A Structural Model of Dense Network Formation*

Angelo Mele†

March 9, 2016

Abstract

This paper proposes an empirical model of network formation, combining strategic and random networks features. Payoffs depend on direct links, but also link externalities. Players meet sequentially at random, myopically updating their links. Under mild assumptions, the network formation process is a potential game and converges to an exponential random graph model (ERGM), generating directed dense networks. I provide new identification results for ERGMs in large networks: if link externalities are non-negative, the ERGM is asymptotically indistinguishable from an Erdos-Renyi model with independent links. We can identify the parameters only when at least one of the externalities is negative and sufficiently large. However, the standard estimation methods for ERGMs can have exponentially slow convergence, even when the model has asymptotically independent links. I thus estimate parameters using a Bayesian MCMC method. When the parameters are identifiable, I show evidence that the estimation algorithm converges in almost quadratic time.

JEL Codes: D85, C15, C73
Keywords: networks, Bayesian estimation, Markov Chain Monte Carlo, graph Limits, large deviations

*I am grateful to Roger Koenker, Ron Laschever, Dan Bernhardt and George Deltas for their inputs at crucial stages of this work. I am grateful to the editor and five outstanding referees that provided invaluable suggestions and comments to previous versions. Thanks to Lingjiong Zhu for his inputs about graph limits and large deviations. I thank Federico Bandi, Alberto Bisin, Stephane Bonhomme, Ethan Cole, Aureo de Paula, Steven Durlauf, Andrea Galeotti, Shweta Gaonkar, Sanjeev Goyal, Dan Karney, Darren Lubotsky, Antonio Mele, Luca Merlino, Ferdinando Monte, Tom Parker, Dennis O'Dea, Micah Pollak, Sergey Popov, Sudipta Sarangi, Giorgio Topa, Antonella Tutino and seminar participants at many seminars and conferences for helpful comments and suggestions. I am grateful to Andriy Norets and Matt Jackson for suggesting references on how to prove and bound convergence of the algorithms. Financial support from the R. Ferber Award, the R. W. Harbeson Memorial Dissertation Fellowship, and the NET Institute Summer Research Grant 2010 is gratefully acknowledged. All remaining errors are mine.

†*Address*: Johns Hopkins University - Carey Business School, 100 International Drive, Baltimore, MD 21202. *Email*: angelo.mele@jhu.edu

1 Introduction

Social networks are important determinants of individuals' socioeconomic performance. An increasing amount of evidence shows that the number and composition of social ties affects employment prospects, school performance, risky behavior, adoption of new technologies, diffusion of information and health outcomes.[1]

The literature on strategic models of network formation provides a framework to interpret the observed network as the equilibrium of a game.[2] However, the estimation and identification of strategic models is challenging. First, network formation models usually have multiple equilibria, because linking generates externalities that are not fully accounted for by individuals. In addition, the number of possible network configurations increases exponentially with the number of players, creating a curse of dimensionality. Second, in most empirical applications the econometrician has access to data on a single network snapshot. While a network may contain a large number of links, these are highly correlated because of the strategic and interdependent decisions of players. Therefore it is necessary to develop non-standard inference and asymptotics for such class of models.[3]

I propose a model of network formation that combines features from the strategic and random network formation literature.[4] Players' utilities depend on payoffs from direct links, but also link externalities (e.g. reciprocity, indirect friends, popularity, etc). The network formation is dynamic: in each period a player meets another agent and decides whether to form a new link, keep an existing link or do nothing. This process generates a sequence of directed dense graphs.[5]

The paper contributes to the economic literature on empirical network models by establishing several results. First, under mild restrictions on the preferences, the network formation process is a potential game: there exists a potential function that summarizes all the incentives of the players in any state of the network.[6]

Second, I prove that the model converges to a unique stationary equilibrium distribution over networks, i.e. the likelihood of observing a specific network realization in the long-run. Assuming that the observed network data is a draw from the stationary distribution, the structural parameters can be estimated using only one network observation. Furthermore, the likelihood is identical to the exponential random graph model (ERGM), a popular empirical model used by social scientists and statisticians in applications.[7]

[1] For example, see the contributions of Topa (2001); Laschever (2009); Cooley (2010); De Giorgi et al. (2010); Nakajima (2007); Bandiera and Rasul (2006); Conley and Udry (forthcoming); Golub and Jackson (2011); Acemoglu et al. (2011).

[2] See Jackson (2008), Jackson and Wolinsky (1996), Bala and Goyal (2000), Currarini et al. (2009), Currarini et al. (2010), De Marti and Zenou (2009), Echenique et al. (2006) for examples.

[3] Several authors have recently contributed to this problem. See Chandrasekhar and Jackson (2014), Leung (2014*a*), DePaula et al. (2014), Menzel (2015) for details.

[4] See Jackson (2008) for a review of network formation models.

[5] A dense graph is such that the number of links scales quadratically with the number of players.

[6] See Monderer and Shapley (1996) for an analytis of potential games.

[7] Snijders (2002) is a good introduction to the ERGMs. See also Koskinen (2008), Caimo and Friel (2010),

Third, I establish new asymptotic results for the class of directed ERGM models with homogeneous players, using a mix of graph limits, large deviations and variational methods for the exponential family.[8] When the number of players becomes large and *when all the link externalities are non-negative*, the model is asymptotically indistinguishable from a directed Erdos-Renyi graph. As a consequence the links are asynmptotically independent and the externalities are not identified in this region of the parameter space. On the other hand, *when at least one of the link externalities is negative and sufficiently large,* the model does not converge asymptotically to a directed Erdos-Renyi graph and the link externalities can be identified.

Fourth, I show that the standard estimation algorithm used by the ERGM practitioners has convergence problems.[9] The likelihood of the model depends on an intractable normalizing constant, that cannot be computed exactly because of the curse of dimensionality. The ERGM literature proposes to approximate the normalizing constant by simulations, using a *local* Markov Chain Monte Carlo sampler. I extend the techniques of Bhamidi et al. (2011) to prove that even in the simplest case of non-negative externalities, when the model has asymptotically independent links, the local MCMC algorithm has exponentially slow convergence for a significant portion of the parameter space, making estimation impractical in many cases of interest. Our identification results provide an explanation for such poor performance of the sampler: in such regions of the parameters, the likelihood is bimodal and the sampler may spend exponentially long time in a local maxima.

Finally, the parameters of the model are estimated using an approximate exchange algorithm (Murray et al. (2006)), with artificial data and medium size networks.[10] I use a double Metropolis-Hastings step to sample from the parameter space and show that the sampler is ergodic and provides samples from the correct posterior distribution. In the estimation exercise I focus on the region of parameters with negative and sufficiently large externalities, where the model's parameters can be identified.[11] I show by simulations that in finite networks one may encounter additional computational problems. While the estimates are precise for a large region of the parameter space, the estimated posterior becomes extremely imprecise for very large negative externalities. For such parameters, the sufficient statistics corresponding to the externalities hit their lower bound (zero), and the output of the simulation is irregular and skewed, making precise estimation impossible.

Chandrasekhar and Jackson (2014).

[8] The use of these techniques is relatively recent. See Diaconis and Chatterjee (2011), Chatterjee and Varadhan (2011), Radin and Yin (2013), Aristoff and Zhu (2014) for several contributions in applied probability. Lovasz (2012) is a good and extensive summary of the literature on graph limits. Most of the literature focuses on undirected graphs, and our extension to directed networks is non-trivial. See Appendix D for details.

[9] See Snijders (2002) for a summary of the simulation methods used in the ERGM literature.

[10] I use networks of 100 or 200 nodes, comparable for example to the school friendship networks in Add Health.

[11] I report additional simulations in Appendix E and there are more results in previous version of this paper.

This work contributes to the economic literature on empirical models of network formation in several dimensions. The challenges that lead to multiple equilibria and the curse of dimensionality have been addressed in different ways, e.g. modeling the network formation as a sequential process (Christakis et al. (2010)), restricting the type of externalities considered (Miyauchi (2012), or using subnetworks as the unit of analysis (Sheng (2012), Chandrasekhar and Jackson (2014)). Others have focused on the observable implications of homophily (Boucher (2013)) or modeled the network formation as a game with imperfect information (Leung (2014*b*)). My model considers a sequential network formation process with complete information and restricts the preferences to guarantee the existence of a potential function. While the characterization using potential games has been considered in previous work (Jackson and Watts (2001), Gilles and Sarangi (2004), Butts (2009)), I show that this modeling strategy reduces the computational complexity of the simulations, because allows us to simulate changes in the potential levels, without keeping track of all the players.

The closest work is Christakis et al. (2010). In their model myopic players meet sequentially and choose which links to form by maximizing current utility. The sequence of meetings is unobservable, and therefore must be integrated out in the likelihood. This computational challenge is addressed with an MCMC scheme that samples from the space of meeting sequences. To limit the computational burden they assume that individuals can meet only once, and linking decisions are permanent. My model is similar in spirit, but I make assumptions on the meeting technology that guarantee existence of a closed form solution for the stationary equilibrium distribution of networks. Players meet often and have the opportunity to revise their links frequently. In addition, I provide a complete characterizations of the strategic equilibrium, the convergence properties of the estimation algorithms and I use graph limits to establish several identification results.

Modeling the network formation externalities jointly with unobserved heterogeneity is challenging. Indeed, Graham (2014) provides frequentist inference for a model with unobserved heterogeneity, but rules out the network formation externalities that are crucial in our model. I abstract from unobserved heterogeneity, which can be included in our model with substantial additional computational effort. However, it is not clear whether it is possible to separately identify unobserved heterogeneity from externalities using a single observation of the network (Graham (2014)).

The literature considers identification in two settings. In the *many networks* asymptotics, the researcher observes multiple networks (Miyauchi (2012), Sheng (2012), Badev (2013)). In the *large network* asymptotics the econometrician observes only one single network, perhaps large (Chandrasekhar and Jackson (2014), Graham (2014), Leung (2014*a*), DePaula et al. (2014), Menzel (2015)). My model is identified in the many networks framework under usual regularity conditions, because the likelihood belongs to the exponential family.[12]. The case of large networks is more complicated and non-standard. I combine tools from the graph limits literature (Diaconis and Chatterjee (2011), Lovasz (2012), Radin and Yin (2013)), large deviations (Chatterjee and Varadhan (2011)) and variational methods (Wainwright and Jordan (2008)) for the exponential family to characterize the behavior of the model in

[12]See Lehman (1983)

large networks. This allows me to make substantial progress on the identification of structural parameters. Some of these techniques can be used to extend the results to the case of heterogenous players. However, these extensions face additional technical complications that are beyond the scope of this paper (Mele and Zhu (2015)).

