Theorem 4.2). A similar point is made in Bhamidi, Bresler, and Sly (2011) (see their Theorem 7).³⁴

An early alternative estimation strategy, adapted to ERGMs by Strauss and Ikeda (1990) (but originally suggested by Besag (1975)), is to rely on the (log) pseudo-likelihood

$$\sum_{\{i,j\}} \ln \mathbb{P}(W_{ij} = 1 | W_{-ij} = w_{-ij}; \alpha),$$

where $\mathbb{P}(W_{ij} = 1 | W_{-ij} = w_{-ij}; \alpha)$ denotes the probability that the link W_{ij} is formed conditional on the remainder of the network, which I denote by W_{-ij} . Upon inspection, it can be seen that this does not correspond to the likelihood function for the model unless links form independently. In fact, if the dependence is not sufficiently weak, this estimator is bound to produce unreliable estimates (e.g., Robins, Snijders, Wang, Handcock, and Pattison (2007)).

If links are independently formed, the objective function is such that $\mathbb{P}(W_{ij} = 1 | W_{-ij} = w_{-ij}; \alpha) = \mathbb{P}(W_{ij} = 1; \alpha)$. In this case, one can easily focus on dyads (i.e., pairs of vertices and the existing links between them) and incorporate covariates. This dyadic model has often been used in the social sciences to study network links (see Wasserman and Faust (1994)). One well-known dyadic model is that of Holland and Leinhardt (1981), where the authors focus on directed links. Their model, which they call the p_1 model, postulates that:

$$\mathbb{P}(W_{ij} = W_{ji} = 1) \propto \exp(\alpha^{\text{rec}} + 2\alpha + \alpha_i^{\text{out}} + \alpha_i^{\text{in}} + \alpha_j^{\text{out}} + \alpha_j^{\text{in}})$$

³⁴These difficulties in distinguishing the model from an Erdős-Rényi one also lead to identification issues as pointed out in Mele (2015), which analyses an ERGM obtained from a network formation model (see discussion below). As he indicates, these identification challenges are less troublesome when multiple networks are observed (as in Nakajima (2007)), but are germane when identification relies on a single network.

and

$$\mathbb{P}(W_{ij} = 1, W_{ji} = 0) \propto \exp(\alpha + \alpha_i^{\text{out}} + \alpha_j^{\text{in}}).$$

Here, the parameter α_i^{out} encodes the tendency of node i to send out links irrespective of the target (its “gregariousness”), and α_j^{in} captures node j ’s tendency to receive links regardless of the sender’s identity (its “attractiveness”). The parameter α^{rec} registers the tendency for directed links to be reciprocated: large, positive values of α^{rec} will increase the likelihood of symmetric adjacency matrices. (Note that when $\alpha_i^{\text{out}} = \alpha_i^{\text{in}} = \alpha^{\text{rec}} = 0$ for all i , links form independently with probability given by $\exp(\alpha)/(1+\exp(\alpha))$, and the model would correspond to a logit.) It is straightforward to add dyad-specific covariates to the specification above.

Generalizations and special cases for this model have been suggested and extensively analyzed. Hoff (2005), for example, considers an augmented model where *multiplicative* interactions between individual unobserved factors are added to the probability specification above (i.e., $\mathbf{z}_i \times \mathbf{z}_j$, where \mathbf{z}_i is a vector of i -specific factors), and those plus the additive “gregariousness” and “attractiveness” features defined previously (i.e., α_i^{out} and α_i^{in}) are modeled as random effects. Building on tools from the (large- N, T) panel data literature (Fernandez-Val and Weidner (2014)) and focusing on the additive structure above with possibly additional observed covariates, Dzemski (2014) on the other hand treats α_i^{out} and α_i^{in} as fixed effects, hence allowing for an arbitrary correlation between those and with any observed characteristic in the model.³⁵ He provides a test of the model based on the prevalence of transitive triads (i.e., vertex triples where links are transitive) and an application to the microfinance-related networks collected and analyzed in Banerjee, Chandrasekhar, Duflo, and Jackson (2014) (among other papers).³⁶ Graham (2014) investigates a similar model (with observed covariates), but for undirected networks (see also Charbonneau (2014)). There, the undirected links are formed with a probability that is proportional to $\exp(\alpha + \alpha_i + \alpha_j)$. This is related to the Rasch model (Rasch (1960)) and can be seen as a pairwise stable arrangement when direct transfers are possible (see Bloch and Jackson (2007)). In the absence of covariates, its MLE large sample properties (for dense networks) are analyzed by Chatterjee, Diaconis, and Sly (2011) and Yan and Xu (2013), who call it the β -model. In this case, the distinction between “sender productivity” and “receiver attractiveness” for a given node disappears, but the parameters α_i can be interpreted as the proclivity by node

³⁵Hoff (2005) parameterized the correlation between α_i^{out} and α_i^{in} , whereas Dzemski (2014) can allow for more general dependencies.

³⁶Interestingly, the estimated distribution of “gregariousness” and “attractiveness” appear to cluster in a few groups, suggesting group-level heterogeneity.

i to establish connections. Those parameters are also treated as fixed effects. Aside from providing large sample characterizations for the ML estimator, Graham (2014) also analyzes the conditional ML estimator constructed using sufficient statistics for α_i , allowing him to “condition those parameters out” and circumvent the incidental parameters problem when estimating the observable covariate coefficients.

Instead of zooming in on individual pairs and the links established between them, Chandrasekhar and Jackson (2014) proposed a framework focusing on additional classes of subgraphs, which they call subgraph generation model (SUGM). The model specifies a set of K subgraphs, $(G_l)_{l=1}^K$, possibly involving a different number of nodes each and probabilities for each one of those. For example, one could specify $K = 2$ and G_1 to be the class of (undirected) edges between two nodes, taking probability p_1 , and G_2 to be the class of (undirected) triagles, to form with probability p_2 . (I focus on undirected graphs, but their analysis can also be extended to directed ones.) A possible narrative for this specification could be that the establishment of certain connections only requires dyads (e.g., a tennis match), whereas others elicit participation from triples (e.g., a proper rock’n’roll band). One version of their model has subgraphs form at random and produce a graph realization defined by the union of edges formed by the initial subgraph draws. Note that some edges drawn in the initial protocol will be redundant. The edge $\{i, j\}$ from the first class of subgraphs may form independently as well as part of a triangle, say $\{i, j, k\}$, with probability $p_1 \times p_t$: Two bandmates in a trio may also be tennis partners. (Chandrasekhar and Jackson (2014) also consider a protocol where subgraphs are formed sequentially, avoiding redundancies.) Furthermore, realized isolated edges, triagles, and more generally modeled subgraphs can possibly get “meshed” in the final observed network. For instance, a triangle involving nodes i, j and k could be the outcome of independently formed edges $\{i, j\}$, $\{j, k\}$, and $\{k, i\}$ (which occurs with probability p_1^3), a genuine triangle $\{i, j, k\}$ (happening with probability p_t), or a combination of independent edges and triangles involving those nodes (and possibly others). Disentangling the count of subgraphs in the model that are genuinely formed or just happenstance from the composition of other subgraphs can be done by noting that the count of each subgraph $(G_l)_{l=1}^K$ is a mixture of both genuinely and incidentally formed subgraphs. This provides a system of equations that can be solved for the parameters of interest. This system of equations also produces a simple method of moments estimator for the desired quantities, and Chandrasekhar and Jackson (2014) provide large sample characterizations for the estimator. The probabilities for each subgraph can also be made to depend on covariates. Finally, the authors also relate SUGM to mutual consent and search intensity network formation models. They provide an

application to microfinance-related network data collected by the authors in several Indian villages (e.g., Banerjee, Chandrasekhar, Duflo, and Jackson (2014)).

4.2 Strategic Network Formation

Most of the literature above relates to statistical models not (or at least not directly) related to economic models of network formation. Here I consider econometric models where agents purposefully form networks according to an explicit equilibrium notion and a payoff structure whereby node i 's utility function $U_i(g)$ depends on the network g and vertex and/or pairwise-specific variables, observable and unobservable to the econometrician. (I omit observable covariates below for simplicity). Though more general specifications are possible, one common specification (for undirected networks) involves variations of

$$U_i(g) \equiv \sum_{j \neq i} W_{ij} \times (u + \epsilon_{ij}) + |\cup_{j: W_{ij}=1} N_j(g) - N_i(g) - \{i\}| \nu + \sum_j \sum_{k > j} W_{ij} W_{ik} W_{jk} \omega, \quad (9)$$

where $N_i(g)$ denotes the set of nodes directly connected to node i and $|\cdot|$ is the cardinality of a given set. The vector $\epsilon_i \equiv (\epsilon_{ij})_{j \neq i}$ enumerates link-specific payoff shifters. I also retain the notation of using $W_{ij} \in \{0, 1\}$ to denote the establishment of a link from i to j . The first term in this payoff function enumerates the utility of direct connections and is indexed by the parameter u . The second term represents the payoff from indirect connections (one link away), and ν is the payoff per such an indirect link. The last term expresses any utility from mutual connections between vertices directly linked to node i . Whereas specific implementations may differ, they typically involve terms relating to each of these three aspects (direct connections, indirect connections, and mutual connections). Similar specifications exist for directed networks. Finally, the analysis may also differ depending on whether utility is transferable or not (see Jackson (2009) for a definition). Here I focus on non-transferable utility models, although transferable utility models have also been examined (see Lee and Fong (2011)).

One class of models focuses on an iterative network formation protocol and includes the papers by Christakis, Fowler, Imbens, and Kalianaraman (2010), Mele (2015) and Badev (2013). At each iteration of the meeting protocol, a pair of individuals and the relevant unobservable errors are drawn and (myopically) determine the formation or dissolution of an edge according to the payoff structure. This relates, for example, to earlier analyses on stochastic best response dynamics by Blume (1993) (for non-cooperative games) and, in the

context of network formation models, to Watts (2001) and Jackson and Watts (2002). A meeting protocol is also employed in the precursor article by Currarini, Jackson, and Pin (2009) (which focuses on direct links). (I mention in passing that, given the nature of the myopic sequential optimization process, a structural interpretation of the unobserved taste shocks and meeting protocol as a component of the data-generating process will be more or less adequate, depending on the empirical context.) The first paper above considers an undirected network. While Christakis, Fowler, Imbens, and Kalianaraman (2010) do not explicitly consider an equilibrium notion, if unobservable preference shocks are absent (or fixed) throughout the meeting protocol, the (undirected) network would converge to a pairwise stable configuration if one exists (see below for definition) or cycle if one does not exist (see Watts (2001)). In the models above, new unobservables are drawn at each meeting, and one can view these “perturbations” as producing a version of the stochastic stability analysis in Jackson and Watts (2002).

