Strategic Random Networks and Tipping Points in Network Formation*

Benjamin Golub† Yair Livne‡

November 22, 2010

Abstract

Agents invest costly effort to socialize. Their effort levels determine the probabilities of relationships, which are valuable for their direct benefits and also because they lead to other relationships in a later stage of "meeting friends of friends". In contrast to many network formation models, there is fundamental uncertainty at the time of investment regarding which friendships will form. The equilibrium outcomes are random graphs, and we characterize how their density, connectedness, and other properties depend on the economic fundamentals. When the value of friends of friends is low, there are both sparse and thick equilibrium networks. But as soon as this value crosses a key threshold, the sparse equilibria disappear completely and only densely connected networks are possible. This transition mitigates an extreme inefficiency.

Keywords: network formation, random graphs, random networks, phase transition
JEL Classification Number: D85

*We thank Daron Acemoglu, Abhijit Banerjee, Jeremy Bulow, Arun Chandrasekhar, Matthew Elliott, Alex Frankel, Alex Hirsch, André Henriques, Yuichiro Kamada, Willemien Kets, Scott Kominers, Ruth Kricheli, Antoine Lallour, Carlos Lever, James Liang, Pietro Majer, and Michael Ostrovsky for helpful conversations. We are especially indebted to Matthew O. Jackson and Andrzej Skrzypacz for their invaluable guidance. We gratefully acknowledge financial support from the NET Institute. Golub gratefully acknowledges financial support from the Jaedicke fellowship at the Stanford Graduate School of Business. Livne gratefully acknowledges financial support from the Bonner fellowship at the Stanford Graduate School of Business.

†Graduate School of Business, Stanford University. Email: bgolub@stanford.edu, http://www.stanford.edu/∼bgolub/

‡Graduate School of Business, Stanford University. Email: ylivne@stanford.edu

Social and economic institutions are embedded in the fabric of social networks — the patterns of relationships in society. Why is this fabric sometimes thick and sometimes sparse? How does this depend on the economic fundamentals? What are the welfare consequences? What are the effects of interventions?

A simple but useful way of looking at networks is that they generate value for the agents who populate them in two distinct ways. First, the agents benefit from relationships directly — for example, because those relationships allow for cooperation in repeated interactions between the two agents involved. But the agents also benefit from social ties because the social ties may lead to other relationships, or facilitate the transmission of favors or information from other places in the network.

The purpose of this paper is to explain how the level of each of these values affects the structure and efficiency of equilibrium networks—an issue that is currently poorly understood. This is important for two reasons. Substantively, the answer can help explain why some groups have dense networks while others have sparse ones; it can also help us understand what kinds of interventions substantially improve the efficiency of network formation. Methodologically, it shows the value of a particular fully rational random graph approach to modeling network formation, and highlights the difference between direct and indirect interaction. This may be useful in building reduced form models of network formation as ingredients in other applications.

To develop our model, we build on standard random graph models, which have enough flexibility to be consistent with observed networks. To make the theory economic, we add rational foundations to these models by viewing link probabilities not as exogenous parameters, but as the outcomes of strategic investments. In the Related Literature section below, we discuss how this model goes beyond existing work. The main innovation is that by adding randomness, we gain a great deal of tractability and predictive power at the aggregate level relative to models which work in a deterministic setting.

The model works as follows. A large group of people meet each other for the first time. They simultaneously select levels of socializing effort during an initial period of mingling, such as the first few weeks of an academic program. Interactions take time or some other resource, and agents have costs that are convex in the total amount of this resource they expend. The costs are also proportional to a privately known cost parameter. The probability of the formation of a valuable relationship between two particular people is increasing in their effort levels during this phase of initial meetings. Once the mingling is over, the early social network forms: each link, independently, is realized or not with the appropriate probability. At this point, agents begin reaping the benefits of their investments. Afterwards, agents meet some of the friends of their friends, forming further relationships, which also confer utility. Agents' strategic choices are their effort levels in the mingling stage, and our equilibrium concept assumes they make these knowing how much others are investing, though not what network will form.

The model is intended to capture three key features of network formation. First, the process of forming new relationships exhibits a substantial amount of fundamental uncertainty. When investing effort in socializing, agents can prevent a relationship (by investing

	Low effort	High effort
Occurs when the value of friends of friends is:	below threshold τ_{eq}	below threshold τ_{eq} or above threshold τ_{eq}
Number of friends per agent	converging to a constant	growing as n
Connectedness	fragmented	fully connected
Diameter	∞	2 or 3

Table 1: A summary of the properties of the two equilibrium regimes; n is the population size.