The model presented here generates a *dense* network, i.e. the probability of linking does not converge to zero as the number of players grows large (Diaconis and Chatterjee (2011), Lovasz (2012), Graham (2014)). Chandrasekhar and Jackson (2014) show that when we impose sparsity, estimation of structural parameters is simpler in many specifications. De-Paula et al. (2014) and Menzel (2015) show that sparsity is crucial for identification. In this model, I can impose a certain degree of sparsity by forcing a link externality to be negative: I show that such model does not converge to an independent links model, thus allowing identification of the link externalities. I also show that in finite networks too much sparsity may generate computational problems if it implies that a sufficient statistic of the network is equal to zero: in such case precise estimation is impossible. Badev (2013) extends our model to include both binary actions and network formation, with an application to smoking among teenagers. Hsieh and Lee (2012) and Goldsmith-Pinkham and Imbens (2013) consider similar models.

2 A Model of Network Formation

2.1 Setup

Let $\mathcal{I} = \{1, 2, ..., n\}$ be the set of agents, each identified by a vector of A (exogenous) characteristics $X_i = \{X_{i1}, ..., X_{iA}\}$, e.g. gender, wealth, age, location, etc. Let the matrix $X = \{X_1, X_2, ..., X_n\}$ collect the vectors of characteristics for the population and let $\mathcal{X}$ denote the set of all possible matrices X. Time is discrete.

The social network is represented as a $n \times n$ binary matrix $G \in \mathcal{G}$, where $\mathcal{G}$ is the set of all $n \times n$ binary matrices. The entry g_{ij} is equal to 1 if individual i forms a connection to individual j, and 0 otherwise; by convention $g_{ii} = 0$, for any i. The network G is *directed*, i.e. $g_{ij} = 1$ does not necessarily imply $g_{ji} = 1$.[13]

Let the *realization* of the network at time t be denoted as g^t and the *realization* of the link between i and j at time t be g^t_{ij}. The network including all the current links but g^t_{ij}, i.e. $g^t \backslash g^t_{ij}$, is denoted as g^t_{-ij}; while g^t_{-i} denotes the network matrix excluding the i-th row (i.e. all the links of player i).

[13] The assumption of directed networks is not crucial to many of the results.

2.1.1 Preferences

The utility of player i from a network g and population attributes $X = (X_1, ..., X_n)$ at parameter $\theta = (\theta_u, \theta_m, \theta_v, \theta_w)$ is given by

$$U_i(g, X; \theta) = \underbrace{\sum_{j=1}^{n} g_{ij} u_{ij}^{\theta_u}}_{\text{direct links}} + \underbrace{\sum_{j=1}^{n} g_{ij} g_{ji} m_{ij}^{\theta_m}}_{\text{mutual links}} + \underbrace{\sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{jk} v_{ik}^{\theta_v}}_{\text{indirect links}} + \underbrace{\sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{ki} w_{kj}^{\theta_w}}_{\text{popularity}} \quad (1)$$

where $u_{ij}^{\theta_u} \equiv u(X_i, X_j; \theta_u)$, $m_{ij}^{\theta_m} \equiv m(X_i, X_j; \theta_m)$, $v_{ij}^{\theta_v} \equiv v(X_i, X_j; \theta_v)$ and $w_{ij}^{\theta_w} \equiv w(X_i, X_j; \theta_w)$ are (bounded) real-valued functions of the attributes. The utility of the network is the sum of the net benefits received from each link. The total benefit from an *additional link* has four components.

When player i creates a link to agent j, he receives a *direct* net benefit $u_{ij}^{\theta_u}$ that includes both costs and benefits from the relationship. The net benefit can possibly be negative, e.g. when only homophily enters payoffs of direct links, the net utility $u_{ij}^{\theta_u}$ is positive if i and j belong to the same group, while it is negative when they are of different types.

Players value linking externalities, i.e. links formed by other players. A player receives additional utility $m_{ij}^{\theta_m}$ if the link is mutual; a connection has different value when the other party reciprocates.

Players value the composition of indirect connections. When i is deciding whether to create a link to j, she observes j's connections and their socioeconomic characteristics. Each of j's links provides additional utility $v(X_i, X_k; \theta_v)$ to i. To be concrete, suppose there are only two types: A and B. In this model, an agent who has the opportunity to form an additional link, values a type-A individual with three links to type-B agents as a different good than a type-A individual with two type-A connections and one type-B connection.[14] In other words, individuals value both *exogenous* heterogeneity and *endogenous* heterogeneity: the former is determined by the socioeconomic characteristics of the agents, while the latter arises endogenously with the process of network formation. In the baseline version of the model I assume that only indirect links are valuable and they are perfect substitutes: individuals do not receive utility from two-links-away contacts.[15]

The fourth component corresponds to a *popularity effect*. If individual i forms a link to j, he automatically creates an indirect link for all the agents that already have a link to i. Thus i generates an externality (positive or negative) for each k that formed a link to him in previous periods. This externality makes i more or less popular.

I impose an additional assumption on the functional forms of the utility components, which provides important equilibrium restrictions. I assume that the utility $m_{ij}^{\theta_m}$ obtained

[14] A similar assumption is used in De Marti and Zenou (2009) where the agents' cost of linking depends on the racial composition of friends of friends. Their model is an extension of the connection model of Jackson and Wolinsky (1996), and the links are formed with mutual consent. The corresponding network is undirected.

[15] This benchmark model can be extended to incorporate additional utility components, as shown below.

from mutual links is symmetric, and that the utility of an indirect link $v_{ij}^{\theta_v}$ has the same functional form as the utility from the popularity effect $w_{ij}^{\theta_v}$.

ASSUMPTION 1 ***(Preferences)*** *The preferences satisfy the following restrictions*

$$\begin{aligned} m\left(X_i, X_j; \theta_m\right) &= m\left(X_j, X_i; \theta_m\right) \text{ for all } i, j \in \mathcal{I} \\ w\left(X_k, X_j; \theta_v\right) &= v\left(X_k, X_j; \theta_v\right) \text{ for all } k, j \in \mathcal{I} \end{aligned}$$

The symmetry in $m_{ij}(\theta_m)$ does not imply that a mutual link between i and j gives both the same utility. If i and j have a mutual link, they receive the same common utility component $(m_{ij}(\theta_m))$ but they may receive different payoffs from direct or indirect links. Two individuals with the same exogenous characteristics $X_i = X_j$ who form a mutual link receive the same $u_{ij}(\theta_u)$ and $m_{ij}(\theta_m)$, but they may have different payoffs from the additional link because of the composition of their indirect contacts and their popularity. Therefore, the first part of the assumption is necessary for identification of the utility from indirect links and popularity.

The second part of the assumption imposes an identifying restriction to the externality generated by i when creating a link to j: any individual k that has formed a link to i, has an additional indirect contact, i.e. j, who agent k values by an amount $w\left(X_k, X_j; \theta_w\right)$. When $w\left(X_k, X_j; \theta_v\right) = v\left(X_k, X_j; \theta_v\right)$, an individual i values his popularity effect as much as k values the indirect link to j, i.e., i internalizes the externality he creates.

Assumption 1[16] is the main ingredient that allows me to characterize the network formation as a potential game (see also Butts (2009) and Chandrasekhar and Jackson (2014) for similar characterizations).

PROPOSITION 1 ***(Existence of a Potential Function)*** *Under Assumption 1, the deterministic component of the incentives of any player in any state of the network are summarized by a* ***potential function****, $Q : \mathcal{G} \times \mathcal{X} \to \mathbb{R}$*

$$Q\left(g, X; \theta\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}u_{ij}(\theta_u) + \sum_{i=1}^{n}\sum_{j>i}^{n} g_{ij}g_{ji}m_{ij}(\theta_m) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\sum_{\substack{k=1 \\ k\neq i,j}}^{n} g_{ij}g_{jk}v_{ik}(\theta_v), \qquad (2)$$

and the network formation game is a Potential Game.

Proof. See Appendix A ■

[16]The first part of the assumption is a normalization of the utility function that allows identification for the utility of indirect links and popularity. The second part of the assumption is an identification restriction, that guarantees the model's coherency in the sense of Tamer (2003). In simple words, this part of the assumption guarantees that the system of conditional linking probabilities implied by the model generates a proper joint distribution of the network matrix.Similar restrictions are also encountered in spatial econometrics models (Besag, 1974) and in the literature on qualitative response models (Heckman, 1978; Amemiya, 1981)

The intuition for the result is simple.[17] Under the restrictions of Assumption 1, for any player i and any link g_{ij} we have

$$Q\left(g_{ij}, g_{-ij}, X; \theta\right) - Q\left(1 - g_{ij}, g_{-ij}, X; \theta\right) = U_i\left(g_{ij}, g_{-ij}, X; \theta\right) - U_i\left(1 - g_{ij}, g_{-ij}, X; \theta\right)$$

Consider two networks, $g = (g_{ij}, g_{-ij})$ and $g' = (1 - g_{ij}, g_{-ij})$, that differ only with respect to one link, g_{ij}, chosen by individual i: the difference in utility that agent i receives from the two networks, $U_i\left(g, X; \theta\right) - U_i\left(g', X; \theta\right)$, is exactly equal to the difference of the *potential* function evaluated at the two networks, $Q\left(g, X; \theta\right) - Q\left(g', X; \theta\right)$. That is, the potential is an aggregate function that summarizes both the state of the network and the deterministic incentives of the players in each state.

Characterizing the network formation as a potential game facilitates the analysis and the simulations. To compute the equilibria of the model, there is no need to keep track of each player's behavior: the potential function contains all the relevant information.[18]

2.1.2 Network Formation Process

The process of network formation follows a *stochastic best-response dynamics* (Blume (1993)), generating a Markov chain of networks. The main ingredients of this process are random meetings and utility maximization. The implicit assumption is that meetings are very frequent, and players can revise their linking strategies often.

Meeting Technology. At the beginning of each period a player i is randomly selected from the population, and he meets individual j, according to a meeting technology. The *meeting process* is a stochastic sequence $m = \{m^t\}_{t=1}^{\infty}$ with support $\mathcal{I} \times \mathcal{I}$. The realizations of the meeting process are ordered pairs $m^t = \{i, j\}$, indicating which agent i should play and which link g_{ij} can be updated at period t.[19]

The probability that player i is randomly chosen from the population and meets agent j is defined as

$$\Pr\left(m^t = ij | g^{t-1}, X\right) = \rho\left(g^{t-1}, X_i, X_j\right) \tag{3}$$

[17] See Monderer and Shapley (1996) for definitions and properties of potential games.

[18] This property is key for the analysis of networks with many players: the usual check for existence of profitable deviations from the Nash equilibrium can be performed using the potential, instead of checking each player's possible deviation in sequence. The computation of all profitable deviations for each player involves $n(n-1)2^{n(n-1)}$ operations: each player has $n-1$ possible deviations, there are n players and a total of $2^{n(n-1)}$ possible network configurations. As it is shown below (Proposition 2), when the game is a potential game, the computation of all Nash equilibria is equivalent to finding the local maxima of the potential function. This corresponds to evalutating the potential function for all the $2^{n(n-1)}$ possible network structures. The latter task involves fewer operations by a factor of $n(n-1)$, thus decreasing the computational burden.