Mele (2015) and Badev (2013), on the other hand, study directed networks (Badev (2013) extends Mele (2015) to model link formation and behavioral choices—smoking—simultaneously). Under some conditions, Mele (2015) demonstrates the existence of a potential function to characterize the Nash equilibria of the model absent the unobservable error. (Badev (2013), who also relies on a potential function, on the other hand, introduces and focuses his analysis of the unobservable-error-free model on the concept of k -Nash stability, building on previous work by Bala and Goyal (2000).) Mele (2015) then shows that the meeting protocol and myopic best response dynamic form an ergodic Markov chain on the space of networks and converges to a unique invariant distribution (with modes at the maximands of the potential function alluded to above). In fact, when the unobservables are assumed to be i.i.d. extreme value distributed, the limiting distribution corresponds to that of an ERGM.³⁷ Given the practical difficulties in the estimation of ERGMs pointed out previously, I should mention that Mele (2015) also proposes a modified MCMC procedure and analyzes the procedure for this particular model along the lines of Bhamidi, Bresler, and Sly (2011), demonstrating that the slow convergence regions in the parameter space are relatively small for parsimonious parameterizations of the utility function where only direct links matter (see Mele (2015) for further details). Intuitively, the most parsimonious parameterization of an ERGM would correspond to an Erdős-Rényi model, for which the parameter space would be in the “high temperature” regime.

All three models above are fit to a network of friendships obtained from the AddHealth

³⁷The connection between potential games and ERGMs was independently noticed by Butts (2009).

data and use MCMC methods to produce Bayesian estimates of the parameters of interest. Whereas Christakis, Fowler, Imbens, and Kalianaraman (2010) use data from one school network, Mele (2015) estimates his model on three school networks, and Badev (2013) uses data from 14 school networks. (The AddHealth data has a total of 16 schools for which all information was collected.) The goodness of fit analysis provided by Christakis, Fowler, Imbens, and Kalianaraman (2010) demonstrate that a model with utility functions extending to indirect connections matches patterns of the observed networks well.

A different class of models focuses instead on a static framework. For undirected networks, a common solution concept adopted in those papers is that of pairwise stability (see Jackson and Wolinsky (1996)). A pairwise stable network is one for which:

$$\{i, j\} \in g \Rightarrow U_i(g) \geq U_i(g - \{i, j\}) \quad (10)$$

and

$$\{i, j\} \notin g \Rightarrow U_i(g \cup \{i, j\}) < U_i(g) \text{ or } U_j(g \cup \{i, j\}) < U_j(g). \quad (11)$$

The notation $g - \{i, j\}$ stands for the graph obtained by the deletion of the link $\{i, j\}$ from g , and $g \cup \{i, j\}$ denotes the graph obtained by adding the edge $\{i, j\}$ to g . This solution concept incorporates the idea that any link can be severed unilaterally, but the formation of a link requires mutual consent. Other solution concepts exist for undirected networks and related solution concepts can be employed for directed networks (Nash stability, for example) and when transfers are possible (see Jackson (2009)). As pointed out above, pairwise stable networks would be rest points for link formation sequences produced via a meeting protocol if the payoff structure does not change at each new meeting. In that case, the realized sequence of meetings could be seen as a selection among the possible stable networks which the approaches described below try to be agnostic about. The articles described above circumvent this issue by introducing noise in the meeting process (as unobservables are drawn anew at each meeting opportunity). A (latent) meeting protocol would still need to be specified.³⁸ Furthermore, the multiplicity partly reappears as the MCMC procedures used in the estimation tend to have the most difficulty in regions of the parameter space where the distribution of networks is multimodal (i.e., the “low temperature” regions), which in turn correspond to there being multiple (Nash or Nash stable) equilibria in the underlying game

³⁸In the directed network case, Badev (2013) shows nevertheless that, under certain restrictions on the class of meeting protocols, the invariant network distribution does not depend on the specification of the meeting process.

(without econometric errors).

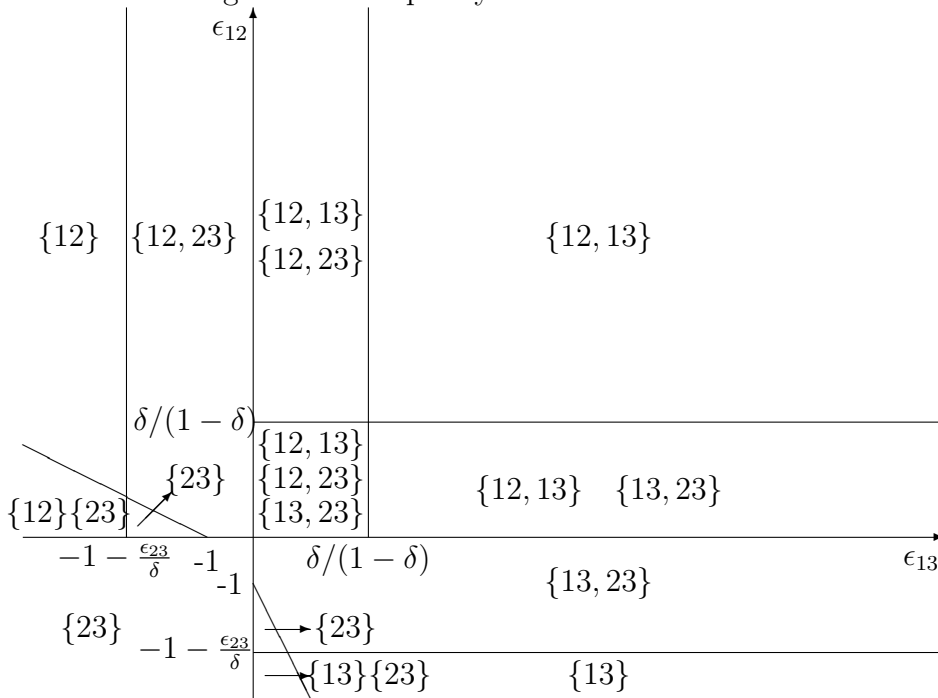
Aside from the proliferation of possible networks as the number of nodes grows, an added difficulty in the analysis of such models is the possible multiplicity of equilibria for given realizations of the payoff-relevant variables. To illustrate this, consider a simple three-node graph with payoffs given by

$$U_i(g) = \sum_{j \in \{1, \dots, n, j \neq i\}} \delta^{d(i,j;g)-1} (1 + \epsilon_{ij}) - |N_i(g)|,$$

where $\delta \in (0, 1)$ and $d(i, j; g)$ is the shortest distance between i and j in the graph g . This is an econometric version of the connections game in Jackson and Wolinsky (1996) where player i collects $1 + \epsilon_{ij}$ if directly connected to node j and $\delta^{d(i,j;g)-1} (1 + \epsilon_{ij})$ if indirectly connected to j and any direct connection cost her one util. The set of possible links is $\{12, 13, 23\}$, and there are eight possible networks (2^3). (To economize on notation, I use ij instead of $\{i, j\}$ to denote the edge between nodes i and j .) To visualize the multiplicity of solutions, I map the possible pairwise stable networks for a realization of ϵ s where $\epsilon_{ij} = \epsilon_{ji}$ for any i, j in the $(\epsilon_{12} \times \epsilon_{13})$ -space for $0 < \epsilon_{23} < \delta/(1 - \delta)$ (see appendix for a more detailed description).

To emulate the approach usually adopted in the empirical games literature (see de Paula (2013)), one could generate bounds on the parameters of interest (i.e., δ in this model) by noting that the model implies probability bounds for each (pairwise stable) network to be observed. For example, the probability of observing the network $\{12, 13\}$ is bounded below by the probability that it arises as a unique equilibrium (e.g., the probability of the NE corner of the figure) and bounded below by that quantity plus the probability that it is a pairwise stable network (but not the only one) (e.g., the positive quadrant). The first bound corresponds to the possibility that this network is never selected when other equilibrium networks are possible and the upper bound corresponds to the opposite scenario where this network is always selected. Such bounds would depend on the parameters of the model. *If one has access to a sample of networks*, one could estimate the identified set of parameters by collecting all those parameters for which the bounds contain the observed frequency of networks. It should be apparent that this task becomes computationally quite complex as the number of nodes increases: the dimensions of both the latent variable space and possible networks to be checked for stability increase relatively quickly. (Remember that with 24 nodes, the number of graphs is more than the estimated number of atoms in the observed universe.)

Figure 4: Multiplicity of Pairwise Stable Networks



Note: The figure shows the pairwise stable networks for different realization of ϵ_{12} and ϵ_{13} , assuming that $0 < \epsilon_{23} < \delta/(1 - \delta)$. A detailed description is given in Appendix B.

To ameliorate the computational difficulties highlighted above, Sheng (2014) focuses instead on subnetworks (i.e., subsets of vertices and the edges among them), checking whether those subgraphs are consistent with pairwise stability (with or without transferable utility) for undirected networks. A lower bound on the probability of observing a particular subgraph to a pairwise stable network is the probability that this subgraph be the only one to satisfy the pairwise stability conditions given any (potentially not pairwise stable) complementary network (i.e., the network after deletion of the subgraph’s edges from the overall graph, including the edges incident with the subgraph nodes). An upper bound is the probability that this subgraph satisfies the pairwise stability condition given some (potentially not pairwise stable) complementary network. As explained by Sheng (2014), these bounds are not sharp (even when considering subnetworks only) since the upper bound is larger than the probability that the subgraph is part of a pairwise stable network, and the lower bound is lower than the probability that it is a subgraph to a unique pairwise stable network given the payoff structure. Consider, for instance, the game depicted in Figure (4), and take the subgraph $\{12\}$. Because this link satisfies the pairwise stability condition when $\epsilon_{12} > 0$, given no links between 1 and 3 and between 2 and 3, the upper bound will be larger than the region where $\{12\}$ is a subgraph to any pairwise stable network (which would exclude the triangle where $\{23\}$ is the only pairwise stable network). Since the set of complementary networks may still be sizeable, the author imposes additional restrictions (payoffs of a link only depend on the remaining graph up to each player’s immediate neighborhood and a cap on the number of links each node can form). Using these bounds, usual techniques in the estimation of partially identified models can then be implemented, and the dimensionality of the problem is reduced from the cardinality of the vertex set to the count of vertices in the subgraphs analyzed. One possible issue is that the restrictions imposed may sacrifice identification power when networks involve a large number of nodes and yield larger bounds.