Figure 1: Examples showing typical networks formed in equilibrium with $n = 400$ agents in (a) a high-effort equilibrium and (b) a low-effort equilibrium. The high-effort network has a single component and many links per node, whereas the low-effort network is highly fragmented.

nothing), and they can increase its probability (by increasing their investments), but in general are not be able to guarantee it. Otherwise, we allow for a very general specification of how socializing efforts translate into a relationship probability. Second, in contrast to many network models, agents pay not only for maintenance of links but for the effort it takes to form them — effort which is sometimes futile due to accidents of fate. Again, we allow for a fairly general specification of the costs. Third, as first modeled by Jackson and Rogers (2007), there is both a random element to socializing ("meeting strangers") and a natural source of dependence and clustering that comes from "meeting friends of friends". That is, agents who are friends are more likely than randomly selected agents to have friends in common — one of the robust tendencies of social networks.

The most stylized aspect of the model is the strict separation into a mingling phase, before any links are realized, and a period of meeting friends of friends after an early network is formed. Clearly, in reality these processes overlap somewhat, and a richer model would feature a more gradual transition. Still, we think the timing does capture something important, and that the tractability gained by this assumption outweighs the realism lost.

Adding best responses to a standard random graph model shows how network phenomena like the degree distribution and connectedness relate to economic fundamentals. It also reveals that there are completely new qualitative phenomena that arise when agents best-respond to each other in this setting.

The first main result is that, when the overall cost of resources is not too convex, equilibrium networks come in two varieties: a connected, high-effort regime, or a fragmented, low-effort one. These regimes are extremely different, and which equilibria are present depends on the value of friends of friends[1] — in particular, how it compares to a certain threshold called τ_{eq}. The properties of the regimes are summarized in Table 1, and some illustrative examples are shown in Figure 1. When friends of friends are sufficiently valuable, with their value exceeding the threshold, agents in equilibrium are guaranteed to devote a lot of resources to socializing, and the expected number of friends each has scales as the population size. Networks in this regime are connected with very high probability as the population grows large — indeed, there is a path of length at most three between any two agents. In contrast, when the value of friends of friends falls just slightly below the threshold, another equilibrium exists in which agents socialize significantly less, and the resulting networks consist of many disconnected pieces. The expected number of friends per agent tends to a constant as the network grows large. Thus, in a finite network, arbitrarily small changes in economic fundamentals can lead to arbitrarily large jumps in equilibrium levels of social activity — a result that has not been obtained before, to our knowledge, in an equilibrium network formation setting.

The second main result focuses on efficiency. Assuming that agents' costs depend only on their own efforts, the only externalities in the model are positive: investing in links creates value for others without imposing any costs on them. Thus, any equilibrium will involve weakly too little socializing. Still, some equilibria are vastly more efficient than others. In the areas of the parameter space where there are multiple types of equilibria, the high-effort equilibria realize more value than the low-effort ones by arbitrarily large factors in large societies. Thus, temporary interventions that don't permanently change any of the key parameters can lead to vast changes in the welfare obtained. Moreover, increasing the value agents expect to get from meeting friends of friends can remove the most inefficient equilibria entirely.

The paper is organized as follows. In Section 1, we discuss how our approach relates to the literature. Next, in Section 2, we formally lay out the model. In Section 3, we examine equilibrium and efficiency. In Section 4, we show that analyzing the model as if agents mingle uniformly (without targeting effort depending on others' labels) is not restrictive, assuming there are at least some search costs that agents must pay if they wish to interact non-uniformly. Section 5 concludes.

1 Related Literature

The importance of the basic problem of how social networks form has been widely recognized in economics[2], and the study of rational network formation has a rich history. One strand

[1]The expected value of a friend of friend is the probability of befriending that person times the value of the relationship conditional on it being formed.

[2]Social networks affect economic outcomes in a multitude of ways. They influence decisions and outcomes relating to employment (Topa, 2001), investment (Duflo and Saez, 2003), risk-sharing (Ambrus, Mobius, and Szeidl 2010), education (Calvó-Armengol, Patacchini, and Zenou, 2009), and crime (Glaeser, Sacerdote, and Scheinkman, 1996), to name just a few of their effects. See Granovetter (2005) for a broad survey of the effects of social networks.

of this literature, starting with Myerson (1991) and continuing with Jackson and Wolinsky (1996), Bala and Goyal (2000), and Hojman and Szeidl (2008), among many others, has studied the stability of certain networks to unilateral and bilateral deviations which translate deterministically into changes in the network. The literature is surveyed extensively by Jackson (2005) and Jackson (2008). While this has been a very important approach for understanding aspects of network formation, the approach implicitly assumes that agents know the network insofar as that is important for their deviations, an assumption that is often implausible. In addition, the use of pairwise stability as an equilibrium concept tends to result in very large sets of possible eqiulibria, without any sense of which are more likely. This makes it difficult to take the predictions to the data. In our model, in contrast to these, any network has a positive probability of appearing in equilibrium, but for given parameter values, some networks are overwhelmingly more likely than others. This allows us to make concrete predictions about aggregate behavior, and also makes this type of model a natural fit for structural estimation.

Recently, there has been a growing recognition that an approach featuring stochastic network formation is necessary. We briefly review some of the most recent and influential papers, and explain why our approach is different.