[19] Several models incorporate a meeting technology in the network formation process. Jackson and Watts (2002) assume individuals meet randomly according to a discrete uniform distribution. Currarini et al. (2009) introduce a matching process that is biased towards individuals of the same type. Christakis et al. (2010) develop a dynamic model, where the sequence of meetings determines which players have the opportunity to form a link in each period.

where $\sum_{i=1}^{n}\sum_{j=1}^{n}\rho\left(g,X_i,X_j\right)=1$ for any $g\in\mathcal{G}$. The meeting probability depends on the current network g (e.g. the existence of a common link between i and j) and the characteristics of the pair. This general formulation includes meeting technologies with a bias for same-type individuals as in Currarini et al. (2009). The simplest example of meeting technology is an i.i.d. discrete uniform process with $\rho\left(g^{t-1},X_i,X_j\right)=\frac{1}{n(n-1)}$. An example with bias for same-type agents is $\rho\left(g^{t-1},X_i,X_j\right)\propto\exp\left[-d\left(X_i,X_j\right)\right]$, where $d\left(\cdot,\cdot\right)$ is a distance function.

To analyze the long run behavior of the model, I impose more structure on the meeting technology.[20]

ASSUMPTION 2 ***(Meeting Process)*** *The meeting probability between i and j does not depend on the existence of a link between them, and each meeting has a positive probability of occurring, i.e. $\rho(g^{t-1},X_i,X_j)=\rho(g_{-ij}^{t-1},X_i,X_j)>0$ for any $ij\in\mathcal{I}\times\mathcal{I}$*

The meeting process is such that any player can be chosen and any pair of agents can meet. This assumption guarantees that any equilibrium network can be reached with positive probability. For example, a discrete uniform distribution satisfies this assumption. The other restriction is for identification purposes: if we allow ρ to depend on the current link between i and j, we cannot write the likelihood in closed form. Using data from a single network observation it is impossible to identify the function ρ unless we make very restricting assumptions.

Utility Maximization. Conditional on the meeting $m^t=ij$, player i updates the link g_{ij} to maximize his current utility, taking the existing network g_{-ij}^t as given. I assume that the agents do not take into account the effect of their linking strategy on the future evolution of the network. The players have *complete information*, since they can observe the entire network and the individual attributes of all agents.[21] Before updating his link to j, individual i receives an idiosyncratic shock $\varepsilon\sim F\left(\varepsilon\right)$ to his preferences that the econometrician cannot observe. This shock models unobservables that could influence the utility of an additional link. Player i links agent j at time t if and only if it is a best response to the current network configuration, i.e. $g_{ij}^t=1$ if and only if

$$U_i\left(g_{ij}^t=1,g_{-ij}^{t-1},X;\theta\right)+\varepsilon_{1t}\geq U_i\left(g_{ij}^t=0,g_{-ij}^{t-1},X;\theta\right)+\varepsilon_{0t}. \tag{4}$$

I assume that when the equality holds, the agent plays the status quo.[22] The network formation process generates a Markov chain of networks, with transition probabilities determined by the meeting process and agents' linking choices.

[20] Christakis et al. (2010) assume that individuals can meet only once and their links remain in place forever. Their assumption is convenient when estimating a large network, but it does not allow the characterization of the stationary equilibrium.

[21] More precisely, to make a decision about linking, the player needs to observe his in-links and the out-links of his friends.

[22] This assumption does not affect the main result and is relevant only when the distribution of the preference shocks is discrete.

The following standard parametric assumption on the shocks allows me to characterize the stationary distribution and transition probabilities.

ASSUMPTION 3 ***(Idiosyncratic Shocks)*** *The shock follows a Type I extreme value distribution, i.i.d. among links and across time.*

2.2 Equilibrium Analysis

A Nash equilibrium is a network in which any player has no profitable deviations from his current linking strategy, when randomly selected from the population. We can show that the set of Nash networks corresponds to the local maxima of the potential function. Suppose that the current network is a Nash network. As a consequence, if a player deviates from the current linking strategy, he receives less utility.[23] Since the change in utility for any agent is equivalent to the change in potential, any deviation from the Nash network must decrease the potential. It follows that the Nash network must be a local maximizer of the potential function over the set of networks that differ from the current network for at most one link.

In the absence of preference shocks, the consequences of assumptions 1 and 2 are that the model will evolve according to a Markov Chain, converging to one of the Nash networks with probability one (see formal details in Appendix A). Suppose a player is drawn from the meeting process. Such agent will play a best response to the current network configuration. Therefore, his utility cannot decrease. This holds for any player and any period. It follows that the potential is nondecreasing over time. Since there is a finite number of possible networks, in the long run, the sequence of networks must reach a local maximum of the potential, i.e., a Nash equilibrium.

Under Assumptions 1-3, the network evolves as a Markov chain with transition probabilities given by the conditional choice probabilities and the probability law of the meeting process m^t. One can easily show that the sequence $[g^0, g^1, ..., g^t]$ is *irreducible* and *aperiodic*.[24] The following theorem summarizes the main theoretical result.

THEOREM 1 ***(Uniqueness and Characterization of Stationary Equilibrium)*** *The network formation game, under Assumptions 1-3, converges to a unique stationary distribution* $\pi(g, X; \theta)$

$$\pi\left(g, X; \theta\right) = \frac{\exp\left[Q\left(g, X; \theta\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X; \theta\right)\right]}, \tag{5}$$

where $Q\left(g, X; \theta\right)$ *is the potential function (2).*

[23] When the utility from the equilibrium and the deviation is the same, the agent plays the status quo, i.e., the Nash strategy.

[24] Intuitively, since the meeting probability $\Pr\left(m^t = ij\right) > 0$ for all ij, there is always a positive probability of reaching a new network in which the link g_{ij} can be updated. The logistic shock assumption implies that there is always a positive probability of switching to another state of the network, thus eliminating absorbing states.

Proof. In Appendix A ■

The first part of the proposition follows directly from the irreducibility and aperiodicity of the Markov process generated by the network formation game. The uniqueness of the stationary distribution is crucial in estimation, since one does not need to worry about multiple equilibria. Furthermore, the stationary equilibrium characterizes the likelihood of observing a specific network configuration in the data. As a consequence, I can estimate the structural parameters from observations of only *one network at a specific point in time*, under the assumption that the observed network is drawn from the stationary equilibrium.

The second part of the proposition provides a closed-form solution for the stationary distribution. The latter can be interpreted as the probability of observing a specific network structure, when the network is observed in the long run. In the long run, the system of interacting agents will visit more often those states/networks that have high potential. Therefore a high proportion of the possible networks generated by the network formation game, will correspond to Nash networks.

The likelihood of the model belongs to the exponential family and coincides with an Exponential Random Graph Model (ERGM): the latter is a statistical model of network formation, with complex dependencies among links. The ERGM class of models posits that the probability of observing a specific network is proportional to an exponential function of a linear combination of network statistics. Exponential random graphs have been successfully used to fit social network data, providing a useful benchmark for alternative models.[25]

COROLLARY 1 ***(Exponential Random Graphs)***
Let Assumptions 1-3 hold. If the utility functions are linear in parameters, the stationary distribution $\pi(g, X; \theta)$ *describes an* ***exponential random graph***

$$\pi(g, X; \theta) = \frac{\exp\left[\theta' \mathbf{t}(g, X)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[\theta' \mathbf{t}(\omega, X)\right]}, \tag{6}$$

where $\theta = (\theta_u, \theta_m, \theta_v)'$ *is a (column) vector of parameters and* $\mathbf{t}(g, X)$ *is a (column) vector of canonical statistics.*

Proof. See Appendix A ■

The vector $\mathbf{t}(g, X) = (t_1(g, X), ..., t_K(g, X))$ is a vector of sufficient statistics for the network formation model. This vector may include the number of links, the number of whites-to-whites links, the number of male-to-female links and so on.

[25] Frank and Strauss (1986) developed the theory of Markov random graphs. These are models of random network formation in which there is dependence among links: the probability that a link occurs depends on the existence of other links. Wasserman and Pattison (1996) generalized the Markov random graphs to more general dependencies, developing the Exponential Random graph models. Snijders (2002) reviews these models and the related estimation techniques.

We can interpret *some specifications* of ERGMs as the stationary equilibrium of a strategic game of network formation, where myopic agents follow a stochastic best response dynamics and utilities are linear functions of the parameters.

2.3 Extensions and discussion

Additional utility components. It is possible to modify the baseline utility function (1) to include additional components. For example, one may be interested in studying preferences that include utility from cyclic triangles effects, i.e. individual i links to j, j connects to k and k links to i. The latter can be modeled as a component of the utility τ that varies with the characteristics of the three players involved in the relationships, i.e. $\tau(X_i, X_j, X_k; \theta_\tau)$ for all $i, j, k \in \mathcal{I}$. The utility is easily modified by including a term $\sum_{j=1}^{n} g_{ij} \sum_{k \neq i,j} g_{jk} g_{ki} \tau_{ijk}(\theta_\tau)$. However, to guarantee the existence of a potential function, we need to restrict τ in analogous way as in Assumption 1: the function τ must satisfy $\tau_{ijk}(\theta_\tau) = \tau_{i'j'k'}(\theta_\tau)$ for any i', j', k' permutation of i, j, k. The potential is easily computed as before, by adding the term $\frac{1}{3} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \sum_{k \neq i,j} g_{jk} g_{ki} \tau_{ijk}(\theta_\tau)$.

In general, it is possible to include additional utility components to (1) as long as we can find restrictions on the payoffs that guarantee the existence of a potential function. Some examples are provided in the proofs of Appendix D.

Undirected networks. The model is concerned about directed networks, but this is not essential to most of the characterizations. The results about the existence of the potential game, the existence and characterization of the stationary distribution and the relation with the ERGM model can be extended to undirected networks with minimal modifications (see Chandrasekhar and Jackson (2014) or Mele and Zhu (2015)).[26] Most of the asymptotic and convergence results in the next section hold also for undirected networks (see Diaconis and Chatterjee (2011)).

Sparsity. The model generates *dense networks*, i.e. each player can potentially form all his $n - 1$ links. This means that as $n \to \infty$ the unconditional probability of a link does not become vanishingly small (see Lovasz (2012)). Chandrasekhar and Jackson (2014) show that assuming sparsity reduces the computational complexity of estimation and it implies good statistical properties (e.g. consistency). I show below that our model with negative linking externalities is compatible with a certain degree of sparsity.

[26] It is also possible to include binary actions (e.g. decision to smoke) into the model, as in Badev (2013).

