

NET Institute*

www.NETinst.org

Working Paper #10-21

September 2010

Strategic Random Networks

Benjamin Golub and Yair Livne
Stanford University

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, <http://www.NETinst.org>, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

Strategic Random Networks*

Benjamin Golub[†]

Yair Livne[‡]

September 30, 2010

Abstract

To study how economic fundamentals affect the formation of social networks, a model is needed that (i) has agents responding rationally to incentives (ii) can be taken to the data. This paper combines game-theoretic and statistical approaches to network formation in order to develop such a model. Agents spend costly resources 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 second stage of “meeting friends of friends”. The model predicts random graphs with tunable degree distributions and clustering, and characterizes how those statistics depend on the economic fundamentals. When the value of friends of friends is low, equilibrium networks can be either sparse or thick. But as soon as this value crosses a key threshold, the sparse equilibrium disappears 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 Abhijit Banerjee, Jeremy Bulow, Arun Chandrasekhar, Matthew Elliott, Alex Frankel, Alex Hirsch, André Henriques, Yuichiro Kamada, Ruth Kricheli, Antoine Lallour, Carlos Lever, James Liang, Pietro Majer, Michael Ostrovsky, and seminar participants 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 (www.NETinst.org). 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 and on the accidents of history? What are the welfare consequences? What are the effects of interventions?

An economic theory to address these questions should have two key properties. First, it should have rational agents responding to incentives. Second, its predictions should be consistent with observations — indeed, it should be possible to use observations to estimate the key economic parameters, which can then be used to understand how the observed phenomena came about and to perform policy analyses.

The goal of this paper is to develop such a theory. For the sake of realism and econometric usefulness, 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 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 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

	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.

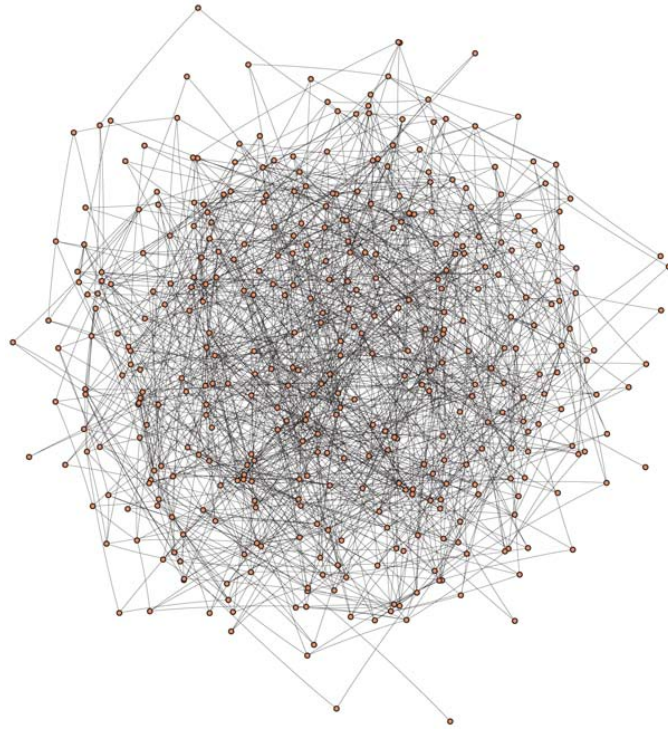
form them — effort which is sometimes futile due to accidents of fate. Again, we allow for a fairly general specification of the costs. Third, as first modeled by [Jackson and Rogers \(2007\)](#), there is both a random element to socializing (“meeting strangers”) and a natural source of dependence and 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.

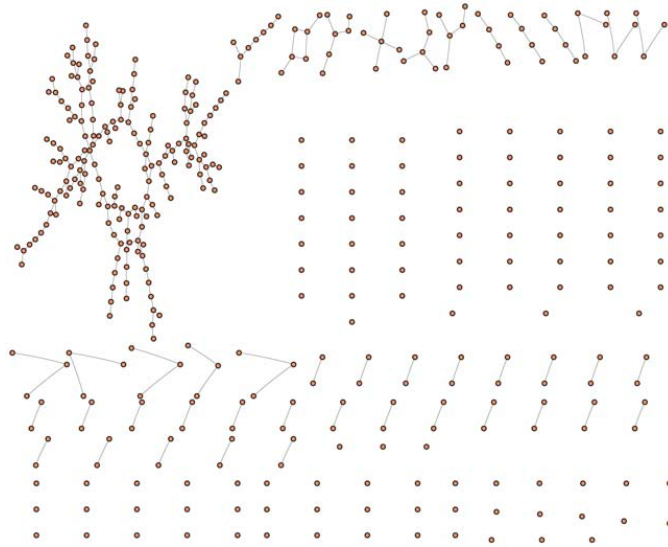
Adding best responses to a standard random graph model shows how network phenomena like the degree distribution and connectedness relate to economic fundamentals. It also reveals that there are completely new qualitative phenomena that arise when agents best-respond to each other in this setting.