Cabrales, Calvó-Armengol, and Zenou (2009) were among the first to argue that an approach inspired by random networks may provide a useful angle on the theory of network formation. Their modeling takes a mean-field perspective, assuming that agents in a community have weak links with everyone; the link strengths may then informally be interpreted as link probabilities. Our approach is similar in spirit, but seeks to model the network more realistically, viewing the existence of a relationship as a discrete random variable (though the relationship may also have a strength dimension). We view this difference as essential for empirical applications, since, in practice, a link is typically observed to exist or not. Moreover, the work by Cabrales, Calvó-Armengol, and Zenou (2009) does not consider the value of friends of friends (except through the effect on one's friends' behavior). Indeed, in the mean-field model, it is not obvious how to define a friend of a friend, since all pairs of agents are linked to the same extent. By modeling links as discrete random variables and calculating the probabilities of various paths in the network, we are able to account for both direct *and* indirect relationships simply and naturally.

Currarini, Jackson, and Pin (2009a; 2009b) analyze a model in which agents sequentially meet others at random, optimizing their search process to acquire a desirable mix of friends. They are able to use this to estimate, for example, the relative effects of choice and chance in accounting for homophily. The main innovation of our approach is that agents care not only about the composition of the social circle they acquire in the initial meetings process (as in the CJP papers) but also about the benefits they may expect from friends of friends they meet later. We view this ingredient as an essential feature of any rational network model, since it is clear that often agents do take such benefits into account when "networking". By including this element, we will be able to address many of the central questions of network formation in a richer setting, which may change some of the important estimates and conclusions.

Finally, Christakis et al. (2010) have recently proposed a model of network formation suitable for structural econometrics. The model is based on myopic decisions in sequential meetings. While the agents in this model do potentially value benefits that flow indirectly through the network, the model is not a rational equilibrium theory in the classical sense, because it assumes somewhat *ad hoc* limits on the agents' reasoning. We seek to develop a model where agents are behaving optimally in view of the (limited) information they have, without losing the tractability obtained in the work of Christakis et. al.

2 The Environment

Players and Types We now define the game $\Gamma(n)$. The set of agents is $\mathcal{N} = \{1, \ldots, n\}$, with $n \geq 4$. Agents have types, which relate to their costs of interaction. Types are independently and identically distributed according to a commonly known distribution. Its support is $\mathcal{C} = \{c_1, ..., c_m\} \subseteq (0, \infty)$, and the distribution is given by a probability vector $\boldsymbol{\pi}$, so that π_k is the probability of type c_k.

Timing All the strategic decisions take place at step 2; we break down the mechanics of the environment into several additional steps.

1. Each agent $i \in \mathcal{N}$ has his type[3] C^i drawn by nature and learns only his own C^i.
2. Simultaneously, each agent $i \in \mathcal{N}$ chooses a number $z^i \in [0, 1]$ called the *socializing effort*. Agents pay costs up-front for their effort, specified below.
3. The early social network is realized: we denote it by an n-by-n symmetric matrix $\mathbf{G}^E$. The indicator variable of the presence of the link $\{i, j\}$ is written $G^E_{ij} = G^E_{ji} \in \{0, 1\}$. The links form independently, with $\mathbf{P}(G^E_{ij} = 1) = p(z^i, z^j)$. The number $p(z^i, z^j)$ measures the probability that i and j become linked given their efforts. The assumptions made about the function $p : [0, 1] \times [0, 1] \to [0, 1]$, which is a parameter of the socializing technology, are discussed below.
4. Meetings take place between agents who do not know each other but are connected through mutual friends in the early-stage network. For every $i, j, \ell \in \mathcal{N}$ such that $G^E_{ij} = 0$ and $G^E_{i\ell} = G^E_{\ell j} = 1$, there is a Bernoulli random variable $M_{ij;\ell}$ which is, intuitively, the indicator of the event "i and j meet through the mutual friend ℓ". This variable takes the value 1 with probability $q > 0$, and 0 otherwise. The $M_{ij;\ell}$ are all independent.
5. The graph of late relationships $\mathbf{G}^L$ is realized by setting

$$G^L_{ij} = G^L_{ji} = \begin{cases} 1 & \text{if } G^E_{ij} = 0 \text{ and } M_{ij;\ell} = 1 \text{ for at least one } \ell \in \mathcal{N} \\ 0 & \text{otherwise.} \end{cases}$$

[3]We use subscripts to denote types and superscripts to denote random variables (and their realizations) that belong to an agent with a paritcular index.

The *final network* is $\mathbf{G}$, the sum of $\mathbf{G}^E$ and $\mathbf{G}^L$.

In this description, we have assumed that agents choose one socializing effort for the whole group. In Section 4, we enrich the game to one in which discrimination is allowed and show that, if there are small costs associated with seeking out particular agents, this assumption is not restrictive.