3 Estimation and Identification

3.1 Computational Problem

Estimation and inference are complicated by the structure of the likelihood function, which is known up to the normalizing constant $c(\mathcal{G}, X, \theta) = \sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X, \theta)]$. To compute the latter constant at parameter vector θ for a network of n players, we would need to sum the value of the potential function over all $2^{n(n-1)}$ possible network configurations. For example, if $n = 10$ players, there are $2^{90} \backsimeq 10^{27}$ network configurations. A supercomputer that can compute 10^{12} potential functions in one second would take almost 40 million years to compute the constant. Therefore standard maximum likelihood maximization routines are impractical. A standard Bayesian estimation approach would encounter the same challenges. Let $p(\theta)$ be the prior distribution, and let the likelihood function of the observed data (g, X) be the long-run stationary distribution of the model $\pi(g, X, \theta)$. The posterior distribution of θ is

$$p(\theta|g, X) = \frac{\pi(g, X, \theta)\, p(\theta)}{\int_\Theta \pi(g, X, \theta)\, p(\theta)\, d\theta}. \tag{7}$$

Using a standard Metropolis-Hastings algorithm to sample from this posterior, we would have to compute ratios

$$\begin{aligned} \alpha(\theta, \theta') &= \min\left\{1, \frac{p(\theta'|g, X)\, q_\theta(\theta|\theta')}{p(\theta|g, X)\, q_\theta(\theta'|\theta)}\right\} \\ &= \min\left\{1, \frac{\exp[Q(g, X, \theta')]}{\exp[Q(g, X, \theta)]} \frac{c(\mathcal{G}, X, \theta)}{c(\mathcal{G}, X, \theta')} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)}\right\}. \end{aligned}$$

that contain the normalizing constant $c(\mathcal{G}, X, \theta)$, and thus cannot be computed.

3.2 Network simulations

This computational problem is common to many models in the statistical literature. The usual approach suggested in the ERGM literature is to provide an approximation of the normalizing constant and the likelihood, using Markov Chain Monte Carlo simulation methods (Snijders (2002)). For a fixed parameter value θ, the algorithm simulates a Markov chain of networks whose unique invariant distribution is (5).

ALGORITHM 1 ***Metropolis-Hastings for Network Simulations***
Fix a parameter vector θ. At iteration r, with current network g_r

1. *Propose a network g' from a proposal distribution $g' \sim q_g(g'|g_r)$*
2. *Accept network g' with probability $\alpha_{mh}(g_r, g')$*

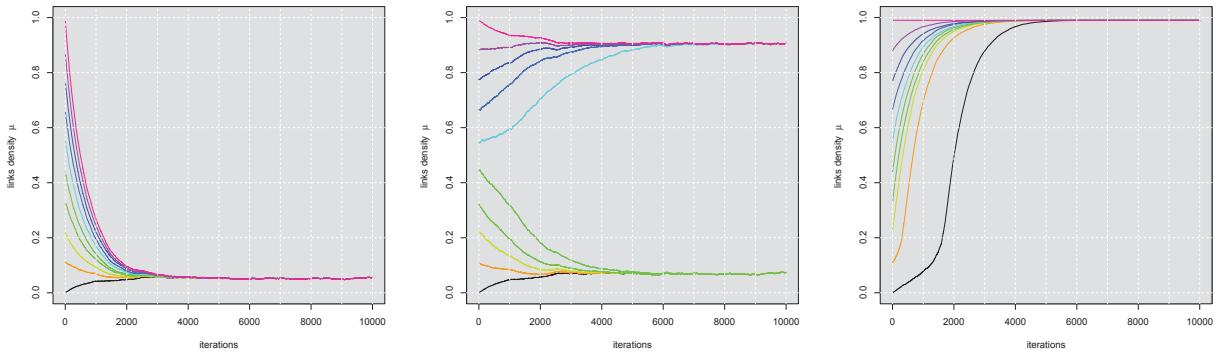
$$\alpha_{mh}(g_r, g') = \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g_r, X, \theta)]} \frac{q_g(g_r|g')}{q_g(g'|g_r)}\right\} \tag{8}$$

The main advantage of this simulation strategy is that the acceptance ratio $\alpha_{mh}(g_r, g')$ does not contain the normalizing constant $c(\mathcal{G}, X, \theta)$. Each quantity in the acceptance ratio can be computed exactly. The Metropolis-Hastings structure of the algorithm guarantees convergence. Standard results[27] show that the chain generated by the algorithm converges uniformly to the likelihood of the model.

However, in practice the convergence can be slow. The standard version of this algorithm is a *local sampler*: at each iteration, we select a random player i with probability $1/n$, we then select another player j with probability $1/(n-1)$, and we update the link g_{ij} according to the Metropolis-Hastings ratio (8). To be concrete, let's implement the local sampler in a special case with homogeneous players, that includes only direct utility and indirect utility.

$$\pi_n(g; \alpha, \beta) = \frac{\exp\left\{\left[\alpha \sum_{i=1}^n \sum_{j=1}^n g_{ij} + \beta \sum_{i=1}^n \sum_{j=1}^n \sum_{k \neq i}^n g_{ij} g_{jk}\right]\right\}}{c(\alpha, \beta, \mathcal{G}_n)} \tag{9}$$

Figure 1: Network simulations at different parameter values



(A) (B) (C)

Traceplots of simulations of model (9) using Algorithm 1 with local chains. The simulations are obtained for a network with $n = 100$ players, with parameters $\alpha = -3$ and $\beta = \{1/n, 3/n, 7/n\}$ (Panel (A), (B) and (C) respectively). Each simulation is started at 10 different starting networks, each corresponding to a directed Erdos-Reny network with probability of link $\mu = \{0, .111, .222, .333, .444, .555, .666, .777, .888, 1\}$.

I simulate this model using the local sampler just described. In Figure 1 I show the trace plot of algorithm 1 for three different parameter vectors: $\alpha = -3$ and $\beta = \{1/n, 3/n, 7/n\}$ (Panels (A), (B) and (C) respectively). I start the simulations at 10 different starting values, each corresponding to a directed Erdos-Renyi with probability of linking $\mu = \{0, 0.111, 0.222, 0.333, 0.444, 0.555, 0.666, 0.777, 0.888, 1\}$. In the figures I show the link density of each iteration.[28] The network has $n = 100$ players.

[27] See Meyn and Tweedie (2009), Levin et al. (2008)

[28] The traceplot for the density of indirect links (the second network statistics) has similar pattern.

The simulations (A) with parameters $(\alpha, \beta) = (-3, 1/n)$ converge to a very sparse network; while the simulations (C) with parameters $(\alpha, \beta) = (-3, 7/n)$ converge to a very dense network. On the other hand, when we consider simulations in (B) with parameters $(\alpha, \beta) = (-3, 3/n)$, we observe that the chains started at relatively dense networks converge to a very dense network with density of links $\mu_2 \approx 0.92$, while chains started at relatively sparse networks converge to a sparse network, with link density $\mu_1 \approx 0.07$.

This is a phenomenon that practitioners have encountered in the ERGM literature and in statistical physics models.[29] The model seems to put very large probability mass on few networks, an issue called degeneracy. In the next section I provide several theoretical results that explain the such simulation problems.

3.3 Large network analysis

There are two ways to study the asymptotic properties of empirical network formation models. First, we can consider a sample of independent networks and study the properties of the model as the number of observed networks grows large (*many networks asymptotics*). Second, we can consider a single network observation, and a sequence of graphs whose number of players n grows large *(large networks asymptotics*). The former case is relatively standard and follows from the theory of exponential families under usual regularity conditions.[30] Identification of the parameters is also standard.

The case of large networks is relatively more complicated, and only recently gained attention in the literature.[31] It is also the most relevant case in empirical applications, because the econometrician usually observes a single network in the data.I provide a detailed asymptotic analysis of the model in the homogeneous players case, using a mix of graph limits theory, large deviations and mean-field approximations for the exponential family.[32]

Consider a sequence of directed graphs g_n, where the number of nodes grows large, $n \to \infty$. To consider such network limits, I re-scale the potential function, to avoid exploding terms as $n \to \infty$: each aggregate utility term is scaled by a factor $n^{v(H)}$, where $v(H)$ is the number of vertices involved in the utility term.

Consider the model

$$\pi_n(g; \alpha, \beta) \quad = \quad \frac{\exp\left\{n^2\left[\alpha \frac{\sum_{i=1}^n \sum_{j=1}^n g_{ij}}{n^2} + \beta \frac{\sum_{i=1}^n \sum_{j=1}^n \sum_{k \neq i}^n g_{ij} g_{jk}}{n^3}\right]\right\}}{c(\alpha, \beta, \mathcal{G}_n)} \tag{10}$$

[29] See Snijders (2002), Butts (2009), Koskinen (2008) for examples.

[30] See Lehman (1983), Sheng (2012), Badev (2013).

[31] See Chandrasekhar and Jackson (2014), Graham (2014), Leung (2014*a*), DePaula et al. (2014), Ridder and Sheng (2015), Menzel (2015) for recent contributions.

[32] The explanation that follows is relatively informal, and I leave the technical details about graph limits, large deviations and mean-field approximations in Appendix D.

Notice that the above model is equivalent to the original model (9) with parameter β re-scaled by n.[33] Let's define the re-scaled network statistics

$$t(H_1, g) \equiv \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \quad \text{and} \quad t(H_2, g) \equiv \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \neq i}^{n} g_{ij} g_{jk}$$

and the corresponding re-scaled potential function $\mathcal{T}(g)$

$$\mathcal{T}(g) = \alpha t(H_1, g) + \beta t(H_2, g) \tag{11}$$

I can then rewrite the model's likelihood as

$$\pi_n(g; \alpha, \beta) = \frac{\exp\left\{n^2 \left[\alpha t(H_1, g) + \beta t(H_2, g)\right]\right\}}{c(\alpha, \beta, \mathcal{G}_n)} = \exp\left\{n^2 \left[\mathcal{T}(g) - \psi_n\right]\right\} \tag{12}$$

where the log-normalizing constant ψ_n is defined as

$$\psi_n = \frac{1}{n^2} \log \sum_{g \in \mathcal{G}_n} \exp\left[n^2 \mathcal{T}(g)\right] \tag{13}$$

The following theorems characterize this model as n becomes large, providing an asymptotic approximation of the normalizing constant and a discussion of identification.[34] Theorem 2, 3 and 4 below are special cases of a more general result. Indeed, I can show that for a large class of models, the normalizing constant in (13) solves a variational problem in the space of probability functions on the unit square (Theorem 10 in Appendix D). Using such general result I obtain the following theorems.