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 friends¹ — 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

¹The 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.



(a)



(b)

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.

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 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 economics², and the study of rational network formation has a rich history. One strand of this literature, starting with [Myerson \(1991\)](#) and continuing with [Jackson and Wolinsky](#)

²Social 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 ([Calvo-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.

(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). This approach implicitly assumes that agents know the network insofar as that is important for their deviations, and delivers very specific and often stark predictions about network structure. While this has been an extremely important approach for understanding aspects of network formation, a different model is appropriate for the first-meeting setting that we focus on, as well as for generating the random graphs that are our equilibrium predictions. In our model, in contrast to these, any network has a positive probability of appearing in equilibrium, though some are much less likely than others; moreover, agents are fully aware of the randomness that generates this and take it into account when optimizing. This makes the present 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.

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 $\mathcal{N} = \{1, \dots, n\}$, 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 $\mathcal{C} = \{c_1, \dots, c_m\} \subseteq (0, \infty)$, 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 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 $G_{ij}^E = G_{ji}^E \in \{0, 1\}$. The links form independently, with $\mathbf{P}(G_{ij}^E = 1) = p(z^i, z^j)$. 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 $p : [0, 1] \times [0, 1] \rightarrow [0, 1]$, 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 $G_{ij}^E = 0$ and $G_{i\ell}^E = G_{\ell j}^E = 1$, 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

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

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:

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

Here c_i , the type of agent i , is an agent-specific coefficient capturing the cost of social interaction. The number $f(z^i, z^j)$ measures the quantity of resources i spends interacting with j given their efforts. The assumptions made about the function $f : [0, 1] \times [0, 1] \rightarrow [0, \infty)$, 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 value³ v_1 from any *early friend* (a j such that $G_{ij}^E = 1$) and a value v_2 from each *late friend* (a j such that $G_{ij}^L = 1$). 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

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

Parameters of the Socializing Technology The probability function for forming early links, $p : [0, 1]^2 \rightarrow [0, 1]$, is assumed to be an analytic⁴, 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

³Long-term maintenance costs can be modeled by reducing the values of links appropriately.

⁴An analytic function is a function with a Taylor series which converges to it uniformly.

effort, so $p(0, x) = 0$ for any $x \in [0, 1]$.⁵ 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 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 $\sigma^i \in [0, 1]^m$ 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 σ^i prescribes for type c_k of agent i . A strategy profile $\sigma = (\sigma^i)_{i \in \mathcal{N}}$ 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 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 $\Gamma(n)$.

⁵This 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.

3.2 Equilibrium Existence

In this game there may exist a trivial symmetric equilibrium – the one in which all agents invest effort level 0. The first result establishes the existence of a more interesting 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 -low if $\max_k F_k(\mathbf{x}) \leq \beta$. For a given n , define an equilibrium \mathbf{x} of $\Gamma(n)$ to be β -high 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