Preferences Agent i's costs take the form:

$$\frac{c^i}{\alpha}\left(\sum_{j\neq i} f\left(z^i, z^j\right)\right)^{\alpha}.$$

Here c^i, the type of agent i, is an agent-specific coefficient capturing the cost of social interaction. The number $f\left(z^i, z^j\right)$ measures the quantity of resources i spends interacting with j given their efforts. The assumptions made about the function $f : [0,1] \times [0,1] \to [0,\infty)$, which, like p, is a parameter of the socializing technology, are discussed below. Finally $\alpha > 1$ measures the convexity of resource costs.

Agent i gains a value[4] v_1 from any *early friend* (a j such that $G^E_{ij} = 1$) and a value v_2 from each *late friend* (a j such that $G^L_{ij} = 1$). We assume that $v_1 > v_2 \geq 0$. The difference in values comes from the extra time spent with the early friend. Thus, the utility of agent i after all the uncertainty is resolved can be written as

$$u^i(\mathbf{z}) = v_1 \cdot \#[\text{early friends}] + v_2 \cdot \#[\text{late friends}] - \frac{c^i}{\alpha}\left(\sum_{j\neq i} f\left(z^i, z^j\right)\right)^{\alpha}.$$

Parameters of the Socializing Technology The probability function for forming early links, $p : [0,1]^2 \to [0,1]$, is assumed to be an analytic[5], symmetric function of two variables, which is strictly increasing in both efforts in the interior of the unit square, and concave. We assume that a link cannot form between two agents if one of them is not investing any effort, so $p(0,x) = 0$ for any $x \in [0,1]$.[6] Finally, we assume that at 0 the cross partial of p is positive: $\frac{\partial^2}{\partial x \partial y} p(0,0) > 0$. This implies that for very low effort levels agents' efforts are complementary.

The resource function $f(x,y)$ is similarly assumed to be analytic. We require that an agent cannot impose costs unilaterally upon another agent who is not investing any effort in the relationship, which translates into: $f(0,x) = 0$ for any $x \in [0,1]$. We also assume that the marginal resources required to interact with an agent who does not invest are low ($f_1(0,0) = 0$), but increasing ($f_{11}(0,0) > 0$). We also assume that f is convex over its domain. This general formulation of the resource function allows for settings where agents' expended

[4]Long-term maintenance costs can be modeled by reducing the values of links appropriately.

[5]An analytic function is a function with a Taylor series which converges to it uniformly.

[6]This can be greatly generalized: for all the results to hold it suffices that the leading term in the Taylor expansion of $p(x,y)$ is xy. This allows for technologies where links can form with only unilateral investment.

resources depend only on their own efforts, but also for settings where these resources may depend on the interaction between the two efforts. Finally, we assume that $f_1(x, y) > 0$ if both $x, y > 0$, so that the marginal cost of additional own effort is strictly positive if both sides of the relationship invest effort.

3 Equilibrium and Welfare

We now turn to analyzing the model, first focusing on equilibrium behavior of the strategic agents, and then on its efficiency. All proofs are provided in the appendix.

3.1 Definitions and the Equilibrium Concept

A *pure strategy* for i is a vector $\boldsymbol{\sigma}^i \in [0,1]^m$ specifying how much effort i selects for every type he might be (recall m is the cardinality of the type space $\mathcal{C}$). We denote by σ_k^i the effort that the strategy $\boldsymbol{\sigma}^i$ prescribes for type c_k of agent i. A strategy profile $\boldsymbol{\sigma} = (\boldsymbol{\sigma}^i)_{i\in\mathcal{N}}$ is *symmetric* if σ_k^i does not depend on i, so that the action one plays depends only on one's type, not one's label.

We will be focusing on symmetric Bayesian Nash equilibria of the game, and we will ignore the uninteresting equilibrium, which exists for some specifications of the costs and benefits, in which everyone puts in no effort.

Definition. *The word "equilibrium" will mean, unless otherwise stated, "symmetric Bayesian Nash equilibrium different from the no-effort equilibrium", though sometimes we will emphasize these features in the statements of results.*

We will denote an equilibrium strategy for i by $\mathbf{x}^i$ and an equilibrium strategy profile by $\mathbf{x} = (\mathbf{x}^i)_{i\in\mathcal{N}}$. The notation $\mathbf{x}(n)$ will refer to an equilibrium of $\Gamma(n)$.

3.2 Equilibrium Existence

The first result establishes the existence of an equilibrium.

Theorem 1. *Fix $n \geq 4$. There exists a symmetric interior equilibrium of $\Gamma(n)$ in which all linking probabilities are positive. This is a strict equilibrium, in the sense that each agent has a unique best response.*

3.3 Equilibrium in Large Populations

We now analyze the properties of equilibria when n is large. For the remainder of this section, we focus on the case in which $f(x, y) = f(x)$: that is, the resources required to socialize depend only on one's own effort, and not on the effort of others. We believe analogues of most of the results could be obtained without this assumption, but it makes the analysis and intuition much simpler in places.