THEOREM 2 ***(Non-negative link externalities)*** *Model (10) with non-negative link externalities $\beta \geq 0$ exhibits the following behavior for $n \to \infty$*

1. *The asymptotic normalizing constant ψ solves*

$$\psi \equiv \lim_{n \to \infty} \psi_n = \max_{\mu \in [0,1]} \left\{\alpha \mu + \beta \mu^2 - \mu \log \mu - (1 - \mu) \log(1 - \mu)\right\} \tag{14}$$

2. *The networks generated by the model are indistinguishable from a directed Erdos-Renyi graph with linking probability μ^*, defined as follows:*

[33] This is important when one runs the simulations using the usual ERGM form. For example, one need to use $\beta^o = \frac{\beta}{n}$ for simulations using the `ergm` package in the software R. The same is true for the replication routines of this paper.

[34] I use the approach developed in Radin and Yin (2013) and Aristoff and Zhu (2014) to study the maximization problem implied by the simplified variational problem.

(a) If the maximization (14) has a unique solution, then μ^* solves

$$\mu = \frac{\exp\left[\alpha + 2\beta\mu\right]}{1 + \exp\left[\alpha + 2\beta\mu\right]} \tag{15}$$

and satisfy $2\beta\mu(1-\mu) < 1$, for almost all $\alpha \in \mathbb{R}$ and $\beta \geq 0$.

(b) If the maximization (14) has two solutions, then μ^* is picked randomly from some probability distribution over μ_1^* and μ_2^*, such that $\mu_1^* < 0.5 < \mu_2^*$ and both solve equation (15) and satisfy $2\beta\mu(1-\mu) < 1$.

Proof. See Theorem 11 and Theorem 19 in Appendix D. ■

The first part of the theorem provides a consistent estimate for the log-normalizing constant of model (10), as the solution of a maximization problem. This formulation is the asymptotic analogous of the variational representation of the discrete exponential family in mean parameterization, as shown in Wainwright and Jordan (2008).

The second part of Theorem 2 shows that when $\beta \geq 0$, a realization of the model with parameters (α, β) will be indistinguishable from the realization of a model with parameters $(\alpha', 0)$ where $\alpha' = \log \frac{\mu^*}{1-\mu^*}$ and μ^* is the maximizer of (14) and solves equation (15). Indeed, both vectors of parameters correspond to the same directed Erdos-Renyi model for large n.[35] Moreover, if the maximization problem (14) has two maxima, the parameters (α, β) can generate two completely different networks, one with link density $\mu_1^* < 0.5$ and one with link density $\mu_2^* > 0.5$.[36] Such behavior of the model has been observed by practitioners (see Snijders (2002) for example) using simulation methods, and it was proven analytically for undirected networks in Diaconis and Chatterjee (2011). Our theorem extends their result to directed networks.

There are two main corollaries of Theorem 2: first, the externality cannot be identified when $\beta \geq 0$; second the network simulation algorithm developed in the ERGM literature and discussed in the previous section (ALGORITHM 1) is not necessary for this region of the parameter space. The Erdos-Renyi graphs can be easily simulated using random Bernoulli draws.

On the other hand, when the link externalities are negative and sufficiently large in magnitude, we can identify the parameters, as shown in the next result.

THEOREM 3 ***(Negative link externalities)*** *If $\beta < 0$ and sufficiently large in magnitude, the model in (10) is asymptotically different from a directed Erdos-Renyi model.*

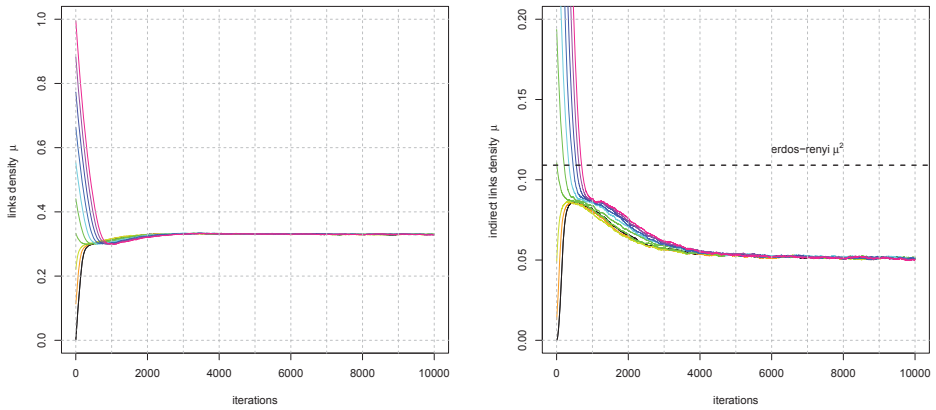
Proof. See Theorem 14 in Appendix D. ■

[35] The model with homogeneous players and only positive externalities violates the condition of *expectation-identification* in Chandrasekhar and Jackson (2014). The condition requires that different parameters correspond to different expected network statistics. This is clearly violated in this special case.

[36] In the applied mathematics and physics literature, such sets of parameters are crucial because they generate a phase transition. See for example Radin and Yin (2013).

The problem of asymptotic identification is generated by positive externalities: a model with *sufficiently large negative externalities* generates graphs that do not converge asymptotically to directed Erdos-Renyi networks. For a sufficiently large negative link externality, model (10) generates networks that are more sparse than an Erdos-Renyi graph. Sparsity indeed has been shown to be an important ingredient for identification in network formation models (see Chandrasekhar and Jackson (2014) for example). Furthermore, for $\beta < 0$ the likelihood is unimodal.

Figure 2: Model with negative externalities does not converge to Erdos-Renyi graphs



(A) (B)

The 2 panels show simulations of model (10) with parameters $(\alpha, \beta) = (5, -10)$. The network has $n = 300$ players, and I run the simulation for 1500000 iterations, sampling every 150 iterations. In Panel (A) I show the convergence of the direct links density to $\mu = 0.3302742$. If model (10) converges to an Erdos-Renyi model, then the density of indirect links should be $\mu^2 = 0.109081$, shown as the horizontal dashed line in Panel (B). Therefore the model does not converge to a model with independent links.

The simulations in Figure 2 show evidence that the model with $\beta < 0$ does not converge to an Erdos-Renyi model in the large n limit. I use 10 different starting values, corresponding to Erdos-Renyi graphs with linking probability μ equi-spaced on the unit interval, for a network of size $n = 300$.[37] I report simulations for $(\alpha, \beta) = (5, -10)$, converging to a network density of $\mu = 0.3302742$ (Figure 2(A)). If the model converges to an Erdos-Renyi graph, the density of indirect links should be $\mu^2 = 0.109081$. Figure 2(B) proves that this is not the case. Indeed our model converges to a different density of indirect links, smaller than the corresponding Erdos-Renyi indirect link density.

The results of Theorems 2 and 3 apply to more general models. Consider a model that includes the effect of common links, i.e. cyclic triangles, with re-scaled potential

$$T(g) = \alpha t(H_1, g) + \beta t(H_2, g) + \gamma t(H_3, g) \tag{16}$$

[37] The theoretical results approximate networks of size $n > 50$ quite well.

where the network statistics $t(H_1, g)$ and $t(H_2, g)$ are the same as in model (10) and $t(H_3, g) = n^{-3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \neq i}^{n} g_{ij} g_{jk} g_{ki}$. The following result holds.

THEOREM 4 *Consider model (16) as $n \to \infty$:*

1. ***(Non-negative externalities)*** *If $\beta \geq 0$ and $\gamma \geq 0$, the asymptotic normalizing constant ψ solves*

$$\psi \equiv \lim_{n \to \infty} \psi_n = \max_{\mu \in [0,1]} \left\{ \alpha\mu + \beta\mu^2 + \gamma\mu^3 - \mu \log \mu - (1-\mu) \log(1-\mu) \right\} \tag{17}$$

and the model is asymptotically indistinguishable from a directed Erdos-Renyi graph. The linking probability μ^* is the maximizer of (17). If the maximization problem (17) has multiple solutions, then μ^* is picked randomly from some probability distribution over the maximizers.

2. ***(Negative externalities)*** *If at least one of the externalities is negative (i.e, $\beta < 0$ or $\gamma < 0$) and sufficiently large, then model (16) does not converge asymptotically to a directed Erdos-Renyi graph. In such case, the externalities can be identified.*

Proof. These statements are proven in Theorem 12, Theorem 17, Theorem 18 and Theorem 16 in Appendix D. ■

The generalization to additional externalities with alternative utility subgraphs is straightforward, but tedious. I provide some examples in Appendix D.

The main lesson from this analysis is that models with homogeneous players including *only positive externalities* converge asymptotically to trivial Erdos-Renyi models and are essentially ill-identified in the large n limit. However, as long as *at least one externality is negative and sufficiently large*, the model does not degenerate into a trivial independent-links model.

While it was not possible to prove similar results for the more general model with heterogeneous players, a conjecture is that the sign of the linking externalities is crucial for identification in these class of models.[38]

3.4 Convergence of network simulations

The analysis in the previous section shows that in many cases we can approximate the likelihood of our model using the likelihood of an Erdos-Renyi graph, which is easy to estimate and simulate.

Using similar techniques we can prove that the standard local sampler used in the ERGM

[38] I am not aware of any result in the literature on graph limits that allows for covariates. Preliminary results are contained in Mele and Zhu (2015).

literature (shown above) may have very slow convergence, even in this simple case. To be precise, I prove that there is a V-shaped region of the parameter space (see Figure 3(A)) in which the sampler has exponentially slow convergence to stationarity, making estimation impractical.

THEOREM 5 ***(Convergence of local sampler with non-negative externalities)*** *Consider model (16), with probability of meeting $\rho_{ij} = 1/(n(n-1))$. Fix any $\gamma \geq 0$. Then in the case of non-negative externalities $\beta \geq 0$, there exist a V-shaped region of the parameter space, delimited by functions $S_\gamma(\phi_1(\alpha))$ and $S_\gamma(\phi_2(\alpha))$ such that*

1. *If (α, β) belongs to the V-shaped region, the model converges to stationarity in e^{Cn^2} steps, where $C > 0$ is a constant. This result extends to any local sampler.*

2. *Otherwise, the model converges in $Cn^2 \log n$ steps, where $C > 0$ is a constant.*

Proof. Follows from the proof of Theorem 4 above, and Theorems 5 and 6 in Bhamidi et al. (2011) ■

Figure 3(A) shows the V-shaped region delimited by the functions $S_\gamma(\phi_1(\alpha))$ and $S_\gamma(\phi_2(\alpha))$, for $\gamma = 0$. The derivation of such functions is shown in Appendix D.

The intuition of the result is simple. In the V-shaped region the problem (17) has two *local maxima*: the sampler spends exponentially long time at one of the local maxima. In other words, once the sampler reaches a local maximum, there is probability e^{-Cn^2} to escape such state of the network. As a consequence, the sampling is practically infeasible with a local sampler. An increase in γ increases the area of exponentially slow convergence. For a visualization see proof of Theorem 17 in Appendix D.