The computational burden can also be alleviated by exploring particular features of the model. Miyauchi (2014) studies a model with non-transferable utility, where payoffs are supermodular, and pairwise stable networks consequently correspond to a fixed point of a monotone mapping. The supermodularity condition requires that $U_i(g) - U_i(g - \{ij\})$ if $\{ij\} \in g$ and $U_i(g \cup \{ij\}) - U_i(g)$ for $\{ij\} \notin g$ be increasing in the adjacency matrix for every pair of nodes i and j . (The connections game depicted in Figure (4) does not satisfy this condition.) Since the set of fixed points in this case possesses a minimal and maximal element by Tarski’s fixed-point theorem, these can be used directly to formulate a computationally tractable estimator for the identified set. Miyauchi (2014) provides Monte Carlo experiments

and an empirical illustration using the AddHealth data. Boucher and Mourifié (2013) analyze a very similar framework, assuming high-level conditions on pointwise identification through a pseudo-likelihood objective function and (weak) dependence across agents in observed pairwise stable networks. Weak dependence and other conditions (e.g., homophily and diversity) are further explored in Leung (2015a) to study large sample properties of estimators in a network formation model where a link is formed whenever those involved receive a positive “joint surplus” (which may depend on other edges in the network). The assumptions in his model (which can also accommodate non-transferable utilities) deliver a sparse network as the number of nodes grows, and he provides an application to a network of physician referrals.

Sheng (2014) and Miyauchi (2014) focus on a sample scheme where a number of networks of (at most) moderate size are observed. An alternative empirical scenario is one in which one has access to very few networks (perhaps only one) and many nodes. de Paula, Richards-Shubik, and Tamer (2015) developed an algorithm to compute the identified set for the preference parameters in the context of large (pairwise stable) networks (without transferable utility). They approximate this large community by a continuum of nodes, where agents can only form a finite number of links, and payoffs depend only on a finite number of direct and indirect connections.³⁹ A pairwise network will then be represented by a continuous graph with bounded degrees (since nodes have a finite number of incident edges), corresponding to a sparse network (see previous discussion about the empirical plausibility of such social and economic graphs). Such mathematical objects, called *graphings* in the applied mathematics literature, are sometimes used to approximate large networks as limits for large discrete graphs under a well-defined convergence metric (e.g., Lovasz (2012) for a recent survey).⁴⁰ To further reduce the dimensionality of the problem, they also assume that unobservable taste shocks depend only on the covariates of putative connections and not on their identity. If covariates have a finite support, individual nodes can be classified into a finite (albeit possibly large) number of “network types,” which provide a description of an individual node’s local network. In the example introduced above, for instance, if 1, 2, and 3 stand for possible characteristics of a node (from a continuum of vertices) and individuals can only establish

³⁹Formally, one needs to be careful in working with the continuum. Since unobservables are assumed to be independently drawn across individual nodes, measurability complications need to be handled using results such as those presented in, e.g., Uhlig (1996) or Sun (2006). Another issue that appears in working in coalitional games on the continuum, when coalitions themselves can be a continuum, is avoided in this framework by restricting individuals to form a finite number of links where only the characteristics on the potential connection (and not her identity) matter.

⁴⁰The corresponding approximation to dense graphs, known as *graphons*, is used, for example, by Chatterjee and Diaconis (2013) in their asymptotic study of ERGMs.

one connection, a given individual node would have three unobservable taste shocks: one for each one of the possible neighbour covariates. (The payoff to being isolated is normalized to zero.) In this example, a network type would describe the characteristic of a node (1, 2, or 3) and that of this neighbor (1, 2, 3 or whether the node has no neighbors).

The proportion of network types in an observed pairwise stable network is an equilibrium outcome, potentially estimable even from *incompletely observed* networks. Hence, whereas the number of vertices may be overly numerous, the cardinality of the set of network types is controlled. This allows them to verify whether observed networks correspond to pairwise stable networks for a given preference parameter vector using a computationally tractable quadratic program! To do so, they first define a partition of the space of unobservable variables. A set in this partition, called a “preference class,” corresponds to the set of network types for which an individual agent with given realization for the unobservable preference shocks would not be inclined to drop a connection from. In the running example used here, those individuals for whom all the taste shocks are positive would prefer a link to any individual (regardless of whether her covariate is 1, 2, or 3) to being isolated. Her preference class would comprise network types connected to each of the three characterizing labels and the isolated type (since there are no links to drop in that case). Then, one can presumably allocate nodes to network types, and the proportion of nodes in a preference class allocated to a specific network type is called by them an “allocation parameter”.

Given preference parameters, a pairwise stable network will correspond to allocation parameters satisfying certain restrictions. The requirement that a link in a pairwise stable network should be beneficial to both parties involved (i.e., (10)) is encoded in the definition of preference classes (since by definition connections characterizing network types in a preference class would not be dropped) and allocations from any preference class to network types not in that class are set to zero. de Paula, Richards-Shubik, and Tamer (2015) also show that the requirement that absent links be detrimental to at least one of the parties involved (i.e., (11)) implies that a quadratic form (on the vector of allocation parameters) be equal to zero. (This necessary condition is also sufficient in certain models and can thus yield sharp identified sets.) Using the restriction that the proportion of network types corresponds to observed ones, which guarantees that the supply and demand of links are balanced, and positivity constraints, the authors express the verification in terms of a quadratic program attaining a minimand equal to zero. Hence, instead of checking for pairwise stability among all possible networks involving a potentially large number of vertices and realizations of the unobservables for a given parameter value, the verification task is reduced to the solution of a quadratic

program defined on the vector allocation parameters and whether the attained minimum is equal to zero and is computationally appealing. For example, de Paula, Richards-Shubik, and Tamer (2015) present a simulation study on a model based on (9), where individuals can form up to three connections on 500 nodes, and covariates take two values. The evaluation of the quadratic program takes on average less than 30 seconds.⁴¹ It should also be pointed out that the computations above rely on the estimated proportion of network types. This can be regarded as an aggregate over observed pairwise stable networks with sampling variability stemming from differently selected equilibrium networks or as an estimate of network-type proportions from an incompletely observed network. As long as a distributional theory is provided for these statistics, the distributional features of the structural parameters can be obtained since the quadratic programming provides a mapping between the two.

The models above presume a complete information framework, where payoffs to all players involved are known to each one. An alternative strand of models focuses on directed networks and an incomplete information environment, whereby an individual node’s preference shocks are unobservable, not only to the econometrician but also to other nodes. This is analyzed, for example, in Gilleskie and Zhang (2009) and Leung (2015b). Both articles employ a multi-step estimation strategy where equilibrium beliefs are estimated at a first step from linking decisions and used in the estimation of payoffs in a second stage. The first step is made possible because private information is independent across agents, and beliefs about other agents’ linking decision do not depend on one’s private information. Analogous multi-stage conditional choice probability-based strategies have been employed in the dynamic programming discrete choice and empirical games literature under similar assumptions (e.g., de Paula (2013)), and Gilleskie and Zhang (2009) use a related framework, so only direct neighbors enter the utility function. Their main goal is to empirically study peer effects in smoking behavior (as in Badev (2013)) while allowing for links to be formed purposefully by the agents involved. As in previous studies, they employed the AddHealth data in their analysis. Leung (2015b) offers a related estimator focusing on the network formation. Here, though, the payoff structure involves the usual graph theoretic configurations as in (9), and the statistical analysis is performed for a small number of large networks instead of a large number of small

⁴¹Miyauchi (2014) considers a simulation study on a similar payoff structure (setting $\nu = 0$) and covariates taking four values (gender \times race). The simulation studies are not directly comparable since they employ different machines, and the model there is simpler in some dimensions (no preference for indirect connections, other than for mutual connections), but more complex in others (dimension of covariate support). For 200 nodes (as opposed to 500 nodes in de Paula, Richards-Shubik, and Tamer (2015)), the evaluation of the model for a single parameter takes about 2286 seconds (≈ 38 minutes) for 100 sampled networks and using 100 simulations for the construction of his objective function.

games (as in Bisin, Moro, and Topa (2011) and Menzel (2015)). An empirical illustration using microfinance-related data from Banerjee, Chandrasekhar, Duflo, and Jackson (2014) is also given. Both papers employ the assumption that a unique equilibrium is present in the data (even if the payoff structure is amenable to multiple solutions), which is not uncommon in the empirical games literature.

Finally, the models discussed previously are static or, at best, myopically dynamic models of network formation. Whereas farsighted models of network formation exist in the theoretical literature (e.g., Jackson (2009)), partly due to data (un)availability and/or the computational complexities even in static settings, they have not been very explored in the applied econometric literature. Alternative models for network formation, based on non-cooperative equilibrium concepts and exploring dynamics with forward-looking behavior, have nevertheless been proposed and could be used in data scenarios where network evolution is observable. Lee and Fong (2011), for instance, propose a dynamic network formation model for bipartite networks, where agents are split into two groups and across-group connections are established. There, payoffs are transferable, and a link is interpreted as a negotiation channel entailing a bargaining game over a contemporaneous surplus and quantifies the value of a particular edge. Connections are established via a link announcement game, where negotiations are open if two parties announce a putative link with each other. Because costs of establishing a negotiation link depend on the previous state of the bargaining connections, the model lends itself to dynamic incentives, and the authors focus on Markov perfect equilibria for the network formation process. This brings them closer to the empirical dynamic games literature and the estimation strategies suggested there (using methods developed after the paper by Hotz and Miller (1993) for individual dynamic decisions and adapted to strategic interactions by various authors in the 2000s). Lee and Fong (2011) provide an illustration on insurer-provider contracting in health care through a series of simulations studies. Similar ideas appear in Johnson (2012).

I end this subsection by pointing to surveys in statistics and econometrics that encompass the class of network formation models in this and the previous subsections. Those surveys are distinct in focus and serve as an adequate complement to the discussion provided above. Those are Kolaczyk (2009), Goldenberg, Zheng, Fienberg, and Airoldi (2009), and Hunter, Krivitsky, and Schweinberger (2012) (in statistics) and Graham (2015) and Chandrasekhar (2015) (in econometrics).

4.3 Network Formation and Outcomes

Whereas network formation may be of interest in its own right, the models discussed above can be seen as a stepping stone for modeling outcomes discussed in the previous section, since the decision to establish links may be informed by any outcomes determined via the ensuing social structure. As mentioned earlier, for example, Gilleskie and Zhang (2009) and Badev (2013) studied econometric models for simultaneous link formation and discrete behavior (i.e., smoking behavior).