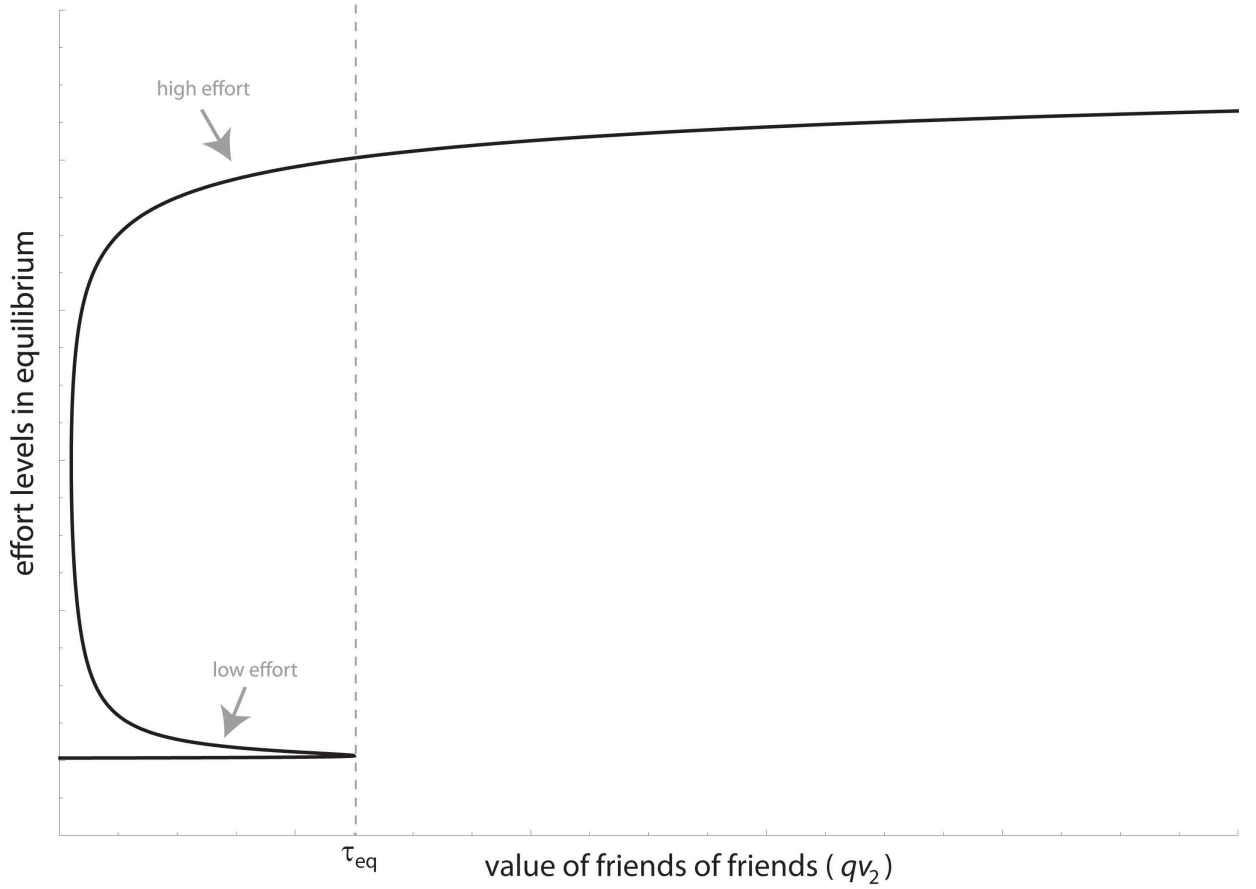


Figure 2: The equilibrium correspondence (in the case $1 < \alpha < 2$) as the value of friends of friends is varied. When $qv_2 \leq \tau_{eq}$, there are low-effort equilibria as well as a high effort one. When $qv_2 > \tau_{eq}$, then there is only a high-effort equilibrium.

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 -high or β -low, and there is at least one of each kind.*
2. *if $qv_2 > \tau_{eq}$, then any nonzero equilibrium of $\Gamma(n)$ is γ -high.*

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 -low.*

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 $\beta > 0$ and $\bar{\epsilon} > 0$. Then there is some N so that if $n \geq N$ and \mathbf{x} is a β -low 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 $|\epsilon| < \bar{\epsilon}$.

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, \dots, i_ℓ 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 -low equilibrium and \mathbf{x}_H is a γ -high 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 history-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

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

Here $\Delta : [0, 1]^{n-1} \rightarrow \mathbb{R}$ 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 , $\min_{j,\ell} \left| \frac{\partial \Delta}{\partial z^{ij} \partial z^{\ell\ell}} \right| \geq n^{-1/2} \log^7 n$ 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. PATAACCHINI, 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.

Appendix A: Proofs

Existence of Equilibrium

Proof of Theorem 1 The strategy of the proof is as follows. We define an auxiliary game $\Gamma_\delta(n)$ in which every type is constrained to play effort at least δ . Then: (1) We use a standard existence theorem to obtain the existence of a symmetric equilibrium in this game. (2) We verify that for some small enough δ^* , no type is playing effort level δ^* in $\Gamma_{\delta^*}(n)$. (3) Then we show that this equilibrium survives as δ gets smaller than δ^* .

An important ingredient in this is the following lemma, which we prove at the end.