There will be two types of equilibria. In one regime, agents will have a number of friends that is of the same order as the population size. In another regime, they will have a number of friends that does not scale with the population size. To state this formally we define $F_k(\mathbf{x})$ to be the expected number of friends (degree) in the final network for an agent with cost type c_k when the equilibrium $\mathbf{x}$ is played. This allows us to formally define "high" and "low" equilibria:

Definition **1.** Define an equilibrium $\mathbf{x}$ to be β-sparse if $\max_k F_k(\mathbf{x}) \leq \beta$. For a given n, define an equilibrium $\mathbf{x}$ of $\Gamma(n)$ to be β-dense if $\min_k F_k(\mathbf{x}) \geq \beta n$.

With these definitions in hand, first we treat the most interesting case: the one in which $1 < \alpha < 2$. In this case, both high and low equilibria are possible. The key quantity for characterizing which can occur is the *value of friends of friends*. Formally defined as qv_2, this is the probability that i befriends j through a particular intermediary ℓ, multiplied by the value to i of the relationship with j, conditional on it being realized. The important comparison will be between qv_2 and a positive number called τ_{eq}, which depends on the parameters of the model other than q and v_2. It is defined by equation (12) in the appendix, and can be solved for explicitly.

The next theorem classifies the equilibria in the case $1 < \alpha < 2$. If $qv_2 \leq \tau_{\text{eq}}$, then there are both high- and low- effort equilibria. Otherwise, there are only high-effort ones. Figure 2 illustrates the large-sample result by plotting the equilibrium correspondence in a particular example.

Theorem **2.** *Assume $1 < \alpha < 2$. Then there exist some $\beta, \gamma > 0$ and some N so that for any $n \geq N$,*

1. *if $qv_2 \leq \tau_{eq}$, then every equilibrium of $\Gamma(n)$ is either γ-dense or β-sparse, and there is at least one of each kind.*

2. *if $qv_2 > \tau_{eq}$, then any nonzero equilibrium of $\Gamma(n)$ is γ-dense.*

Now we treat the case of highly convex costs, $\alpha > 2$. In this case, there are only low equilibria.

Theorem **3.** *Assume $\alpha > 2$. Then there exists some $\beta > 0$ and some N so that for any $n \geq N$, every equilibrium is β-sparse.*

Finally, we establish that every low-effort equilibrium has a simple structure: agents invest in inverse proportion to a power of their costs.

Theorem **4.** *Assume $\alpha > 1$. Fix $\beta > 0$. Then, for every $\bar{\epsilon} > 0$ there is some N so that if $n \geq N$ and $\mathbf{x}$ is a β-sparse equilibrium of $\Gamma(n)$, we have*

$$\frac{F_k(\mathbf{x})}{F_j(\mathbf{x})} = \left(\frac{c_j}{c_k}\right)^{\frac{1}{2\alpha-1}} + \epsilon$$

where $|\epsilon| < \bar{\epsilon}$.