Goldsmith-Pinkham and Imbens (2013) model the joint determination of social networks and a continuous outcome (high school grade) using a dyadic edge formation framework for the former and a linear-in-means model for the latter using the AddHealth data. Because the network is observed on two different occasions, they postulate that the links are formed through pairwise stability based on a felicity function that depends solely on direct links (i.e., parameters ν and ω are set to zero in (9)), and the previously observed network as a state variable.⁴² Link formation and outcomes are connected by the presence of individual specific unobservables ξ_i and covariates (in this case, previous grade point average). Both ingredients affect an individual’s outcome directly in the linear-in-means model, and links are formed based on an affinity between the covariates and the individual specific unobservables for the two parties involved in the putative link (which can be seen as a “latent position model” as in Hoff, Raftery, and Handcock (2002)). The presence of this individual effect introduces network endogeneity in the linear-in-means system as links are related to unobservables that determine those very outcomes. In their empirical application, estimated using Bayesian methods, the authors find that the individual specific unobservable driving both network formation and grades determination improves the fit for the network formation model but does not do so appreciably for the estimation of the linear-in-means model.

One natural approach to integrating network formation and outcome interactions takes the payoffs from network formation as indirect utilities, subsuming the potential payoffs from the economic system determining outcomes after the network is formed. One econometric model that can be cast along those lines is Hsieh and Lee (2013) (who also applied their model to the AddHealth, looking at smoking and grades as outcomes of interest). Link formation is formed based on a joint surplus function that depends on the network in a way similar to Mele (2015) and behaviors (as in Badev (2013)), producing an ERG model at the network formation stage (as in the papers cited above). Outcomes are determined

⁴²Since indirect edges are not payoff-relevant and there are no restrictions on the number of links formed, there is a single pairwise stable network given realizations for the unobservable preference shocks.

through a linear-in-means model (for grades) and a Tobit version of the linear-in-means model (for smoking, measured as the frequency in the year preceding the survey). The model is estimated using Bayesian methods, as in those other articles. Hsieh and Lee (2013) found that behaviors appear significantly in link formation and that network interactions matter for both outcomes.⁴³

One important aspect affecting the joint econometric analysis of network formation and interactions is the possibility of multiplicity in network formation and/or at the outcome determination system. As seen earlier, this is a possibility depending on the solution concept adopted and on other details of the framework at hand (e.g., whether utility is transferable or not). In this case, partial identification in either stage — formation or interactions — will likely be transmitted to other parameters of the model. For example, even if point identification at the network interactions model is achieved along the lines highlighted previously, if the network formation protocol is subject to partial identification, parameters in a linear-in-means model will possibly only be partially identified. An illustration of this is given, for example, in Ciliberto, Murry, and Tamer (2015) in the context of an empirical entry-exit games in industrial organization. Aside from joint modeling and estimation of both network formation and interactions, another possibility is to consider instruments for networks as suggested, for example, in Qu and Lee (2015). Of course, if network formation is prone to multiplicity, the model is incomplete without an equilibrium selection rule (see Tamer (2003)). New developments in the partial identification literature may nonetheless prove useful here (e.g., Chesher and Rosen (2014)).

5 Measuring Networks and Outcomes

It should be noted that a few econometric models for network formation highlighted in the previous section presume the availability of data on the complete network. Whereas some strategies do not require observation of the whole network (e.g., in independent dyadic mod-

⁴³Hsieh and Lee (2013), Goldsmith-Pinkham and Imbens (2013), and Badev (2013) use different subsets of the AddHealth data. They also rely on different outcome measurements and covariates—Badev (2013) uses a binary variable related to smoking behavior in the month prior to the survey, whereas Hsieh and Lee (2013) use a multivalued measure for the past year; Badev (2013) incorporates the price of cigarettes, whereas Hsieh and Lee (2013) do not. Furthermore, they all employ different models—Badev’s model for smoking does not correspond to a binary model of the linear-in-means model, for instance, and Goldsmith-Pinkham and Imbens (2013) use network data from two survey waves to construct their model. Hence the estimates across these three models are hard to compare. From a practitioner’s viewpoint, it would be interesting to compare these competing frameworks.

els), others are more demanding. ERG models, for example, are typically not “projective,” which implies that estimators based on subgraphs are not consistent (see Shalizi and Rinaldo (2013)). As indicated by those authors, if incomplete network information is available and “an ERGM is postulated for the whole network, then inference for its parameters must explicitly treat the unobserved portions of the network as missing data (perhaps through an expectation-maximization algorithm), though of course there may be considerable uncertainty about just how much data is missing” (p. 523). On this latter point, see, e.g., Handcock and Gile (2010) or Koskinen, Robins, and Pattison (2010). When repeated outcomes are observed for a given system, the methods suggested in Manresa (2013) and de Paula, Rasul, and Souza (2015) may also be useful for network information retrieval.

For some of the econometric models described above, though, complete observation of the network may not be necessary, and relevant features of the network can be estimated, provided that a suitable sampling scheme is given. Kolaczyk (2009), for example, provides estimators for various network features like the total number of edges or the total number of triangles. These could presumably be used for the estimation of the SUG model proposed by Chandrasekhar and Jackson (2014). The proportion of “network types” contemplated in de Paula, Richards-Shubik, and Tamer (2015) could potentially be estimated along similar lines. A review of available methods for the estimation of graph features from sampled subnetworks is given in Kolaczyk (2009).

In the context of outcome interactions, Moffitt (2001) and Angrist (2014) aptly point out that measurement issues may compromise any identification results assuming no mismeasurement (see also Ammermueller and Pischke (2009)). When using between-group variation in covariates to identify β through its connection to OLS and 2SLS coefficients, measurement errors in covariates will typically produce attenuation in the first, though not the second.⁴⁴ In the covariance restriction case studied by Moffitt (2001) (Proposition 2), if outcomes or covariates are measured with error, measurement error variances will be confounded in the variance of observed outcomes. These are indeed empirically relevant considerations that researchers should be aware of. In his study of randomly assigned roommates, for example, Sacerdote (2001) accounts for the possibility of mismeasured covariates by including classical measurement error, which precludes him from using covariance restrictions as suggested in Moffitt (2001). Ammermueller and Pischke (2009) carefully discuss the consequences of measurement error in covariates in their analysis for peer effects in education. Under classical

⁴⁴As pointed out by Angrist (2014), other considerations may also drive a wedge between those two, even when there are no peer effects.

measurement error in *outcomes*, a related strategy can nevertheless be employed to obtain the endogenous effect. If classical measurement errors are independent across peers in a given group, their covariance washes out as we consider observed outcome covariances, and contrasts in those covariances across groups may still allow for point identification of the endogenous effect. Below I demonstrate this for groups of size two and three (double and triple rooms in Sacerdote’s context). In the proposition below, $\tilde{y}_{i,g}$ denotes the observed outcome for individual i in group g , and $v_{i,g}$ denotes the measurement error.

Proposition 6. *Suppose there are two groups $g = 1, 2$ such that $N_1 = 2$ and $N_2 = 3$. If $|\beta| < 1$, $W_{ij,g} = (N_g - 1)^{-1}$ if $i \neq j$, $W_{ii,g} = 0$, $\mathbb{V}(\epsilon_g | \mathbf{x}_g) = \sigma^2 \mathbf{I}_g$ and $\tilde{y}_{i,g} = y_{i,g} + v_{i,g}$ where $v_{i,g} \perp\!\!\!\perp v_{j,h}$ and $v_{i,g} \perp\!\!\!\perp y_{j,h}$ for any i, j, g and h , then β is identified.*

Of course, that measurement errors be classical is in itself a strong assumption, and Proposition 6 above should be interpreted with that in mind. It also assumes that covariates are *not* measured with error, which was the focus of the above-mentioned concerns. The result nevertheless demonstrates that network interaction models may themselves provide additional structure to be explored in the estimation of measurement error-ridden models.

6 Conclusion

This article provided a (selective) review of recent works on networks and outcomes mediated through networks. This is an area of active research, but much remains to be learned.

Regarding models of network effects on outcomes, I should mention that heterogeneity is an important issue that I have not explicitly discussed in this work. As pointed out by Sacerdote (2010) in his review of peer effects studies in education, a “linear-in-means model masks considerable heterogeneity in the effects experienced by different types of students.” Nonlinear and heterogeneous effects models could prove useful. Tincani (2015) provides an interesting empirical examination of nonlinearities in social interactions. Of course, should nonlinearities be relevant, multiple social equilibria are a possibility. Some of the ideas mentioned in this article allow for heterogeneous peer coefficients (e.g., de Paula, Rasul, and Souza (2015), if adequate data is available) and Masten (2015) is a recent contribution in the development of random coefficient versions to the models considered here. Measurement issues present important practical difficulties. As indicated by Ammermueller and Pischke (2009), more attention to measurement issues could bring in important rewards. As far as network formation models are concerned, econometric methodologies that take into account

sampling and measurement peculiarities are also important in practice. Computational and identification difficulties are also a primary concern, and integrated models of network formation and interactions will typically inherit those.

In both cases, most models examined are static. Dynamic, forward-looking models may be adequate for many applications (e.g., industrial organization and banking), especially as more detailed and abundant data on the evolution of networks and outcomes becomes available.

References

- ACEMOGLU, D., AND J. ANGRIST (2001): “How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws,” in *NBER Macroeconomics Annual*, ed. by B. S. Bernanke, and K. Rogoff, vol. 15 of 9-74. MIT Press.
- ACEMOGLU, D., V. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): “The Network Origins of Aggregate Fluctuations,” *Econometrica*, 80(5), 1977–2016.
- ACEMOGLU, D., A. OZDAGLAR, AND A. TAHBAZ-SALEHI (forthcoming): “Networks, Shocks, and Systemic Risk,” in *The Oxford Handbook on the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, and B. Rogers. Oxford University Press.
- AMMERMUELLER, A., AND J.-S. PISCHKE (2009): “Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study,” *Journal of Labor Economics*, 27(3), 315–348.
- ANGELUCCI, M., AND G. DE GIORGI (2009): “Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption?,” *American Economic Review*, 99(1), 486–508.
- ANGRIST, J. (2014): “The Perils of Peer Effects,” *Labour Economics*, 30, 98–108.
- ANTRAS, P., AND D. CHOR (2013): “Organizing the Global Value Chain,” *Econometrica*, 81(6), 2127–2204.
- ATALAY, E., A. HORTACSU, J. ROBERTS, AND C. SYVERSON (2011): “Network Structure of Production,” *Proceedings of the National Academy of Sciences*, 108(13), 5199–5202.
- BADEV, A. (2013): “Discrete Games with Endogenous Networks: Theory and Policy,” Working Paper, University of Pennsylvania.
- BALA, V., AND S. GOYAL (2000): “A Noncooperative Model of Network Formation,” *Econometrica*, 68(5), 1181–1229.
- BALLESTER, C., A. CALVÓ-ARMENGOL, AND Y. ZENOU (2006): “Who’s Who in Networks. Wanted: The Key Player,” *Econometrica*, 74, 1403–1417.