LEMMA 1. *For any strategies of the other players, the utility function of each player is concave in his own action \mathbf{z}^i .*

In addition to this, $\Gamma_\delta(n)$ is a symmetric game with convex, compact strategy spaces. Nash's theorem for symmetric games (Moulin, 1986, p. 115) furnishes the existence of a symmetric equilibrium $\mathbf{x}^{(\delta)}$ of $\Gamma_\delta(n)$. This gives step (1). Suppose now we can show (2) above, that the equilibrium is away from the lower boundary. Step (3) then also follows from the lemma. In particular: since utility is concave in each z_i^k , each player's marginal utility is decreasing in his own action. Thus, i 's marginal utility in $\Gamma_{\delta^*}(n)$ at $\mathbf{x}^{(\delta)}$ is 0, or he is playing 1 (since we know he is not at the lower boundary). In either case, that response remains optimal in $\Gamma_{\delta^*}(n)$, since the first order condition remains the same and no binding constraint was relaxed.

Now we focus on step (2). Recall that the marginal costs per other agent for an agent of type k is given in equilibrium by:

$$\frac{1}{n-1} \text{MC}_k = \frac{c_k}{n-1} \mathbf{E}_{Y_1, \dots, Y_{n-1}} \left[\left(\sum_{j=1}^{n-1} f_1(x_k, Y_j) \right) \left(\sum_{j=1}^{n-1} f(x_k, Y_j) \right)^{\alpha-1} \right] \quad (1)$$

where Y_1, \dots, Y_{n-1} are i.i.d random variables taking for each $j \in \{1, \dots, m\}$ the value x_j with probability π_j . The normalized marginal benefit for type k in equilibrium is given by:

$$\frac{1}{n-1} \text{MB}_k = \mathbf{E}_Y \left[(v_1 - v_2) P' + v_2 (1 - qR)^{n-3} (P'(1 - qR) + (1 - p)(n-2)qR') \right]$$

with:

$$\begin{aligned} P &= p(x_k, Y) & P' &= p_1(x_k, Y) \\ R &= \mathbf{E}_W [p(x_k, W) p(W, Y) | Y] & R' &= \mathbf{E}_W [p_1(x_k, W) p(W, Y) | Y] \end{aligned}$$

where the W_j are also i.i.d random variables taking for each $j \in \{1, \dots, m\}$ the value x_j with probability π_j .

We claim that if $\mathbf{x}^{(\delta)}$ is a symmetric equilibrium of $\Gamma_\delta(n)$, then $\max_k x_k^\delta$ cannot become arbitrarily small. Suppose otherwise. Using Taylor expansion, we verify that marginal benefits are bounded below by a term linear in $\max_k x_k^\delta$, whereas marginal costs are bounded above by a term that decays faster than linearly in $\max_k x_k^\delta$. Thus, there is some lower bound $\underline{\epsilon} > 0$ so that for all $\delta < \underline{\epsilon}$, there is some type playing at least $\underline{\epsilon}$ in any equilibrium of $\Gamma_\delta(n)$. But then it follows that, once δ is small enough, *no* type can be playing δ , because marginal benefits are bounded below by something linear in $\underline{\epsilon}$ (and independent of δ), whereas marginal costs are decreasing with δ . This completes the second step.

It only remains to prove Lemma 1.

Proof of Lemma 1. It suffices to prove that the action is concave in each z_k^i because the overall expected utility is just a linear combination taken over the various types. Fix a cost type k for agent i and compute all expectations taking this as known. Let Y_j denote the random variable capturing the effort level of agent j . Write

$$\mathbf{E}[u_i(\mathbf{z}^i)] = u_i^E(\mathbf{z}^i; \mathbf{x}^{-i}) + u_i^L(\mathbf{z}^i; \mathbf{x}^{-i}) - u_i^C(\mathbf{z}^i; \mathbf{x}^{-i})$$

where:

$$u_i^E = v_1 \mathbf{E} \left[\sum_{j \neq i} p(z_k^i, Y_j) \right];$$

$$u_i^L = v_2 \mathbf{E} \left[\sum_{j \neq i} [1 - p(z_k^i, Y_j)] \sum_{L \subset N \setminus \{i, j\}} [1 - (1 - q)^{|L|}] \prod_{\ell \in L} [p(z_k^i, Y_\ell) p(Y_\ell, Y_j)] \prod_{\substack{\ell \notin L \\ \ell \neq i, j}} [1 - p(z_k^i, Y_\ell) p(Y_\ell, Y_j)] \right];$$

and

$$u_i^C = \mathbf{E} \left[\frac{c_i}{\alpha} \left(\sum_{j \neq i} f(z_k^i, Y_j) \right)^\alpha \right].$$

Here we have split the expected utility of agent i into three pieces: the benefits coming from early-stage friends, the benefits coming from late-stage friends, and the costs. We know u_i^E is concave because p is concave, and we know $-u_i^C$ is concave because f is convex and $\alpha > 1$. So it remains only to deal with u_i^L .