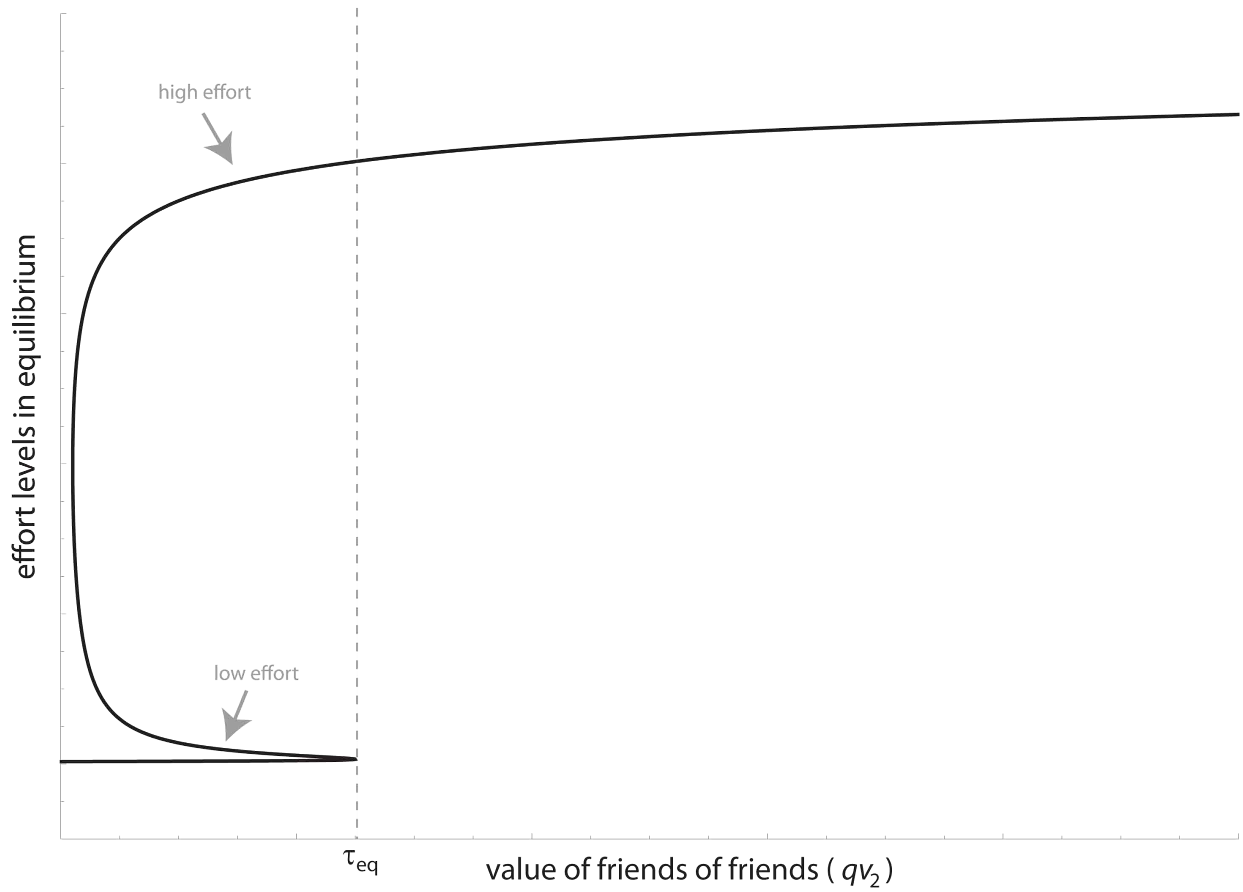

Figure 2: The equilibrium correspondence (in the case $1 < \alpha < 2$) as the value of friends of friends is varied. When $qv_2 \leq \tau_{\mathrm{eq}}$, there are low-effort equilibria as well as a high effort one. When $qv_2 > \tau_{\mathrm{eq}}$, then there is only a high-effort equilibrium.

In the case of low-effort equilibria, agents' degrees (numbers of friends) depend on costs of socializing in a way that can be precisely pinned down. In high-effort equilibria, the connection between costs and degrees is more subtle and is obtained by solving a system of nonlinear equations. Nevertheless, as shown in Lemma 2(2) of the appendix, agents with higher costs choose lower levels of effort in the high equilibrium, too.

3.4 Network Properties

The qualitative difference between the sizes of agents' neighborhoods in the two regimes results in dramatic differences in overall features of the network as a whole. To describe these differences we define the following terms: we say that a network $\mathbf{G}$ is *connected*, if for any two agents i, j there exists a sequence of agents $i_1, \ldots, i_\ell$ linking them, such that $G_{i,i_1} = 1$, $G_{i_\ell,j} = 1$, and for every $1 \leq k \leq \ell - 1$, $G_{i_k,i_{k+1}} = 1$; we say that agents i, j are at *distance k* in a network $\mathbf{G}$ if the shortest path connecting them in $\mathbf{G}$ is of length k; finally, given a network $\mathbf{G}$, we define the *diameter* of $\mathbf{G}$ to be the maximum distance between any two agents in $\mathbf{G}$.

Using classical results from the theory of random graphs, we characterize the macroscopic differences between the two regimes in the following result. We say a statement holds "asymptotically almost surely" (a.a.s.) if it holds with a probability that tends to 1 as n grows.

PROPOSITION **1.** *In the high-effort regime the realized social network is connected asymptotically almost surely, and the diameter of the network is between 2 and 3 asymptotically almost surely. In the low-effort regime the realized network is a.a.s. not connected.*

This result shows that the difference between high-effort and low-effort regimes yields sharp empirical predictions at the macroscopic level. High-effort regime networks are connected with a very high probability, so that any agent is linked, directly or indirectly, to any other agent. Moreover, with a probability that tends to 1, any two agents are at most three steps away from each other. Low-effort networks, on the other hand, are disconnected with arbitrarily high probability once they become large enough.

The proof of this result uses the asymptotic behavior of the linking probabilities $p(x_k, x_j)$; in particular, that these probabilities are decaying as n^{-1} in the low-effort regime and roughly as $n^{-1/2}$ in the high-effort regime. A result of Bollobás (2001) then allows us to characterize the diameter readily.

Proposition 1 implies, together with the previous results, that small changes in the exogenous parameters can cause dramatic differences in the large-scale properties of the resulting social networks. For example, a very slight increase in the probability of meeting friends of friends can lead to the network going from disconnected to very densely connected. As we will show below, this shift is also associated with a sharp rise in efficiency.