- BANERJEE, A., A. CHANDRASEKHAR, E. DUFLO, AND M. JACKSON (2014): “Gossip: Identifying Central Individuals in a Social Network,” MIT Working Paper.
- BARABÁSI, A., AND R. ALBERT (1999): “Emergence of Scaling in Random Networks,” *Science*, 286, 509–512.
- BEKKER, P., AND D. POLLOCK (1986): “Identification of Linear Stochastic Models with Covariance Restrictions,” *Journal of Econometrics*, pp. 179–208.
- BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2013): “Inference Methods for High-Dimensional Sparse Econometric Models,” in *Advances in Economics and Econometrics, Theory and Applications: Tenth World Congress of the Econometric Society*, ed. by D. Acemoglu, and M. Arellano, vol. 3. Cambridge: Cambridge University Press.
- BERGE, C. (1962): *The Theory of Graphs and Its Applications*. Methuen Wiley.
- BERNARD, A. B., A. MOXNES, AND Y. SAITO (2014): “Geography and Firm Performance in the Japanese Production Network,” RIETI Discussion Paper 14-E-034.
- BESAG, J. (1974): “Spatial Interactions and the Statistical Analysis of Lattice Systems (with Discussions),” *Journal of the Royal Statistical Society B*, 36, 196–236.
- (1975): “Statistical Analysis of Non-Lattice Data,” *The Statistician*, 24(3), 179–195.
- BHAMIDI, S., G. BRESLER, AND A. SLY (2011): “Mixing Time of Exponential Random Graphs,” *The Annals of Applied Probability*, 21(6), 2146–2170.
- BISIN, A., A. MORO, AND G. TOPA (2011): “The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions,” NBER Working Paper 17196.
- BLOCH, F., AND M. JACKSON (2007): “The Formation of Networks with Transfers among Players,” *Journal of Economic Theory*, 113(1), 83–110.
- BLUME, L. (1993): “The Statistical Mechanics of Strategic Interaction,” *Games and Economic Behavior*, 5, 387–424.
- BLUME, L., W. BROCK, S. DURLAUF, AND Y. IOANNIDES (2011): “Identification of Social Interactions,” in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. Jackson, vol. 1B, pp. 853–964. North-Holland, Amsterdam.
- BLUME, L., W. BROCK, S. DURLAUF, AND R. JAYARAMAN (2015): “Linear Social Interaction Models,” *Journal of Political Economy*, 123(2), 444–496.
- BOLLOBÁS, B. (2001): *Random Graphs*. Cambridge University Press, Cambridge.
- BONACICH, P. (1972): “Factoring and Weighting Approaches to Status Scores and Clique Identification,” *Journal of Mathematical Sociology*, 2(113-120).
- (1987): “Power and Centrality: A Family of Measures,” *American Journal of Sociology*, 92, 1170–1182.

- BONALDI, P., A. HORTACSU, AND J. KASTL (2014): “An Empirical Analysis of Systemic Risk in the EURO-zone,” University of Chicago Working Paper.
- BOOZER, M. A., AND S. E. CACCIOLA (2001): “Inside the ‘Black Box’ of Project STAR: Estimation of Peer Effects Using Experimental Data,” Yale Economic Growth Center Discussion Paper No. 832.
- BOUCHER, V., Y. BRAMOULLÉ, H. DJEBBARI, AND B. FORTIN (2014): “Do Peers Affect Student Achievement? Evidence From Canada Using Group Size Variation,” *Journal of Applied Econometrics*, 29(1), 91–109.
- BOUCHER, V., AND I. MOURIFIÉ (2013): “My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks,” Université Laval and University of Toronto Working Paper.
- BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): “Identification of Peer Effects Through Social Networks,” *Journal of Econometrics*, 150, 41–55.
- BRAMOULLÉ, Y., R. KRANTON, AND D’AMOURS (2014): “Strategic Interaction and Networks,” *American Economic Review*, 104(3), 898–930.
- BRAUN, M., AND J. MCAULIFFE (2010): “Variational Inference for Large-Scale Models of Discrete Choice,” *Journal of the American Statistical Association*, 105(489), 324–335.
- BRIN, S., AND L. PAGE (1998): “The Anatomy of a Large Scale Hypertextual Web Search Engine,” *Computer Networks*, 30, 107–117.
- BROCK, W., AND S. DURLAUF (2001): “Discrete Choice with Social Interactions,” *Review of Economic Studies*, 68, 235–261.
- (2006): “Multinomial Choice with Social Interactions,” in *The Economy as an Evolving Complex System*, ed. by S. Durlauf, and L. Blume, vol. 3, pp. 175–206. Oxford University Press, New York.
- (2007): “Identification of Binary Choice Models with Social Interactions,” *Journal of Econometrics*, 140(1), 57–75.
- BUTTS, C. (2009): “A Behavioral Micro-foundation for Cross-sectional Network Models,” UC Irvine Working Paper.
- CALVÓ-ARMENGOL, A., E. PATACCHINI, AND Y. ZENOU (2009): “Peer Effects and Social Networks in Education,” *Review of Economic Studies*, 76(4), 1239–1267.
- CARVALHO, V. (2014): “From Micro to Macro via Production Networks,” *Journal of Economic Perspectives*, 28(4), 23–48.
- CARVALHO, V., M. NIREI, AND Y. SAITO (2014): “Supply Chain Disruptions: Evidence from the Great East Japan Earthquake,” Cambridge University Working Paper.

- CHANDRASEKHAR, A. (2015): “Econometrics of Network Formation,” in *Oxford Handbook on the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, and B. Rogers. Oxford University Press.
- CHANDRASEKHAR, A. G., AND M. O. JACKSON (2014): “Tractable and Consistent Random Graph Models,” Working Paper 20276, NBER.
- CHARBONNEAU, K. B. (2014): “Multiple Fixed Effects in Binary Response Panel Data,” Bank of Canada Working Paper 2014-17.
- CHATTERJEE, S., AND A. DEMBO (2014): “Nonlinear Large Deviations,” Stanford University Working Paper, arXiv:1401.3495v5.
- CHATTERJEE, S., AND P. DIACONIS (2013): “Estimating and Understanding Exponential Random Graph Models,” *Annals of Statistics*, 41(5), 2428–2461.
- CHATTERJEE, S., P. DIACONIS, AND A. SLY (2011): “Random Graphs with a Given Degree Sequence,” *The Annals of Applied Probability*, 21(4), 1400–1435.
- CHESHER, A., AND A. ROSEN (2014): “Generalized Instrumental Variable Models,” CeMMAP Working Paper 04/14.
- CHRISTAKIS, N., J. FOWLER, G. IMBENS, AND K. KALIANARAMAN (2010): “An Empirical Model of Strategic Network Formation,” Harvard University Working Paper.
- CHWE, M. (2000): “Communication and Coordination in Social Networks,” *Review of Economic Studies*, 67, 1–16.
- CILIBERTO, F., C. MURRY, AND E. TAMER (2015): “Inference on Market Power in Markets with Multiple Equilibria,” University of Virginia and Harvard University Working Paper.
- CLAUSET, A., C. SHALIZI, AND M. NEWMAN (2009): “Power-Law Distributions in Empirical Data,” *SIAM Review*, pp. 661–703.
- COHEN-COLE, E., A. KIRILENKO, AND E. PATACCHINI (2014): “Trading Network and Liquidity Provision,” *Journal of Financial Economics*, 113, 235–251.
- COLEMAN, J. S. (1964): *Introduction to Mathematical Sociology*. London Free Press Glencoe.
- COMOLA, M., AND S. PRINA (2014): “Do Interventions Change the Network? A Dynamic Peer Effect Model Accounting for Network Changes,” PSE Working Paper.
- CURRARINI, S., M. JACKSON, AND P. PIN (2009): “An Economic Model of Friendship: Homophily, Minorities and Segregation,” *Econometrica*, 77(4), 1003–1045.
- DAVEZIES, L., D’HAULTFOEUILLE, AND D. FOUGÈRE (2009): “Identification of Peer Effects Using Group Size Variation,” *Econometrics Journal*, 12(3), 397–413.
- DE GIORGI, G., M. PELLIZARI, AND S. REDAELLI (2010): “Identification of Social Interactions through Partially Overlapping Peer Groups,” *American Economic Journal: Applied Economics*, 2(2), 241–275.

- DE PAULA, A. (2009): “Inference in a Synchronization Game with Social Interactions,” *Journal of Econometrics*, 148, 56–71.
- (2013): “Econometric Analysis of Games with Multiple Equilibria,” *Annual Review of Economics*, 5, 107–131.
- DE PAULA, A., I. RASUL, AND P. C. SOUZA (2015): “Identifying and Estimating Social Connections from Outcome Data,” UCL and PUC-Rio Working Paper.
- DE PAULA, A., S. RICHARDS-SHUBIK, AND E. TAMER (2015): “Identification of Preferences in Network Formation Games,” UCL Working Paper.
- DE PAULA, A., AND J. A. SCHEINKMAN (2010): “Value Added Taxes, Chain Effects and Informality,” *American Economic Journal: Macroeconomics*, 2, 195–221.
- DE PAULA, A., AND X. TANG (2012): “Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information,” *Econometrica*, 80(1), 143–172.
- DENBEE, E., C. JULLIARD, Y. LI, AND K. YUAN (2014): “Network Risk and Key Players: A Structural Analysis of Interbank Liquidity,” LSE Working Paper.
- DIEBOLD, F. X., AND K. YILMAZ (2015): *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press.
- DIEYE, R., H. DJEBBARI, AND F. BARRERA-OSORIO (2014): “Accounting for Peer Effects in Treatment Response,” IZA Discussion Paper 2014.
- DRUKKER, D., P. EGGER, AND I. PRUCHA (2013): “On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors,” *Econometric Reviews*, 32(5-6), 686–733.
- DZEMSKI, A. (2014): “An Empirical Model of Dyadic Link Formation in a Network with Unobserved Heterogeneity,” University of Mannheim Working Paper.
- ERDÖS, P., AND A. RÉNYI (1959): “On Random Graphs,” *Publicationes Mathematicae Debrecen*, 6, 290–297.
- (1960): “On the Evolution of Random Graphs,” *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 17–61.
- FAFCHAMPS, M., AND S. LUND (2003): “Risk Sharing Networks in Rural Philippines,” *Journal of Development Economics*, 71, 261–287.
- FAN, J., AND R. LI (2001): “Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties,” *Journal of the American Statistical Association*, 96(456), 1348–1360.
- FERNANDEZ-VAL, I., AND M. WEIDNER (2014): “Individual and Time Effects in Nonlinear Panel Models with Large N, T ,” Boston University Working Paper.
- FISHER, F. (1966): *The Identification Problem in Econometrics*. New York: McGraw-Hill.