We do this by conditioning on the realizations of all early links not involving i , a network we call \mathbf{G}_{-i}^E . If each of these conditional expectations is concave in z_k^i , the same is true of the unconditional expectation. Let N_j be the neighborhood of agent j in this network. Then

$$\mathbf{E}[u_i(z_k^i) | \mathbf{G}_{-i}^E] = \mathbf{E} \left[\sum_j 1_{G_{ij}^E=0} \left[1 - \prod_{\ell \in N_j} (1 - qp(z_k^i, Y_\ell)) \right] \middle| \mathbf{G}_{-i}^E \right].$$

We will show the concavity of the expression inside the summation for each separate j . Note that in treating each such term, we may condition on $G_{ij}^E = 0$, because the other realization contributes nothing to the expectation. In that case, we just have to show the concavity of

$$\mathbf{E} \left[1 - \prod_{\ell \in N_j} (1 - qp(z_k^i, Y_\ell)) \middle| \mathbf{G}_{-i}^E, G_{ij}^E = 0 \right].$$

But this is concave because $\prod_{\ell \in N_j} (1 - qp(z_k^i, Y_\ell))$ is convex in z_k^i : why the latter? Because it is the product of nonnegative functions with negative first derivatives and positive second derivatives in z_k^i — one checks by an elementary calculus exercise that this suffices to guarantee convexity. \square

This completes the proof of Theorem 1. ■

Large Population Analysis

We lay out a sequence of technical lemmas from which the proofs of Theorems 2, 3, and 4 follow directly.

The notation \mathbf{c} will refer to the vector of types, ordered from least to greatest cost.

LEMMA 2. *Let $x(n)$ be a sequence of m -vectors of equilibrium intensities, such that x_k is the level of effort invested by an agent with cost coefficient c_k in a mingling equilibrium with n agents, for every $k \in \{1, \dots, m\}$. Then:*

1. $\lim_{n \rightarrow \infty} x_k(n) = 0$.
2. *Agents' investments decay at the same rate, so that there exists constants $0 < d < D < \infty$ such that for any j, k and any n :*

$$d < \frac{x_j(n)}{x_k(n)} < D$$

3. *Assuming that $f(x, y) = f(x)$, agents with lower cost coefficients choose higher effort levels in equilibrium, so far large enough n :*

$$x_1(n) > \dots > x_m(n)$$

Proof. For the first part of the claim, assume that:

$$\limsup_{n \rightarrow \infty} x_k(n) = \max_{1 \leq j \leq m} \limsup_{n \rightarrow \infty} x_j(n) = \epsilon > 0$$

for some $k \in \{1, \dots, m\}$. Slightly abusing notation, restrict to a subsequence so we can write $\lim x(n) = \epsilon$, and let us drop the notation describing the dependence on n for the rest of this proof. Also, perhaps by further restricting to a subsequence, assume that for any n , $x_k \geq \max_{1 \leq j \leq m} x_j$. Recall that the marginal costs per other agent for an agent of type k is given in equilibrium by:

$$\frac{1}{n-1} \text{MC}_k = \frac{c_k}{n-1} \mathbf{E}_{Y_1, \dots, Y_{n-1}} \left[\left(\sum_{j=1}^{n-1} f_1(x_k, Y_j) \right) \left(\sum_{j=1}^{n-1} f(x_k, Y_j) \right)^{\alpha-1} \right] \quad (2)$$