3.5 An Application: Social Networking Technology

These results can shed some light on the recent developments in social networking technologies, and specifically the dramatic rise and substantial impact of online social networks such as Facebook, MySpace, LinkedIn and Twitter. Hundreds of millions of people now use these networks regularly, spending, on average, hours a day on the sites (Boyd and Ellison, (2007); "Facebook: Statistics" (2010)). While these networks offer their users different and perhaps easier forms of connecting with friends, the direct benefits of using them (to browse photographs, exchange messages, etc.) are arguably similar to those of other technologies already in existence. It is clear, though, that these networks specifically and intentionally increase users' benefits from indirect friends. All of the above networks expose a user to the identities of friends of friends, usually providing some information about them, such as occupations, photos, hobbies and interests. Moreover, some of these tools, like LinkedIn, explicitly emphasize friends of friends by showing users how they can connect to certain individuals or organizations through their personal and professional social networks. In the model, this is exactly the kind of change that would push the formation of social networks beyond the critical threshold and into the high-effort regime, and even slight changes can make a big difference. Thus, the theory presented here provides one mechanism for the seemingly outsize impact of these technologies.

3.6 Welfare: Comparing Equilibria

Under the assumption that agents pay only for their own effort levels, there are only positive externalities in the game, and any equilibrium involves weakly too little effort. Despite this, there are huge differences in efficiency between the equilibria when both high- and low-effort equilibria coexist at the same parameter values. In particular, the following proposition follows immediately from the properties deduced in the appendix of the two types of equilibria.

PROPOSITION **2.** *Assume $1 < \alpha < 2$, and fix the $\beta, \gamma > 0$ guaranteed by Theorem 2. Choose any $\epsilon > 0$. Then there is an N so that for $n \geq N$, if $\mathbf{x}_L$ is a β-sparse equilibrium and $\mathbf{x}_H$ is a γ-dense equilibrium, the total utility under $\mathbf{x}_L$ is at most ϵ times that under $\mathbf{x}_H$.*

One important implication of this is that the system can exhibit path-dependence not only in local and large-scale network structure but also in welfare: interventions that move the levels of socializing without changing any underlying parameters can have lasting effects, either increasing or decreasing the welfare by huge factors. Structural estimation of the parameters (especially α) would be important for shedding light on whether this is possible in a given situation.

4 Mingling Evenly as an Equilibrium

In the description of our game, we assumed that agents choose one intensity for socializing within the group in general, without the possibility of discriminating. While this can be motivated as a reasonable restriction based on the difficulty of coordinating and focusing on specific others at the early stages of interactions, as in Cabrales, Calvó-Armengol, and Zenou (2009), we do not have to view this as a restriction. Indeed, we can enrich the model to one in which discrimination is allowed and show that, when there are small search costs, it is equilibrium behavior not to discriminate, but instead to mingle evenly.

To this end, define a new game $\tilde{\Gamma}(n)$. This game is the same as $\Gamma(n)$ except for two changes. Each agent's action is not determined merely a number z_k^i for each of his types, but rather by a set of numbers z_k^{ij} for each type, where j takes on all indices in $\mathcal{N}$ other than i. The probability that i and j are linked given their actions becomes $p(z^{ij}, z^{ji})$, and the resource costs paid become $f(z^{ij}, z^{ji})$, which are subject to the assumptions that we made in describing the model. The other difference is the utility function. We assume now that

$$u^i(\mathbf{z}) = v_1 \cdot \#\text{early friends} + v_2 \cdot \#\text{late friends} - \frac{c^i}{\alpha} \left(\sum_{j \neq i} f\left(z^{ij}, z^{ji}\right) \right)^\alpha - \Delta(\mathbf{z}^i).$$

Here $\Delta : [0,1]^{n-1} \to \mathbb{R}$ is the *discrimination cost*, capturing how difficult it is to set unequal levels of interaction. We assume that $\Delta(\mathbf{z}^i)$ is 0 when $\mathbf{z}^i$ is a constant and that Δ is a convex function, meaning that more evenly mixed interactions are cheaper.

The main result of this section is that when the curvature of Δ is not decaying too fast, as measured by certain conditions on its second derivatives, then socializing evenly is an equilibrium.

THEOREM **5.** *Assume that, for large enough n,*

$$\min_{j,\ell} \left| \frac{\partial \Delta}{\partial z^{ij} \partial z^{i\ell}} \right| \geq n^{-1/2} \log^7 n$$

and that the Hessian of Δ is positive definite. Consider a symmetric nonzero equilibrium of the no-discrimination game $\Gamma(n)$. Then, for n large enough, it is also an equilibrium of the game $\tilde{\Gamma}(n)$.

Using the magnitudes of the entries of the Hessian as a measure of the difficulty of discriminating is a "reduced-form" approach; the aim is not to develop a detailed micro-model of search costs. We would only like to point out that modest search frictions can suffice to ensure that agents find it optimal to interact evenly, so the assumption of even mingling is not too severe a restriction.

We believe that milder assumptions on Δ could give the same result, and we do not know whether not discriminating is an equilibrium for large n when there are no search frictions.