- FORTIN, B., G. LACROIX, AND M.-C. VILLEVAL (2007): “Tax Evasion and Social Interactions,” *Journal of Public Economics*, 91(11-12), 2089–2112.
- FRANK, O., AND D. STRAUSS (1986): “Markov Graphs,” *Journal of American Statistical Association*, 81, 832–842.
- GALBIATI, R., AND G. ZANELLA (2012): “The Tax Evasion Social Multiplier: Evidence from Italy,” *Journal of Public Economics*, 96, 485–494.
- GANDHI, A., S. NAVARRO, AND D. RIVERS (2013): “On the Identification of Production Functions: How Heterogeneous is Productivity?,” University of Wisconsin Working Paper.
- GAUTIER, E., AND A. TSYBAKOV (2014): “High-Dimensional Instrumental Variables Regression and Confidence Sets,” Working Paper CREST.
- GEYER, C. J. (2009): “Likelihood Inference in Exponential Families and Directions of Recession,” *Electronic Journal of Statistics*, 3, 259–289.
- GILBERT, E. (1959): “Random Graphs,” *Annals of Mathematical Statistics*, 30(4), 1141–1144.
- GILLESKIE, D., AND Y. S. ZHANG (2009): “Friendship Formation and Smoking Initiation Among Teens,” University of North Carolina Working Paper.
- GLAESER, E., B. SACERCOTE, AND J. A. SCHEINKMAN (2003): “The Social Multiplier,” *Journal of the European Economic Association*, 1(2-3), 345–353.
- GLAESER, E., B. SACERDOTE, AND J. A. SCHEINKMAN (1996): “Crime and Social Interactions,” *Quarterly Journal of Economics*, 111(2), 507–548.
- GOLDENBERG, A., A. ZHENG, S. E. FIENBERG, AND E. M. AIROLDI (2009): “A Survey of Statistical Network Models,” *Foundations and Trends in Machine Learning*, 2(2), 129–333.
- GOLDSMITH-PINKHAM, P., AND G. IMBENS (2013): “Social Networks and the Identification of Peer Effects,” *Journal of Business and Economic Statistics*, 31(3), 253–264.
- GOULD, P. (1967): “On the Geographical Interpretation of Eigenvalues,” *Transactions of the Insititute of British Geographers*, 42, 53–86.
- GRAHAM, B. (2008): “Identifying Social Interactions through Conditional Variance Restrictions,” *Econometrica*, 76(3), 643–660.
- (2014): “An Empirical Model of Network Formation: Detecting Homophily when Agents are Heterogeneous,” University of California at Berkeley Working Paper.
- (2015): “Methods of Identification in Social Networks,” *Annual Review of Economics*, 7, University of California at Berkeley Working Paper.
- GRANOVETTER, M. (1973): “The Strength of Weak Ties,” *American Journal of Sociology*, 78, 1360–1380.

- GURYAN, J., K. KORY, AND M. J. NOTOWIDIGDO (2009): “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments,” *American Economic Journal: Applied Economics*, 1(4), 34–68.
- HANDCOCK, M. (2003): “Assessing Degeneracy in Statistical Models for Social Networks,” University of Washington Working Paper.
- HANDCOCK, M., AND K. GILE (2010): “Modeling Social Networks from Sampled Data,” *The Annals of Applied Statistics*, 4(1), 5–25.
- HAUSMAN, J., W. NEWEY, AND W. TAYLOR (1987): “Efficient Estimation and Identification of Simultaneous Equation Models with Covariate Restrictions,” *Econometrica*, 55(4), 849–874.
- HOFF, P. (2005): “Bilinear Mixed-Effects Models for Dyadic Data,” *Journal of the American Statistical Association*, 100, 286–295.
- HOFF, P., A. RAFTERY, AND M. HANDCOCK (2002): “Latent Space Approaches to Social Network Analysis,” *Journal of the American Statistical Association*, 97, 1090–1098.
- HOLLAND, P. W., AND S. LEINHARDT (1981): “An Exponential Family of Probability Distributions for Directed Graphs,” *Journal of the American Statistical Association*, 76, 33–65.
- HONG, H., AND J. XU (2014): “Count Models of Social Networks in Finance,” Princeton University Working Paper.
- HONORÉ, B. E., AND A. DE PAULA (2010): “Interdependent Durations,” *Review of Economic Studies*, 77, 1138–1163.
- HOTZ, V. J., AND R. A. MILLER (1993): “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, 60(3), 497–529.
- HSIEH, C.-S., AND L.-F. LEE (2013): “Specification and Estimation of Network Formation and Network Interaction Models with the Exponential Probability Distribution,” OSU Working Paper.
- HUNTER, D. R., P. N. KRIVITSKY, AND M. SCHWEINBERGER (2012): “Computational Statistical Methods for Social Network Models,” *Journal of Computational and Graphical Statistics*, 21(4), 856–882.
- JACKSON, M. (2009): *Social and Economic Networks*. Princeton University Press.
- JACKSON, M., AND B. ROGERS (2007): “Meeting Strangers and Friends of Friends: How Random Are Social Networks?,” *American Economic Review*, 97(3), 890–915.
- JACKSON, M., AND A. WATTS (2002): “The Evolution of Social and Economic Networks,” *Journal of Economic Theory*, 106, 265–295.

- JACKSON, M., AND A. WOLINSKY (1996): “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory*, 71(1), 44–74.
- JOHNSON, T. (2012): “Dynamic Network Formation: Theory and Estimation,” University of Notre Dame Working Paper.
- JORDAN, M. I. (2004): “Graphical Models,” *Statistical Science*, 19(1), 140–155.
- JORDAN, M. I., AND M. J. WAINWRIGHT (2008): “Graphical Models, Exponential Families, and Variational Inference,” *Foundations and Trends in Machine Learning*, 1(1-2), 1–305.
- KATZ, L. (1953): “A New Status Index Derived from Sociometric Analysis,” *Psychometrika*, 18, 39–43.
- KELEJIAN, H. H., AND I. R. PRUCHA (2010): “Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedasticity Disturbances,” *Journal of Econometrics*, 157, 53–67.
- KELEJIAN, H. H., I. R. PRUCHA, AND Y. YUZEFOVICH (2006): “Estimation Problems in Models with Spatial Weighting Matrices which Have Blocks of Equal Elements,” *Journal of Regional Science*, 46(3), 507–515.
- KLINE, B., AND E. TAMER (2011): “Some Interpretation of the Linear-in-Means Model of Social Interactions,” Northwestern University Working Paper.
- KOLACZYK, E. (2009): *Statistical Analysis of Network Data*. Springer-Verlag.
- KOLLER, D., AND N. FRIEDMAN (2009): *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.
- KOPCZUK, W., A. ALSTADSAETER, AND K. TELLE (2014): “Social networks and tax avoidance: Evidence from a well-defined Norwegian tax shelter,” Columbia University Working Paper.
- KOSKINEN, J. H., G. ROBINS, AND P. PATTISON (2010): “Analysing Exponential Random Graph (p-star) Models with Missing Data using Bayesian Data Augmentation,” *Statistical Methodology*, 7, 366–384.
- KRANTON, R., AND D. MINEHART (2001): “A Theory of Buyer-Seller Networks,” *American Economic Review*, 91(3), 485–508.
- LAM, C., AND P. C. SOUZA (2013): “Regularization for High-Dimensional Spatial Models Using the Adaptive LASSO,” LSE Working Paper.
- (2014): “One-Step Regularized Spatial Weight Matrix and Fixed Effects Estimation with Instrumental Variables,” Working Paper LSE.
- LAVY, V., AND E. SAND (2015): “The Effect of Social Networks on Students’ Academic and Non-Cognitive Behavioral Outcomes: Evidence from Conditional Random Assignment of Friends in School,” University of Warwick and Hebrew University Working Paper.

- LAZZATI, N. (2015): “Treatment Response with Social Interactions: Partial Identification via Monotone Comparative Statics,” *Quantitative Economics*, 6, 49–83.
- LEE, L.-F. (2007): “Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects,” *Journal of Econometrics*, 140(2), 333–374.
- LEE, L.-F., X. LIU, AND X. LIN (2010): “Specification and Estimation of Social Interaction Models with Network Structures,” *Econometrics Journal*, 13, 145–176.
- LEE, L.-F., AND J. YU (2010): “Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects,” *Journal of Econometrics*, 154, 165–185.
- LEE, R. S., AND K. FONG (2011): “Markov-Perfect Network Formation: An Applied Framework for Bilateral Oligopoly and Bargaining in Buyer-Seller Networks,” NYU Working Paper.
- LEUNG, M. (2015a): “A Random-Field Approach to Inference in Large Network Models,” Stanford University Working Paper.
- (2015b): “Two-Step Estimation of Network Formation Models with Incomplete Information,” *Journal of Econometrics*, 188(1), 182–195.
- LIN, X., AND L.-F. LEE (2010): “GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity,” *Journal of Econometrics*, 157, 34–52.
- LIU, X., E. PATACCHINI, AND Y. ZENOU (2014): “Endogenous Peer Effects: Local Aggregate or Local Average,” *Journal of Economic Behavior and Organization*, 103, 39–59.
- LOVASZ, L. (2012): *Large Networks and Graph Limits*. American Mathematical Society.
- MANRESA, E. (2013): “Estimating the Structure of Social Interactions using Panel Data,” CEMFI Working Paper.
- MANSKI, C. F. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *The Review of Economic Studies*, 60(3), 531–542.
- (2013): “Identification of Treatment Response with Social Interactions,” *Econometrics Journal*, 16(1), S1–S23.
- MANSKI, C. F., AND J. MAYSHAR (2003): “Private Incentives and Social Interactions: Fertility Puzzles in Israel,” *Journal of the European Economic Association*, 1(1), 181–211.
- MASTEN, M. A. (2015): “Random Coefficients on Endogenous Variables in Simultaneous Equation Models,” Duke University Working Paper.
- MCADAM, D. (1986): “Recruitment to High-Risk Activism: The Case of Freedom Summer,” *American Journal of Sociology*, 92(1), 64–90.
- MELE, A. (2015): “A Structural Model of Segregation in Social Networks,” Johns Hopkins Working Paper.