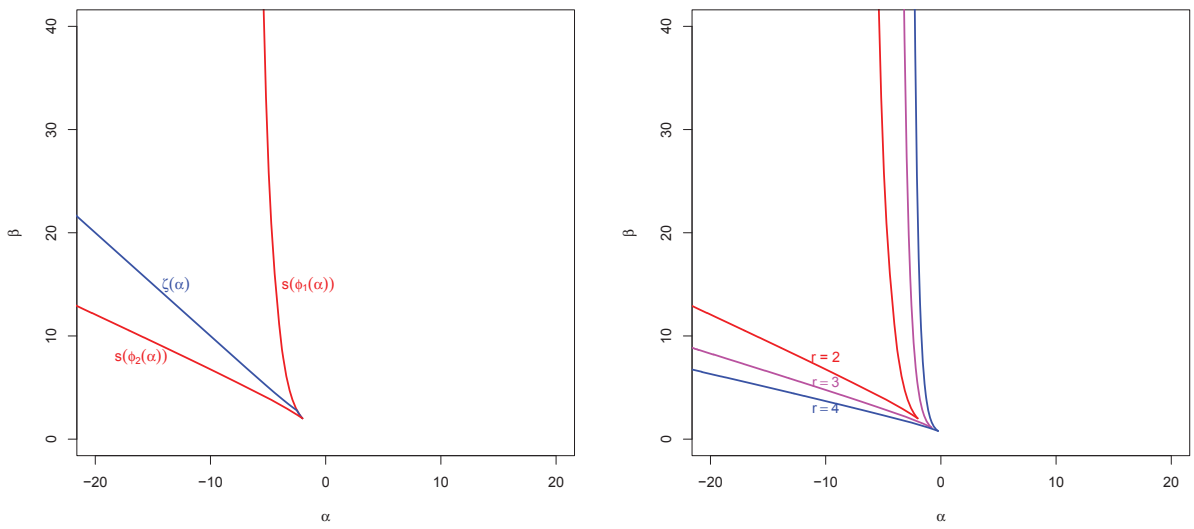
In Figure 3(B) I focus on a model with only two parameters

$$T(g) = \alpha t(H_1, g) + \beta t(H_2, g)$$

and show how the V-shaped region changes when we consider alternative network statistics $t(H_2, g)$. In the figure, r defines the order of interdependencies of the second utility term (the externality): $r = 2$ corresponds to the original model in (10); $r = 3$ corresponds to a model with utility from direct links and common connections (cyclic triangles), i.e. $t(H_2, g) = n^{-3} \sum_i \sum_j \sum_k g_{ij} g_{jk} g_{ki}$; $r = 4$ corresponds to a model with utility from direct links and externality from 4 connections, e.g. 4-cycle with $t(H_2, g) = n^{-4} \sum_i \sum_j \sum_k \sum_l g_{ij} g_{jk} g_{kl} g_{li}$. The general result is that if we increase the order of dependencies the size of the V-shaped region increases. The derivations and analysis with several alternative utility terms are provided in Appendix D.

When the convergence is quadratic (i.e. in order $n^2 \log n$ steps), the sampler is feasible for moderate size networks ($n < 500$). However, this is the case of Theorem 4: the model behaves asymptotically as an Erdos-Renyi model. Therefore the sampler can be simplified drastically to simulate the model as a matrix of Bernoulli variables.

Figure 3: Visualization of the regions described in Theorem 5



(A) (B)

Panel (A) shows the functions $S_\gamma(\phi_1(\alpha))$ and $S_\gamma(\phi_2(\alpha))$ described in Theorem 5. I fix $\gamma = 0$ for this picture. The function $\zeta(\alpha)$ is the value of the externality β for which problem (17) has two global maxima, for a given parameter α. Panel (B) shows how the V-shaped region delimited by $S_\gamma(\phi_1(\alpha))$ and $S_\gamma(\phi_2(\alpha))$ change if we consider the model with direct utility and only one externality, i.e. a model with two parameters only. Here r defines the order of interdependencies of the second utility term (the externality): $r = 2$ corresponds to the original model in Theorem 2; $r = 3$ correspond to a model with direct links utility and utility from common connections (cyclic triangles); $r = 4$ corresponds to a model with direct links utility and utility from 4 common connections (e.g. 4-cycle). If we increase the order of dependencies the size of the region increases. The derivations and additional utility terms are considered in Appendix D.

In Appendix B, I suggest a modification of the local algorithm that allows for large steps. This should improve convergence when the likelihood is bimodal. I show some simulation evidence that this is the case.

4 Simulation and estimation in finite networks

I estimate the posterior distribution of the structural parameters using an approximate version of the *exchange algorithm* (see Murray et al. (2006)). The approximate algorithm uses a double Metropolis-Hastings step to avoid the computation of the normalizing constant $c(\mathcal{G}, X, \theta)$ in the likelihood, as in Liang (2010).[39] Several authors have proposed similar

[39]This improvement comes with a possible cost: the algorithm may produce MCMC chains of parameters that have very poor mixing properties (Caimo and Friel, 2010) and high autocorrelation. I partially correct for this problem by carefully calibrating the proposal distribution. In this paper I use a random walk proposal. Alternatively one could update the parameters in blocks or use recent random block techniques

algorithms in the ERGM literature.[40]

The idea of the algorithm is to sample from an augmented distribution using an auxiliary variable. At each iteration, the algorithm proposes a new parameter vector θ', drawn from a suitable proposal distribution $q_\theta(\theta'|\theta)$; in the second step, it samples a network g' (the auxiliary variable) from the likelihood $\pi(g', X, \theta')$; finally, the proposed parameter is accepted with a probability $\alpha_{ex}(\theta, \theta')$, such that the Markov chain of parameters generated by these update rules, has the posterior (7) as unique invariant distribution.[41]

ALGORITHM 2 ***(APPROXIMATE EXCHANGE ALGORITHM)***
Fix the number of network simulations R. At each iteration t, with current parameter $\theta_t = \theta$ and network data g:

1. *Propose a new parameter θ' from a distribution $q_\theta(\cdot|\theta)$,*
2. *Start **ALGORITHM 1** at the observed network g, iterating for R steps using parameter θ' and collect the last simulated network g'*
3. *Accept parameter θ' with probability $\alpha_{ex}(\theta, \theta', g', g)$ where*

$$\alpha_{ex}(\theta, \theta', g', g) = \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]}\right\}. \tag{18}$$

The main advantage of this algorithm is that all quantities in the acceptance ratio (18) can be evaluated: there are no integrals or normalizing constants to compute. This simple modification of the original Metropolis-Hastings allows sampling from the posterior.

The sampler is likely to accept proposals that move towards high density regions of the posterior, but it is likely to reject proposals that move towards low density regions of the posterior. The formal statement about convergence is contained in the following theorem.

THEOREM 6 *(Ergodicity of the Approximate Exchange Algorithm). The approximate exchange algorithm is ergodic, and it converges to the correct posterior distribution.*

Proof. In Appendix B. ■

In practice, the algorithm produces good samples as long as the number of steps of the network simulation algorithm is big enough and the algorithm is run for a sufficient number

as in Chib and Ramamurthy (2009) to improve convergence and mixing.

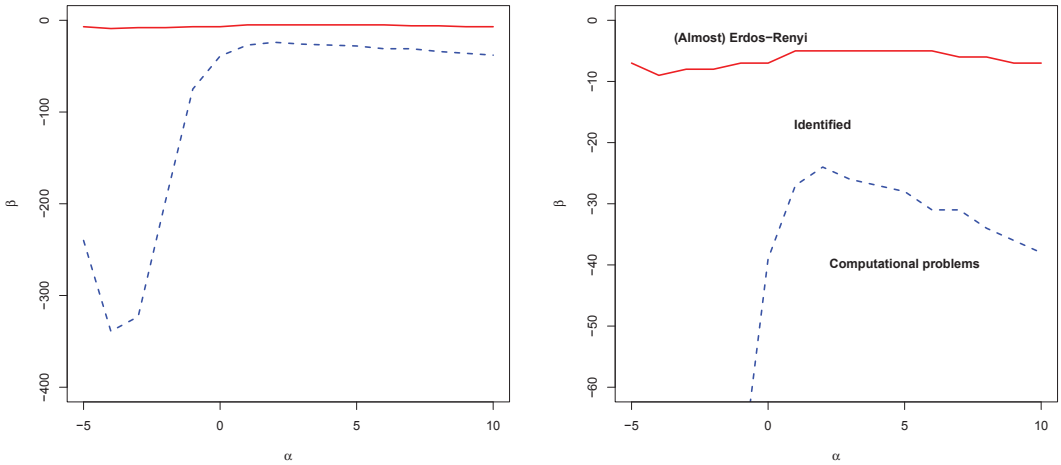
[40] Caimo and Friel (2010) use the exchange algorithm to estimate ERGM. They improve the mixing of the sampler using the snooker algorithm. Koskinen (2008) proposes the Linked Importance Sampler Auxiliary variable (LISA) algorithm, which uses importance sampling to provide an estimate of the acceptance probability. Another variation of the algorithm is used in Liang (2010).

[41] The result in Lemma 1 in Appendix B shows that choosing the observed network as initial network for the simulations guarantees that the *approximate* and the *exact* exchange algorithm have the same acceptance ratio, for any length R of the network simulations. Therefore, the proof of convergence to the correct posterior only needs to show the convergence of the proposal distribution, i.e. convergence of the network simulations to the stationary equilibrium of the model (see details in Appendix B).

of iterations.

In general, for a fixed number of network simulations R, the samples generated by the algorithm will converge to a posterior that is "close" to the correct posterior. As $R \to \infty$ the algorithm converges to the exact exchange algorithm of Murray et al. (2006), producing exact samples from the posterior distribution. However, an higher value of R would increase the computational cost and result in a higher rejection rate for the proposed parameters.

Figure 4: Approximate region of identified parameters for Model (10) when $\beta < 0$



Panel B is a zoom-in of Panel A, to show the details of the graph. The two lines are the boundaries of three regions. For values of β above the red solid line, the model is indistinguishable from an Erdos-Renyi model with independent links. For β below the red solid line and above the blue dashed line, the externality can be identified and estimated using our algorithm. For β below the blue dashed line, the externality cannot be estimated because of computational and identification issues in finite samples.

I test the performance of the estimation algorithms using artificial data.[42] Ideally, we want to compare the results of the approximate exchange algorithm with the exact algorithm. This is feasible for a special case, where preferences depend only on direct and mutual links

[42] All the computations with artificial data are performed in a standard desktop Dell Precision T7620 with 2 Intel Xeon CPUs E5-2697 v2 with 12 Dual core processors at 2.7GHZ each and 64GB of RAM. For replication purposes, there is a package in Github at https://github.com/meleangelo/netnew. In all estimation exercises I use independent normal priors $\mathcal{N}(0, 10)$. The proposal of the exchange algorithm is a random walk $\mathcal{N}(0, \Sigma)$. I repeat the estimation twice: the first time I use a diagonal Σ; in the second round, I use the covariance from the first round as baseline. In all simulations the probability of large steps is 0.001 and a large step updates $0.1n$ links. However, the gains from the modified algorithm are minimal in the region of negative β's because the likelihood is usually unimodal.