where Y_1, \dots, Y_{n-1} are i.i.d random variables taking for each $j \in \{1, \dots, m\}$ the value x_j with probability π_j . We first show that the normalized marginal cost diverges to infinity at a rate which at least polynomial. By the strong law of large numbers, the probability r_n that a proportion of at least $\pi_k/2$ of the random variables Y_1, \dots, Y_{n-1} get the value x_k converges to 1. Therefore:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n-1} \text{MC}_k &\geq \lim_{n \rightarrow \infty} \frac{c_k r_n}{n-1} \left(\frac{\pi_k}{2} (n-1) f_1(x_k, x_k) \right) \left(\frac{\pi_k}{2} (n-1) f(x_k, x_k) \right)^{\alpha-1} \\ &= \lim_{n \rightarrow \infty} (n-1)^{\alpha-1} c_k \left(\frac{\pi_k}{2} \right)^\alpha f_1(\epsilon, \epsilon) f(\epsilon, \epsilon)^{\alpha-1} \end{aligned}$$

where the last equality follows from $r_n \rightarrow 1$ and since f is a smooth function. Since we assume that f_1 is a strictly positive function for values both bounded away from 0, and since $\alpha > 1$ we have that the marginal cost diverges to infinity at a rate which at least polynomial.

Recall that the normalized marginal benefit for type k in equilibrium is given by:

$$\frac{1}{n-1} \text{MB}_k = \mathbf{E}_Y \left[(v_1 - v_2) P' + v_2 (1 - qR)^{n-3} (P'(1 - qR) + (1 - p)(n-2)qR') \right]$$

with:

$$\begin{aligned} P &= p(x_k, Y) & P' &= p_1(x_k, Y) \\ R &= \mathbf{E}_W [p(x_k, W) p(W, Y) | Y] & R' &= \mathbf{E}_W [p_1(x_k, W) p(W, Y) | Y] \end{aligned}$$

where W, Y are i.i.d random variables taking for each $j \in \{1, \dots, m\}$ the value x_j with probability π_j . Denote $\bar{p}_1 = \max_{x, y \in [0, 1]} p_1(x, y)$. This maximum exists since p is a smooth function. Using this notation:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n-1} \text{MB}_k \\ & \leq \lim_{n \rightarrow \infty} (v_1 - v_2) \bar{p}_1 + v_2 \bar{p}_1 \mathbf{E}_Y \left[(1 - q \mathbf{E}_W [p(x_k, W) p(W, Y) | Y])^{n-3} (1 + (n-2)q \mathbf{E}_W [p(W, Y) | Y]) \right] \end{aligned}$$

Since p is an increasing function and since $x_k \geq x_j$ for any j :

$$\begin{aligned} & \leq \lim_{n \rightarrow \infty} (v_1 - v_2) \bar{p}_1 + v_2 \bar{p}_1 \mathbf{E}_Y \left[(1 - q \pi_k p(x_k, x_k) p(x_k, Y))^{n-3} (1 + (n-2)q p(x_k, Y)) \right] \\ & = C + \lim_{n \rightarrow \infty} v_2 \bar{p}_1 \mathbf{E}_Y \left[(1 - q \pi_k p(x_k, x_k) p(x_k, Y))^{n-3} (n-2)q p(x_k, Y) \right] \\ & = C + \lim_{n \rightarrow \infty} v_2 \bar{p}_1 \mathbf{E}_Y [(1 - q \pi_k p(x_k, x_k) p(x_k, Y))^n n q p(x_k, Y)] \end{aligned}$$

for some constant C . Elementary analysis shows that the expression inside the expectation, as a function of the variable $y = p(x_k, Y)$, is maximized when $y = \frac{1}{(n+1)\pi_k q p(x_k, x_k)}$, so the above expression is bounded above by:

$$\leq C + \lim_{n \rightarrow \infty} v_2 \bar{p}_1 \left(1 - \frac{1}{n+1} \right)^{n-3} \frac{n}{(n+1)\pi_k p(x_k, x_k)} = C + \frac{v_2 \bar{p}_1}{e \pi_k p(\epsilon, \epsilon)} < \infty$$

This shows that the marginal benefits for an agent of type k are asymptotically bounded. This cannot hold in equilibrium where the agent interacts with other agents with a positive intensity, and thus we have reached a contradiction. We can therefore conclude that for any $k \in \{1, \dots, m\}$ it holds that $x_k \rightarrow 0$.

For the second part of the proof, let $x(n)$ be a sequence of equilibrium intensities and assume that for some $1 \leq j_1, j_2 \leq m$ it holds that:

$$\liminf_{n \rightarrow \infty} \frac{x_{j_1}}{x_{j_2}} = 0$$

Perhaps by switching to a subsequence, we can assume without loss of generality that this holds as an actual limit. Because the number of permutations over a finite set is finite, we can also assume without loss of generality, perhaps by switching to a further subsequence, that along this subsequence $x_1 \leq x_2 \leq \dots \leq x_m$. This implies that $\lim x_1/x_m = 0$.