- MENZEL, K. (2015): “Inference for Games with Many Players,” NYU Working Paper.
- MIYAUCHI, Y. (2014): “Structural Estimation of a Pairwise Stable Network with Nonnegative Externality,” MIT Working Paper.
- MOFFITT, R. (2001): “Policy Interventions, Low-Level Equilibria and Social Interactions,” in *Social Dynamics*, ed. by S. Durlauf, and P. Young. MIT Press.
- NAKAJIMA, R. (2007): “Measuring Peer Effects on Youth Smoking Behaviour,” *Review of Economic Studies*, 74(3), 897–935.
- NEIDELL, M., AND J. WALDFOGEL (2010): “Cognitive and Noncognitive Peer Effects in Early Education,” *Review of Economics and Statistics*, 92(3), 562–576.
- PENNOCK, D., G. FLAKE, S. LAWRENCE, E. GLOVE, AND C. GILES (2002): “Winners Don’t Take All: Characterizing the Competition for Links on the Web,” *Proceedings of the National Academy of Sciences*, 99(8), 5207–5211.
- POMERANZ, D. (forthcoming): “No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax,” *American Economic Review*.
- QU, X., AND L.-F. LEE (2015): “Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix,” *Journal of Econometrics*, 184(2), 209–232.
- RASCH, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press.
- RINALDO, A., S. E. FIENBERG, AND Y. ZHOU (2009): “On the Geometry of Discrete Exponential Families with Application to Exponential Random Graph Models,” *Electronic Journal of Statistics*, 3, 446–484.
- ROBINS, G., P. PATTISON, Y. KALISH, AND D. LUSHER (2007): “An Introduction to Exponential Random Graph (p^*) Models for Social Networks,” *Social Networks*, 29(2), 173–191.
- ROBINS, G., T. SNIJDERS, P. WANG, M. HANDCOCK, AND P. PATTISON (2007): “Recent Developments in Exponential Random Graph (p^*) Models for Social Networks,” *Social Networks*, 29(2), 192–215.
- ROSE, C. (2015): “Essays in Applied Microeconometrics,” Ph.D. thesis, University of Bristol.
- RUDIN, W. (1987): *Real and Complex Analysis*. McGraw-Hill.
- SACERDOTE, B. (2001): “Peer Effects with Random Assignment: Results for Dartmouth Roomates,” *Quarterly Journal of Economics*, 106(2), 681–704.
- (2010): “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, vol. 3. North-Holland.

- SHALIZI, C., AND A. RINALDO (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,” UCLA Working Paper.
- SIRAKAYA, S. (2006): “Recidivism and Social Interactions,” *Journal of the American Statistical Association*, 101(475), 863–877.
- SNIJDERS, T. (2002): “Markov Chain Monte Carlo Estimation of Exponential Random Graph Models,” *Journal of Social Structure*, 3(2), 1–40.
- SOUZA, P. C. (2014): “Estimating Network Effects without Network Data,” LSE Working Paper.
- STRAUSS, D., AND M. IKEDA (1990): “Pseudolikelihood Estimation for Social Networks,” *Journal of the American Statistical Association*, 85(409), 204–212.
- SUN, Y. (2006): “The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks,” *Journal of Economic Theory*, 126, 31–69.
- TAMER, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *The Review of Economic Studies*, 70(1), 147–165.
- TAO, J., AND L.-F. LEE (2014): “A Reaction Function Model with the First Order Statistic,” *Econometrics Journal*, 17(3), 197–240.
- TIBSHIRANI, R. (1996): “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society B (Methodological)*, 58(1), 267–288.
- TINCANI, M. (2015): “Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence,” UCL Working Paper.
- TOPA, G. (2001): “Social Interactions, Local Spillovers and Unemployment,” *Review of Economic Studies*, 68(2), 261–295.
- UHLIG, H. (1996): “A Law of Large Numbers for Large Economies,” *Economic Theory*, 8, 41–50.
- WASSERMAN, S., AND K. FAUST (1994): *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- WATTS, A. (2001): “A Dynamic Model of Network Formation,” *Games and Economic Behavior*, 34, 331–341.
- XU, X., AND L.-F. LEE (forthcoming): “Maximum Likelihood Estimation of a Spatial Autoregressive Tobit Model,” *Journal of Econometrics*.
- YAN, T., AND J. XU (2013): “A Central Limit theorem in the β -Model for Undirected Random Graphs with a Diverging Number of Vertices,” *Biometrika*, 100(2), 519–524.

ZENOU, Y. (2015): “Key Players,” in *Handbook on the Economics of Networks*, ed. by Y. Bramoullé, B. Rogers, and A. Galeotti. Oxford University Press.

ZHANG, C.-H. (2010): “Nearly Unbiased Variable Selection Under Minimax Concave Penalty,” *Annals of Statistics*, 38(2), 894–942.

ZHENG, T., M. J. SALGANIK, AND A. GELMAN (2006): “How Many People Do You Know in Prison?: Using Overdispersion in Count Data to Estimate Social Structure in Networks,” *Journal of the American Statistical Association*, 101(474), 409–423.

ZOU, H., AND T. HASTIE (2005): “Regularization and Variable Selection via the Elastic Net,” *Journal of the Royal Statistical Society B*, 67(2), 301–320.

A Proofs

A.1 Proof of Proposition 2

If $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I}$, then $\mathbb{V}(\mathbf{y}|\mathbf{x}) = \sigma^2(\mathbf{I} - \beta W)^{-2}$. Since $|\beta| < 1$ and W is (row-)stochastic, we obtain

$$(\mathbf{I} - \beta W)^{-1} = \mathbf{I} + \beta W + \beta^2 W^2 + \dots$$

It can be verified that $W^k, k = 1, 2, \dots$ is symmetric with diagonal elements $(W^k)_{ii} = a_{k-1}$ and off-diagonal entries $(W^k)_{ij} = a_k, i \neq j$, where

$$a_0 = 0, \quad a_{-1} = 1 \quad \text{and} \quad a_k = (a_{k-2} + a_{k-1}(N-2))/(N-1).$$

Let $S = \sum_{k=1}^{\infty} \beta^k W^k$. Then,

$$\begin{aligned} S_{ii} &= \sum_{k=1}^{\infty} \beta^k a_{k-1} = \sum_{\tilde{k}=0}^{\infty} \beta^{\tilde{k}+1} a_{\tilde{k}} = \sum_{\tilde{k}=1}^{\infty} \beta^{\tilde{k}+1} (a_{\tilde{k}-2} + a_{\tilde{k}-1}(N-2))/(N-1) \\ &= \frac{\beta^2}{N-1} \left(1 + \sum_{\tilde{k}=2}^{\infty} \beta^{\tilde{k}-1} a_{\tilde{k}-2} \right) + \frac{N-2}{N-1} \beta \sum_{\tilde{k}=1}^{\infty} \beta^{\tilde{k}} a_{\tilde{k}-1} \\ &= \frac{\beta^2}{N-1} \left(1 + \sum_{\tilde{k}=1}^{\infty} \beta^{\tilde{k}} a_{\tilde{k}-1} \right) + \frac{N-2}{N-1} \beta \sum_{\tilde{k}=1}^{\infty} \beta^{\tilde{k}} a_{\tilde{k}-1} \\ &= \frac{\beta^2}{N-1} (1 + S_{ii}) + \frac{N-2}{N-1} \beta S_{ii} \end{aligned}$$

where the second and fifth equalities set $\tilde{k} = k - 1$ and $\bar{k} = \tilde{k} - 1$ respectively, the third equality acknowledges that $a_0 = 0$ and uses the definition of a_k above, and the last equality uses the definition of S_{ii} . This implies that

$$S_{ii} = \frac{\beta^2}{(N-1) - (N-2)\beta - \beta^2}.$$

(The denominator is non-zero as long as $|\beta| < 1$.) On the other hand, $S_{ij} = \sum_{k=1}^{\infty} \beta^k a_k = \beta^{-1} S_{ii}$ (for $i \neq j$).

Since $(\mathbf{I} - \beta W)^{-1} = \mathbf{I} + \beta W + \beta^2 W^2 + \dots = \mathbf{I} + S$, its diagonal elements are then given by $1 + S_{ii} = [(N-1) - (N-2)\beta] / C$, and its off-diagonal entries are given by $\beta^{-1} S_{ii} = \beta / C$, where $C = (N-1) - (N-2)\beta - \beta^2$.

Using then the fact that $\mathbb{V}(\mathbf{y}|\mathbf{x}) = \sigma^2(\mathbf{I} + S)^2$, the ratio between covariance and variance among observable outcome variables is given by

$$\kappa(\beta, N) \equiv \frac{\mathbb{C}(y_i, y_j|\mathbf{x})}{\mathbb{V}(y_i|\mathbf{x})} = \frac{2\beta(N-1) - (N-2)\beta^2}{((N-1) - (N-2)\beta)^2 + (N-1)\beta^2}. \quad (12)$$

Notice that $\beta = 0 \Rightarrow \mathbb{C}(y_i, y_j|\mathbf{x}) = 0$, and the Cauchy-Schwarz inequality implies that $|\kappa(\beta, N)| \leq 1$. Furthermore, the denominator above is always positive and $\text{sgn}(\kappa(\beta, N)) = \text{sgn}(2\beta(N-1) - (N-2)\beta^2)$. If $N = 2$, the right-hand side in (12) becomes $2\beta/(1 + \beta^2)$, which is increasing in β and the equation above has a unique solution. I will thus focus on $N > 2$. In this case, the numerator is a quadratic, concave polynomial with two roots: $\beta_1 = 0$ and $\beta_2 = 1 + N/(N-2) > 1$. It is then negative for $\beta \in (-\infty, \beta_1) \cup (\beta_2, \infty)$ and positive for $\beta \in (\beta_1, \beta_2)$. Since $|\beta| < 1$, it is then straightforward to see that

$$\beta \geq 0 \Leftrightarrow \kappa(\beta, N) \geq 0.$$

Let $\bar{\kappa}$ be the observed covariance-variance ratio among outcome variables. The parameter β is then a solution to the quadratic equation $p(b; \bar{\kappa}, N) = 0$ obtained from (12), where $p(b; \bar{\kappa}, N)$ is given by

$$\{\bar{\kappa} [(N-2)^2 + (N-1)] + (N-2)\} b^2 - 2(N-1) [1 + \bar{\kappa}(N-2)] b + \bar{\kappa}(N-1)^2.$$

By the Fundamental Theorem of Algebra, there are at most two solutions to the equation

above. As noted earlier, $\text{sgn}(\bar{\kappa}) = \text{sgn}(\beta)$. There are then three cases to consider:

i) If $\bar{\kappa} = 0$, then $\beta = 0$ and the model is identified since the remaining parameters can then be obtained from the reduced-form coefficients.

ii) If $\bar{\kappa} > 0$, then $\beta > 0$, and the coefficient on the quadratic term in the equation above is positive and the polynomial $p(b; \bar{\kappa}, N)$ is convex in b . Furthermore, $p(0; \bar{\kappa}, N) = \bar{\kappa}(N-1)^2 > 0$ (since $\bar{\kappa} > 0$) and $p(1; \bar{\kappa}, N) = N(\bar{\kappa} - 1) \leq 0$ (since $\bar{\kappa} \leq 1$, see above). Together, these imply that both roots are greater than zero, that one of roots is greater than one and the other is less than one. Hence, only one root is positive and below one, and the model is identified as the remaining parameters again can be obtained from the reduced-form coefficients.

iii) If $\bar{\kappa} < 0$, then $\beta < 0$. Notice that the linear coefficient in $p(b; \bar{\kappa}, N)$ is negative when $N = 2$. For $N > 2$, it is positive if, and only if,

$$\bar{\kappa} < \frac{1}{2 - N} < 0.$$

Since $\bar{\kappa}$ is bounded below by $\kappa(-1, N)$ (see Proposition 3) and

$$\kappa(-1, N) - \frac{1}{2 - N} = \frac{N(N - 1)}{\{[(N - 1) + (N - 2)]^2 + (N - 1)\}(N - 2)} > 0,$$

the linear coefficient in $p(b; \bar{\kappa}, N)$ is always negative.