(i.e. excluding friends of friends and popularity effects). These results are in appendix E and show good performance of the algorithm.

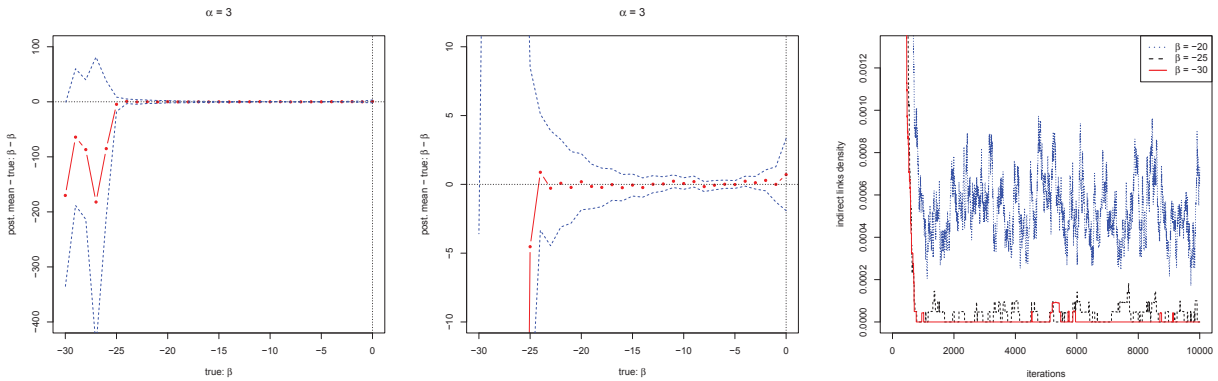
I focus on estimation for the area of parameters that allows identification, i.e. when at least one of the externalities is negative. The theoretical results suggest that in this region we should be able to estimate the externality parameters with precision. I use model (10) to perform the exercise.

The main result is shown in Figure 4(A) (and with more detail in Figure 4(B)). The two lines in the figure delimit three (approximate) regions of the parameters. For values of β above the solid line, the model is indistinguishable from an Erdos-Renyi model, as in the case of positive β. Thus the result in Theorem 2 also translates to negative but small β's.

For β below the solid line and above the dashed line, the parameters are identified. This area correspond to the theoretical result in Theorem 3.

Finally, for values of β below the dashed line, there are some computational problem and estimation becomes impossible: for such values of the externality parameter, the number of indirect links is too close to zero to allow the network simulation algorithm to provide a good sample.

Figure 5: Example estimates for externality $\beta < 0$



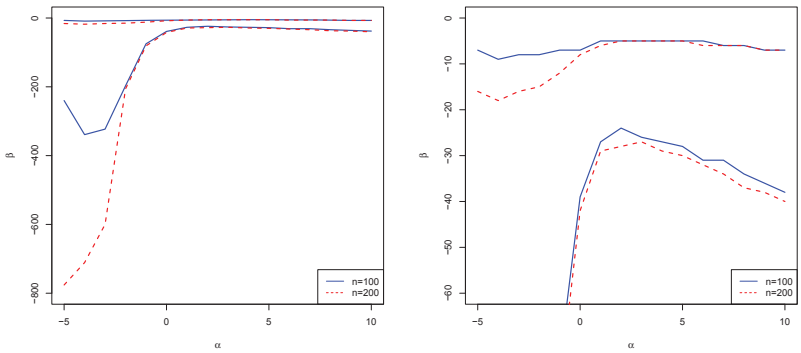
A B C

Panel A shows the difference between the true parameter and the estimated posterior mean (red solid line), with 95% credibility intervals (blue dashed line) for several values of $\beta < 0$. Panel B provides closer detail. The estimates are relatively precise when $\beta > -25$. For $\beta \leq -25$ the posterior mean becomes extremely imprecise and the standard deviation of the posterior is huge. Panel C shows that the problem is the consequence of the simulations hitting lower bound of the network statistics for indirect links. When $\beta = -25$ the simulation output deteriorates so much that estimation becomes impossible.

In Figure 5 I provide a close-up for $\alpha = 3$. Panel A shows the difference between the estimated posterior mean and the true parameter that generates the data (red solid line) and the 95% credibility interval (blue dashed line). If the estimation is precise we expect the difference to be close to zero, with small confidence bands. This is the case for most values of

the externality β. However, for $\beta < -25$ the estimates are so imprecise that the estimation exercise becomes meaningless. This is more evident when we zoom-in the figure in Panel B. The reason for such imprecise estimates is shown in Panel C. When $\beta < -25$ the network simulations hit the lower bound of the indirect links density, i.e. zero. When that happens, the network simulations are extremely inaccurate. In Panel C I show that when $\beta = -20$ (blue dashed) the MCMC has a regular pattern, while for $\beta = -25$ (black dashed) the output of the sampler is highly irregular and skewed. This creates the computational problems in estimation. I performed the same analysis for a grid of parameters with $\alpha \in [-5, 10]$, and $\beta \in [-50, 0]$. The results are contained in the replication files.

Figure 6: Approximate region of identified parameters, $n = 100$ vs $n = 200$



The blue solid line delimits the same regions in Figure 4 for $n = 100$. The red dashed line delimits the three regions for $n = 200$.

In Figure 6, I show that the regions where the parameters are identified change with the size of the network. I compare a model with $n = 100$ (solid blue lines) to a model with $n = 200$ players (dashed red lines). The regions are virtually identical for positive values of α, while they diverge significantly when $\alpha < 0$.

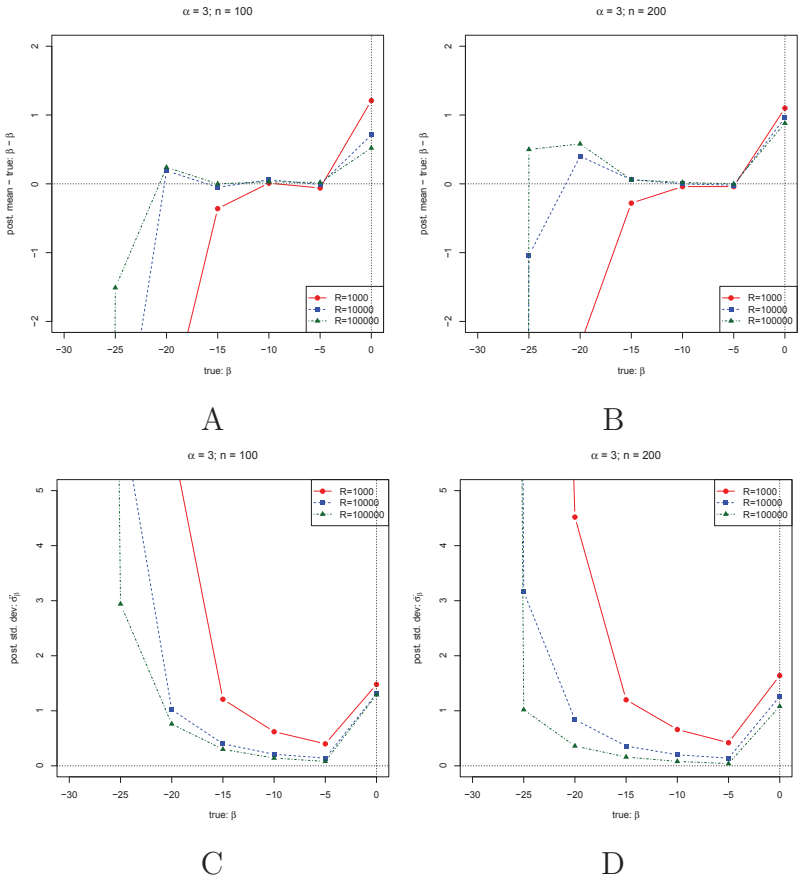
Finally, in Figure 7 I show the effect of the number of network simulations R on the precision of the posterior estimates. As an example I report the results for $\alpha = 3$, but the pattern is the same for other parameters.[43] In Panel A and B, I report the difference between posterior mean and true β for networks of $n = 100$ and $n = 200$ respectively. In Panels C and D I show the posterior standard deviation.

The estimates with $R = 1000$ are relatively imprecise and there is almost no precision gain when we increase networks simulations from $R = 10000$ to $R = 100000$.

On the other hand, the cost of increasing the network simulations is almost linear, e.g.

[43] See the replication files for additional simulations and results.

Figure 7: Estimates and length of network simulations



computational time increase by 10 times when we increase R from 10000 to 100000. Thus we conclude that $R = 10000$ is a good compromise between precision of the estimates and computational cost for networks of this size.[44]

The simulations suggest that convergence is almost quadratic in n in this area of the parameter space. Using the intuition developed in the theoretical result for $\beta \geq 0$, I conjecture that the speed of convergence is relatively fast because in this region the likelihood is unimodal.

We conclude that even in the area considered in Theorem 3, where externalities can be identified, we can encounter estimation problems in finite networks, due to computational approximations. In such regions, estimation becomes impossible.

In Appendix B we provide some modification of the local simulation algorithm that could improve estimation with multimodal likelihoods, and show some results that accelerate convergence of the posterior estimation.

[44]The replication files contain details and estimation times.

5 Conclusions

The model presented in this paper shows that an exponential random graph model (ERGM) can be thought of as equilibrium outcome of a strategic network formation game. My model considers payoffs that depend on direct connections but also link externalities, and constrains the preferences to guarantee the existence of a potential function. I have shown that such restrictions guarantee that the model converges to a unique stationary equilibrium that corresponds to an ERGM.

I contribute to the literature by studying the equilibrium properties of the model in large networks, using a mix of graph limits, large deviations and variational methods for the exponential family. In particular I show that the sign of the linking externalities is crucially related to the identification. When the externalities are all positive, the model is asymptotically indistinguishable from an Erdos-Renyi graph. On the other hand, if at least one of the externalities is negative and sufficiently large, the model does not converge to an Erdos-Renyi graph and the externality can be identified.

I propose a Bayesian MCMC estimation method using an approximate exchange algorithm. Our theoretical identification result shows that negative and sufficiently large externalities can be estimated and identified. However, I show that in finite networks there are some computational problems even in this region of the parameter space, making estimation of the link externalities infeasible in some cases.

In this paper, I have considered approximate estimation through sampling, using a Markov Chain Monte Carlo method to approximate the likelihood and the posterior distribution of the parameters. As an alternative, we could approximate the likelihood using a variational deterministic technique. Some preliminary attempts in this direction are provided in He and Zheng (2013) and Mele (2015), using (structural) mean-field approximations for the exponential family (see Wainwright and Jordan (2008) and Bishop (2006)). An alternative approach is provided in Chandrasekhar and Jackson (2014), by imposing sparsity, which implies good statistical properties of the estimators and improves the tractability of the model.