By the previous part of the proof, we know for that all large enough n , the FOC holds with equality.

Using the same notation as before, for every $k \in \{1, \dots, m\}$ and for any large enough n , we have that:

$$\begin{aligned}
1 &= \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[(v_1 - v_2) P' + v_2 (1 - qR)^{n-3} (P' (1 - qR) + (1 - p)(n - 2)qR') \right]}{\frac{c_k}{n-1} \mathbf{E}_{Y_1, \dots, Y_{n-1}} \left[\left(\sum_{j=1}^{n-1} f_1(x_k, Y_j) \right) \left(\sum_{j=1}^{n-1} f(x_k, Y_j) \right)^{\alpha-1} \right]} \\
&= \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[(v_1 - v_2) P' + v_2 (1 - qR)^{n-3} (P' + (n - 2)qR') \right]}{\frac{c_k}{n-1} \mathbf{E}_{Y_1, \dots, Y_{n-1}} \left[\left(\sum_{j=1}^{n-1} f_1(x_k, Y_j) \right) \left(\sum_{j=1}^{n-1} f(x_k, Y_j) \right)^{\alpha-1} \right]} \quad (3)
\end{aligned}$$

We now aim to “get rid” of the expectations in this limit. Starting with the denominator, note that since f, f_1 are increasing functions:

$$\lim_{n \rightarrow \infty} \frac{1}{n-1} MC_k \leq \lim_{n \rightarrow \infty} \frac{1}{n-1} \left[\left(\sum_{j=1}^{n-1} f_1(x_k, x_m) \right) \left(\sum_{j=1}^{n-1} f(x_k, x_m) \right)^{\alpha-1} \right]$$

Similarly, by the strong law of large numbers the probability r_n that a proportion of at least $\pi_m/2$ of the random variables take the value x_m satisfies $r_n \rightarrow 1$. Hence:

$$\lim_{n \rightarrow \infty} \frac{1}{n-1} MC_k \geq \lim_{n \rightarrow \infty} \frac{\pi_m}{n-1} \left[\left(\sum_{j=1}^{n-1} f_1(x_k, x_m) \right) \left(\sum_{j=1}^{n-1} f(x_k, x_m) \right)^{\alpha-1} \right]$$

We can thus write:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n-1} MC_k &= \lim_{n \rightarrow \infty} \frac{h_n^1}{n-1} \left[\left(\sum_{j=1}^{n-1} f_1(x_k, x_m) \right) \left(\sum_{j=1}^{n-1} f(x_k, x_m) \right)^{\alpha-1} \right] \\
&= \lim_{n \rightarrow \infty} h_n^1 (n-1)^{\alpha-1} f_1(x_k, x_m) f(x_k, x_m)^{\alpha-1} \quad (4)
\end{aligned}$$

for a sequence $h_{n,k}^1 \leq 1$ with $\liminf h_{n,k}^1 > 0$.

Turning to the numerator, by the same type of argument:

$$\lim_{n \rightarrow \infty} \mathbf{E}_Y (v_1 - v_2) P' = \lim_{n \rightarrow \infty} h_n^2 p_1(x_k, Y) \quad (5)$$

with $h_n^2 < v_1 - v_2$ and $\liminf h_{n,k}^2 > 0$, and similarly:

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \mathbf{E}_Y \left[v_2 (1 - qR)^{n-3} (P' + (n - 2)qR') \right] \\
&= \lim_{n \rightarrow \infty} \mathbf{E}_Y \left[\left(1 - h_{n,k}^3 p(x_k, x_m) p(x_m, Y) \right)^n (h_{n,k}^4 p_1(x_k, Y) + h_{n,k}^5 n p_1(x_k, x_m) p(x_m, Y)) \right] \quad (6)
\end{aligned}$$

where $h_{n,k}^3, h_{n,k}^4, h_{n,k}^5 < M$ for some $M < \infty$ and $\liminf h_{n,k}^i > 0$ for each $2 \leq i \leq 5$. We again note that

these sequences may depend on k .