On the other hand, the quadratic coefficient in $p(b; \bar{\kappa}, N)$ is positive if, and only if,

$$\bar{\kappa} > \frac{-(N - 2)}{(N - 2)^2 + (N - 1)}.$$

For $N \geq 4$, it can be seen that

$$\kappa(-1, N) > \frac{-(N - 2)}{(N - 2)^2 + (N - 1)}.$$

In this case, the quadratic polynomial is convex, and its minimizer is positive (since the linear coefficient is negative, and the quadratic coefficient is positive). This implies that only one of its roots can be negative and the model is then identified.

If $N = 3$, the roots to $p(b; \bar{\kappa}, 3)$ are given by $2 \pm 2/\sqrt{1 + \bar{\kappa}}$. (The lower bound on $\bar{\kappa}$ is $\kappa(-1, 3) = -5/11 > -1$, so the denominator will always be non-zero.) Since $-1 < \bar{\kappa} < 0$, $2 - 2/\sqrt{1 + \bar{\kappa}} < 0 < 2 + 2/\sqrt{1 + \bar{\kappa}}$, and only one root is admissible. The model is identified as before. ■

A.2 Proof of Proposition 3

The derivative of $\kappa(\beta, N)$ defined in (12) with respect to β is given by

$$\begin{aligned} \frac{\partial \kappa(\beta, N)}{\partial \beta} &= \frac{(N-1)^2(N-1-\beta^2-\beta(N-2))}{\{((N-1)-(N-2)\beta)^2+(N-1)\beta^2\}^2} \\ &> \frac{(N-1)^2(N-2)(1-\beta)}{\{((N-1)-(N-2)\beta)^2+(N-1)\beta^2\}^2} \\ &\geq 0, \end{aligned}$$

where both inequalities use the assumption that $|\beta| < 1$. Hence, a lower bound for $\kappa(\beta, N) \equiv \mathbb{C}(y_i, y_j | \mathbf{x}) / \mathbb{V}(y_i | \mathbf{x})$ is given by $\kappa(-1, N) = (4 - 3N) / (4N^2 - 11N + 8)$. ■

A.3 Proof of Proposition 6

Because the measurement error is classical and independent across individuals,

$$\mathbb{C}(\tilde{y}_{i,g}, \tilde{y}_{j,g} | \mathbf{x}_g) = \mathbb{C}(y_{i,g}, y_{j,g} | \mathbf{x}_g).$$

From the proof of Proposition 2, we obtain

$$\mathbb{C}(y_{i,1}, y_{j,1} | \mathbf{x}_1) = 2\beta\sigma^2 / (1 - \beta^2)^2$$

for the group with two individuals and

$$\mathbb{C}(y_{i,2}, y_{j,2} | \mathbf{x}_1) = \beta(4 - \beta)\sigma^2 / [(1 - \beta) + (1 - \beta^2)]^2$$

for the group with three individuals. Then,

$$\psi(\beta) = \frac{\mathbb{C}(y_{i,1}, y_{j,1} | \mathbf{x}_1)}{\mathbb{C}(y_{i,2}, y_{j,2} | \mathbf{x}_1)} = \frac{8 + 8\beta + 2\beta^2}{4 + 7\beta + 2\beta^2 - \beta^3}.$$

It can be checked that $\psi'(\beta) < 0$ for $|\beta| < 1$ and consequently $\psi(\beta) > \psi(1) = 1.5$.

Let $\bar{\psi}$ be the observed covariance ratio among outcome variables. The parameter β is then a solution to a cubic equation $q(b; \bar{\psi}) = 0$ obtained from the expression above, where $q(b; \bar{\psi}) \equiv -\bar{\psi}\beta^3 + 2(\bar{\psi} - 1)\beta^2 + (7\bar{\psi} - 8)\beta + 4(\bar{\psi} - 2)$. To show that only one root to this equation is below 1 in absolute value, I make use of Rouché's theorem (see Rudin (1987), pp. 225 and 229). The result is stated for general complex-valued functions. In our context, it establishes that, if the functions f and g are continuous on a compact set C and differentiable on its interior with $|g(x)| < |f(x)|$ on the boundary of C , then f and $f + g$ have the same number of zeros in the interior of C , where each zero is counted as many times as its multiplicity. Taking $f(x) = a_1x$ (so that $f(x) = 0 \Rightarrow x = 0$) and $g(x) = a_0 + a_2x^2 + \dots + a_Kx^K$ and $C = [-1, 1]$, one obtains the following corollary:

If $|a_1| > |a_0| + |a_2| + \dots + |a_K|$, then there is exactly one root for the polynomial $a_0 + a_1x + \dots + a_Kx^K$ with absolute value less than 1.

When $\bar{\psi} \geq 2$, $7\bar{\psi} - 8 > |-\bar{\psi}| + |2(\bar{\psi} - 1)| + |4(\bar{\psi} - 2)| = 7\bar{\psi} - 10$. When $\bar{\psi} < 2$, we have that $|-\bar{\psi}| + |2(\bar{\psi} - 1)| + |4(\bar{\psi} - 2)| = 6 - \bar{\psi}$. Then, $7\bar{\psi} - 8 > 6 - \bar{\psi} \Leftrightarrow \bar{\psi} > 14/8$, which is true since $\bar{\psi} > 1.5$. This implies that there is only one solution with absolute value below one (i.e., β) and completes the proof. ■

B Connections Game

Here I examine the three-player connections game where $u_i(g) = \sum_{j \in \{1, \dots, n, j \neq i\}} \delta^{d(i,j;g)-1} (1 + \epsilon_{ij}) - |N_i(g)|$. The set of possible links is $\{12, 13, 23\}$, and there are 8 possible networks (2^3). However, there are only four distinct topologies, which can be characterized by the number of links (0 to 3). These networks and respective payoffs are characterized below:

(i) One network with 0 links ($g = \emptyset$): $u_i = 0$ for each player. For this network to be pairwise stable, one needs $\epsilon_{ij} < 0$ or $\epsilon_{ji} < 0$ for every pair ij .

(ii) Three networks with 1 link ($g = \{ij\}$): $u_i = \epsilon_{ij}$ and $u_j = \epsilon_{ji}$ for the two connected players, $u_k = 0$ for the isolated player (3 possible networks with distinct isolated players). For this network to be pairwise stable, one needs $\epsilon_{ij} \geq 0$ and $\epsilon_{ji} \geq 0$ (for the ij link); $\epsilon_{ik} < 0$ or $\epsilon_{ki} < 0$ (for the absent ik link); and $\epsilon_{jk} < 0$ or $\epsilon_{kj} < 0$ (for the absent jk link).

(iii) Three networks with 2 links ($g = \{ij, jk\}$): $u_i = \epsilon_{ij} + \delta(1 + \epsilon_{ik})$, $u_k = \epsilon_{kj} + \delta(1 + \epsilon_{ki})$ and $u_j = \epsilon_{ji} + \epsilon_{jk}$ (3 possible networks with different middle players). For this network to be pairwise stable, one needs $\epsilon_{ij} + \delta(1 + \epsilon_{ik}) \geq 0$ and $\epsilon_{ji} \geq 0$ (for the ij link); $\epsilon_{kj} + \delta(1 + \epsilon_{ki}) \geq 0$ and $\epsilon_{jk} \geq 0$ (for the kj link); and $\epsilon_{ik} < \delta(1 + \epsilon_{ik})$ or $\epsilon_{ki} < \delta(1 + \epsilon_{ki})$ (for the absent ik link).

(iv) One network with 3 links ($g = \{12, 13, 23\}$): $u_i = \epsilon_{ij} + \epsilon_{ik}$ and similar expressions for the other players. For this network to be pairwise stable, one needs $\epsilon_{ij} \geq \delta(1 + \epsilon_{ij})$ and $\epsilon_{ji} \geq \delta(1 + \epsilon_{ji})$ for every ij pair.

If $0 < \epsilon_{23} < \delta/(1 - \delta)$ and $\epsilon_{ij} = \epsilon_{ji}$ for every pair, we can establish the conditions on ϵ_{12} and ϵ_{13} for each of the 2^3 possible graphs to be pairwise stable:

\emptyset :	not pairwise stable
$\{12\}$:	$\epsilon_{12} \geq 0; \epsilon_{13} \leq 0; \epsilon_{13} < \epsilon_{23}/\delta - 1$
$\{13\}$:	$\epsilon_{13} \geq 0; \epsilon_{12} \leq 0; \epsilon_{12} < \epsilon_{23}/\delta - 1$
$\{23\}$:	$\epsilon_{12} < 0, \epsilon_{13} < 0$
$\{12, 13\}$:	$\epsilon_{12}, \epsilon_{13} \geq -\delta(1 + \epsilon_{23}); \epsilon_{13}, \epsilon_{12} > 0; \epsilon_{23} < \delta/(1 - \delta)$
$\{12, 23\}$:	$\epsilon_{12}, \epsilon_{23} \geq -\delta(1 + \epsilon_{13}); \epsilon_{12}, \epsilon_{23} > 0; \epsilon_{13} < \delta/(1 - \delta)$
$\{13, 23\}$:	$\epsilon_{13}, \epsilon_{23} \geq -\delta(1 + \epsilon_{12}); \epsilon_{13}, \epsilon_{23} > 0; \epsilon_{12} < \delta/(1 - \delta)$
$\{12, 13, 23\}$:	not pairwise stable