5 Concluding Remarks

This model of network formation with rational agents and uncertainty in the realization of links has two useful properties. First, the networks it predicts have the complex and irregular structure seen in real networks (Newman, 2003); moreover, they correspond to random network models with heterogeneous degrees recently developed in the probability literature (Chung and Lu, 2002; Chung et al., 2004). At the same time, the model does not rely on mechanistic foundations for link formation; the probabilities of links are endogenous choice variables that are selected when agents optimize, trading off the costs of socializing against the expected benefits. From a technical perspective, the fact that there is uncertainty over the precise realizations of the links enables the classification of equilibria into two simple kinds.

The main results of the paper serve as an illustration of the ways in which the simple framework can generate nontrivial predictions about how the economic fundamentals affect equilibrium and efficiency. In the particular application considered here, we showed that small changes in the value of friends-of-friends can change the orders of growth of social activity, the fundamental shapes of equilibrium networks, and the efficiency of outcomes. The framework is capable of accommodating other specifications of costs and benefits –

for instance, ones that have negative externalities or which involve more intricate network properties like transitivity.

It is important for the particular type of analysis we did that agents interact evenly within the population, without targeting their efforts at specific others. We showed in Section 4 that this can be equilibrium behavior given mild search frictions. However, it is not our intent to suggest that uniform mingling is always the reasonable model of relationship formation. Sometimes highly targeted interactions are much more relevant, as in international trade agreements. At other times, agents target their interactions, but do so randomly. Could it be the case that "randomly targeted" interactions yield results similar to the ones seen in this model? We view this as a potentially promising avenue for future theoretical work.

References

Ambrus, A., M. Mobius, and A. Szeidl (2010): “Consumption Risk-sharing in Social Networks,” Preprint, available at
http://www.econ.berkeley.edu/~szeidl/papers/risksharing.pdf

Anon. (2010): “Facebook: Statistics,”
http://www.facebook.com/press/info.php?statistics.

Bala, V. and S. Goyal (2000): “A Noncooperative Model of Network Formation,” *Econometrica*, 68, 1181–1229.

Bollobás, B. (2001): *Random Graphs, Second Edition*, Cambridge: Cambridge University Press.

Boyd, D. and N. B. Ellison (2007): “Social Network Sites: Definition, History, and Scholarship,” *Journal of Computer-Mediated Communication*, 13, 210–230.

Cabrales, A., A. Calvó-Armengol, and Y. Zenou (2009): “Social Interactions and Spillovers,” Preprint, available at http://www.ifn.se/web/yvesz.aspx

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

Christakis, N., J. Fowler, G. W. Imbens, and K. Kalyanaraman (2010): “An Empirical Model for Strategic Network Formation,” Working paper, available at
http://www.economics.harvard.edu/faculty/imbens

Chung, F. and L. Lu (2002): “The Average Distance in Random Graphs with Given Expected Degrees,” *Proceedings of National Academy of Science*, 99, 15879–15882.

Chung, F., L. Lu, and V. Vu (2004): “The Spectra of Random Graphs with Given Expected Degrees,” *Internet Mathematics*, 1, 257–275.

Currarini, S., M. O. Jackson, and P. Pin (2009a): “An Economic Model of Friendship: Homophily, Minorities and Segregation,” *Econometrica*, 77, 1003–45.

——— (2009b): “Identifying the roles of race-based choice and chance in high school friendship network formation,” *Proceedings of the National Academy of Sciences*, 107, 4857–4861.

Duflo, E. and E. Saez (2003): “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence From a Randomized Experiment,” *Quarterly Journal of Economics*, 118, 815–842.

Glaeser, E., B. Sacerdote, and J. Scheinkman (1996): “Crime and Social Interactions,” *Quarterly Journal of Economics*, 111, 507–548.

Granovetter, M. (2005): "The Impact of Social Structure on Economic Outcomes," *The Journal of Economic Perspectives*, 19, 33–50.

Hojman, D. and A. Szeidl (2008): "Core and Periphery in Networks," *Journal of Economic Theory*, 139, 295–309.

Jackson, M. O. (2005): "A Survey of Models of Network Formation: Stability and Efficiency," in *Group Formation in Economics: Networks, Clubs, and Coalitions*, ed. by G. Demange and M. Wooders, Cambridge: Cambridge University Press.

——— (2008): *Social and Economic Networks*, Princeton, N.J.: Princeton University Press.

Jackson, M. O. and B. W. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review*, 97, 890–915.

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

Moulin, H. (1986): *Game Theory for the Social Sciences, 2nd Ed.*, Studies in Game Theory and Mathematical Economics, New York, NY: New York University Press.

Myerson, R. B. (1991): *Game Theory: Analysis of Conflict*, Cambridge, MA: Harvard University Press.

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

Topa, G. (2001): "Social Interactions, Local Spillovers and Unemployment," *Review of Economic Studies*, 68, 261–295.