In the development of a model of empirical network formation, we also need to consider how modeling unobserved heterogeneity affects our results. Graham (2014) includes unobserved heterogeneity in a model with heterogeneous agents, but excludes the link externalities that are central to the model presented here. I can include unobserved heterogeneity in our model, with substantial increase in computational burden. However, it is not clear that we can separately identify externalities and unobserved heterogeneity using only one network realization.

References

Acemoglu, D., M. Dahleh, I. Lobel and A. Ozdaglar (2011), 'Bayesian learning in social networks', *Review of Economic Studies* **forthcoming**.

Amemiya, Takeshi (1981), 'Qualitative response models: A survey', *Journal of Economic Literature* **19**(4), 1483–1536.

Andrieu, C. and G. O. Roberts (2009), 'The pseudo-marginal approach for efficient monte carlo computations', *Annals of Statistics* **37**(2), 697–725.

Aristoff, David and Lingjiong Zhu (2014), On the phase transition curve in a directed exponential random graph model.

Atchade, Yves and Jing Wang (forthcoming), 'Bayesian inference of exponential random graph models for large social networks', *Communications in Statistics - Simulation and Computation* .

Athey, Susan and Guido Imbens (2007), 'Discrete choice models with multiple unobserved choice characteristics', *International Economic Review* **48**(4), 1159–1192.

Badev, Anton (2013), Discrete games in endogenous networks: Theory and policy.

Bala, Venkatesh and Sanjeev Goyal (2000), 'A noncooperative model of network formation', *Econometrica* **68**(5), 1181–1229.

Bandiera, Oriana and Imran Rasul (2006), 'Social networks and technology adoption in northern mozambique', *Economic Journal* **116**(514), 869–902.

Besag, Julian (1974), 'Spatial interaction and the statistical analysis od lattice systems', *Journal of the Royal Statistical Society Series B (Methodological)* **36**(2), 192–236.

Bhamidi, Shankar, Guy Bresler and Allan Sly (2011), 'Mixing time of exponential random graphs', *The Annals of Applied Probability* **21**(6), 2146–2170.

Bishop, Christopher (2006), *Pattern recognition and machine learning*, Springer, New York.

Blume, Lawrence E. (1993), 'The statistical mechanics of strategic interaction', *Games and Economic Behavior* **5**(3), 387–424.

Boeckner, Derek (2013), Directed graph limits and directed threshold graphs. PhD Thesis.

Borgs, C., J.T. Chayes, L. Lovsz, V.T. Ss and K. Vesztergombi (2008), 'Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing', *Advances in Mathematics* **219**(6), 1801 – 1851.

Boucher, Vincent (2013), Structural homophily.

Butts, Carter (2009), Using potential games to parameterize erg models. working paper.

Caimo, Alberto and Nial Friel (2010), 'Bayesian inference for exponential random graph models', *Social Networks* **forthcoming**.

Chandrasekhar, Arun and Matthew Jackson (2014), Tractable and consistent exponential random graph models.

Chatterjee, Sourav and Amir Dembo (2015), Nonlinear large deviations, Working papers, Stanford University.

Chatterjee, Sourav and S.R.S. Varadhan (2011), 'The large deviation principle for the erdos-rnyi random graph', *European Journal of Combinatorics* **32**(7), 1000 – 1017. Homomorphisms and Limits.
URL: *http://www.sciencedirect.com/science/article/pii/S0195669811000655*

Chib, Siddartha and Srikanth Ramamurthy (2009), 'Tailored random block', *Journal of Econometrics* .

Christakis, Nicholas, James Fowler, Guido W. Imbens and Karthik Kalyanaraman (2010), An empirical model for strategic network formation. Harvard University.

Conley, Timothy and Christopher Udry (forthcoming), 'Learning about a new technology: Pineapple in ghana', *American Economic Review* .

Cooley, Jane (2010), Desegregation and the achievement gap: Do diverse peers help? working paper.

Currarini, Sergio, Matthew O. Jackson and Paolo Pin (2009), 'An economic model of friendship: Homophily, minorities, and segregation', *Econometrica* **77**(4), 1003–1045.

Currarini, Sergio, Matthew O. Jackson and Paolo Pin (2010), 'Identifying the roles of race-based choice and chance in high school friendship network formation', *the Proceedings of the National Academy of Sciences* **107**(11), 48574861.

De Giorgi, Giacomo, Michele Pellizzari and Silvia Redaelli (2010), 'Identification of social interactions through partially overlapping peer groups', *American Economic Journal: Applied Economics* **2**(2).

De Marti, Joan and Yves Zenou (2009), Ethnic identity and social distance in friendship formation, Cepr discussion papers, C.E.P.R. Discussion Papers.

DePaula, Aureo, Seth Richards-Shubik and Elie Tamer (2014), Identification of preferences in network formation games. working paper.

Diaconis, Persi and D. Stroock (1991), 'Geometric bounds for eigenvalues of markov chains', *Annals of Applied Probability* **1**(1), 36–61.

Diaconis, Persi and Sourav Chatterjee (2011), Estimating and understanding exponential random graph models.

Echenique, Federico, Roland G. Fryer and Alex Kaufman (2006), 'Is school segregation good or bad?', *American Economic Review* **96**(2), 265–269.

Frank, Ove and David Strauss (1986), 'Markov graphs', *Journal of the American Statistical Association* **81**, 832–842.

Gelman, A., G. O. Roberts and W. R. Gilks (1996), 'Efficient metropolis jumping rules', *Bayesian Statistics* **5**, 599–608.

Gelman, A., J. Carlin, H. Stern and D. Rubin (2003), *Bayesian Data Analysis, Second Edition*, Chapman & Hall/CRC.

Gilles, Robert P. and Sudipta Sarangi (2004), Social network formation with consent, Discussion paper, Tilburg University, Center for Economic Research.

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

Golub, Benjamin and Matthew Jackson (2011), 'Network structure and the speed of learning: Measuring homophily based on its consequences', *Annals of Economics and Statistics* .

Graham, Bryan (2014), An empirical model of network formation: detecting homophily when agents are heterogeneous. working paper.

He, Ran and Tian Zheng (2013), Estimation of exponential random graph models for large social networks via graph limits, *in* 'Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining', ASONAM '13, ACM, New York, NY, USA, pp. 248–255.

Heckman, James J. (1978), 'Dummy endogenous variables in a simultaneous equation system', *Econometrica* **46**(4), 931–959.

Hsieh, Chih-Sheng and Lung-Fei Lee (2012), A structural modeling approach for network formation and social interactions with applications to students' friendship choices and selectivity on activities.

Jackson, Matthew and Alison Watts (2001), 'The existence of pairwise stable networks', *Seoul Journal of Economics* **14**(3), 299–321.

Jackson, Matthew and Allison Watts (2002), 'The evolution of social and economic networks', *Journal of Economic Theory* **106**(2), 265–295.

Jackson, Matthew and Asher Wolinsky (1996), 'A strategic model of social and economic networks', *Journal of Economic Theory* **71**(1), 44–74.

Jackson, Matthew O. (2008), *Social and Economics Networks*, Princeton.

Koskinen, Johan H. (2008), The linked importance sampler auxiliary variable metropolis hastings algorithm for distributions with intractable normalising constants. MelNet Social Networks Laboratory Technical Report 08-01, Department of Psychology, School of Behavioural Science, University of Melbourne, Australia.

Laschever, Ron (2009), The doughboys network: Social interactions and labor market outcomes of world war i veterans. working paper.

Lehman, E. L. (1983), *Theory of Point Estimation*, Wiley and Sons.

Leung, Michael (2014*a*), A random-field approach to inference in large models of network formation. working paper.

Leung, Michael (2014*b*), Two-step estimation of network-formation models with incomplete information. working paper.

Levin, David. A, Yuval Peres and Elizabeth L. Wilmer (2008), *Markov Chains and Mixing Times*, American Mathematical Society.

Liang, Faming (2010), 'A double metropolis-hastings sampler for spatial models with intractable normalizing constants', *Journal of Statistical Computing and Simulation* **forthcoming**.

Liang, Faming, Chuanhai Liu and Raymond J. Carroll (2010), *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*, John Wiley and Sons, Ltd.

Lovasz, L. (2012), *Large Networks and Graph Limits*, American Mathematical Society colloquium publications, American Mathematical Society.

Lovasz, Laszlo and Balazs Szegedy (2007), 'Szemeredi's lemma for the analyst', *GAFA Geometric And Functional Analysis* **17**(1), 252–270.

Mele, Angelo (2015), Approximate variational inference for a model of social interactions. Working Paper.

Mele, Angelo and Lingjiong Zhu (2015), Variational approximations for large strategic ergms.

Menzel, Konrad (2015), Strategic network formation with many agents, Working papers, NYU.

Meyn, Sean and Richard L. Tweedie (2009), *Markov Chains and Stochastic Stability*, Cambridge University Press.

Miyauchi, Yuhei (2012), Structural estimation of a pairwise stable network formation with nonnegative externality.

Monderer, Dov and Lloyd Shapley (1996), 'Potential games', *Games and Economic Behavior* **14**, 124–143.

Murray, Iain A., Zoubin Ghahramani and David J. C. MacKay (2006), 'Mcmc for doubly-intractable distributions', *Uncertainty in Artificial Intelligence* .

Nakajima, Ryo (2007), 'Measuring peer effects on youth smoking behavior', *Review of Economic Studies* **74**(3), 897–935.

Radin, Charles and Mei Yin (2013), 'Phase transitions in exponential random graphs', *The Annals of Applied Probability* **23**(6), 2458–2471.

Ridder, Geert and Shuyang Sheng (2015), Estimation of large network formation games, Working papers, UCLA.

Robert, Christian P. and George Casella (2005), *Monte Carlo Statistical Methods (Springer Texts in Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Rossi, P., R. McCulloch and G. Allenby (1996), 'The value of purchase history data in target marketing', *Marketing Science* **15**(4), 321–340.

Sheng, Shuyang (2012), Identification and estimation of network formation games.

Snijders, Tom A.B (2002), 'Markov chain monte carlo estimation of exponential random graph models', *Journal of Social Structure* **3**(2).

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

Topa, Giorgio (2001), 'Social interactions, local spillovers and unemployment', *Review of Economic Studies* **68**(2), 261–295.

Wainwright, M.J. and M.I. Jordan (2008), 'Graphical models, exponential families, and variational inference', *Foundations and Trends@ in Machine Learning* **1**(1-2), 1–305.

Wasserman, Stanley and Philippa Pattison (1996), 'Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p*', *Psychometrika* **61**(3), 401–425.