We now want to use the first part of the lemma, combined with the analyticity of p, f to get more control on these expressions. Since f and p are both analytic functions, we can write:

$$f(x, y) = \sum_{i,j \geq 0} a_{ij} x^i y^j \quad p(x, y) = \sum_{i,j \geq 0} b_{ij} x^i y^j$$

Since we assume that costs cannot be imposed upon others, we have that $a_{0j} = 0$ for any $j \geq 0$. We also directly assume that $a_{10} = 0$, and that either $a_{11} > 0$ or $a_{20} > 0$. By Lagrange's formula for the remainder, this allows us to describe f in a small enough neighborhood of 0 by:

$$f(x, y) = F_1(x, y)x^2 + F_2(x, y)xy$$

where F_1, F_2 are positive bounded functions, with at least one of them bounded away from 0, and with:

$$\lim_{(x,y) \rightarrow 0} F_1(x, y) = a_{20} \quad \lim_{(x,y) \rightarrow 0} F_2(x, y) = a_{11}.$$

Similarly, we can write (in a neighborhood of 0):

$$f_1(x, y) = F_3(x, y)x + F_4(x, y)y$$

with F_3, F_4 positive and bounded and at least one of them bounded away from 0 (corresponding to this property for F_1, F_2), and with:

$$\lim_{(x,y) \rightarrow 0} F_3(x, y) = 2a_{20} \quad \lim_{(x,y) \rightarrow 0} F_4(x, y) = a_{11}.$$

Similarly, for p , since by our assumptions it follows that $b_{00} = b_{10} = b_{01} = b_{20} = b_{02} = 0$ and $b_{11} > 0$, we can write in a neighborhood of 0:

$$p(x, y) = P_1(x, y)xy$$

with $P_1(x, y)$ strictly positive, bounded and bounded away from 0, and with:

$$\lim_{(x,y) \rightarrow 0} P_1(x, y) = b_{11}$$

Similarly, we have that:

$$p_1(x, y) = P_2(x, y)y$$

with P_2 having the same properties. Plugging all this into the FOC together with (4), (5) and (6), gives:

$$1 = \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[\tilde{h}_{n,k}^2 Y + \left(1 - \tilde{h}_{n,k}^3 x_k (x_m)^2 Y \right)^n \left(\tilde{h}_{n,k}^4 Y + \tilde{h}_{n,k}^5 (x_m)^2 Y \right) \right]}{(n-1)^{\alpha-1} \left(f_{n,k}^3 x_k + f_{n,k}^4 x_m \right) \left(f_{n,k}^1 (x_k)^2 + f_{n,k}^2 x_k x_m \right)^{\alpha-1}} \quad (7)$$

where:

$$\tilde{h}_{n,k}^2 = \frac{h_{n,k}^2}{h_{n,k}^1} P_1(x_k, Y) \quad \tilde{h}_{n,k}^3 = h_{n,k}^3 P(x_k, x_m) P(x_m, Y)$$

$$\tilde{h}_{n,k}^4 = \frac{h_{n,k}^4}{h_{n,k}^1} P_1(x_k, Y) \quad \tilde{h}_{n,k}^5 = \frac{h_{n,k}^5}{h_{n,k}^1} n P_1(x_k, x_m) P(x_m, Y)$$

are bounded positive random variables which are also strictly bounded away from 0, and:

$$f_{n,k}^i = F_i(x_k, x_m) \text{ for } 1 \leq i \leq 4$$

Consider now the limit in (7) for $k = m$:

$$\begin{aligned} 1 &= \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[\tilde{h}_{n,m}^2 Y + \left(1 - \tilde{h}_{n,m}^3 (x_m)^3 Y \right)^n \left(\tilde{h}_{n,m}^4 Y + \tilde{h}_{n,m}^5 (x_m)^2 Y \right) \right]}{(n-1)^{\alpha-1} \left((f_{n,m}^3 + f_{n,m}^4) x_m \right) \left((f_{n,m}^1 + f_{n,m}^2) (x_m)^2 \right)^{\alpha-1}} \\ &\leq \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[\tilde{h}_{n,m}^2 Y + \left(1 - \tilde{h}_{n,m}^3 x_1 (x_m)^2 Y \right)^n \left(\tilde{h}_{n,m}^4 Y + \tilde{h}_{n,m}^5 (x_m)^2 Y \right) \right]}{(n-1)^{\alpha-1} \left((f_{n,m}^3 + f_{n,m}^4) x_m \right) \left((f_{n,m}^1 + f_{n,m}^2) (x_m)^2 \right)^{\alpha-1}} \\ &= C \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[\tilde{h}_{n,1}^2 Y + \left(1 - \tilde{h}_{n,1}^3 x_1 (x_m)^2 Y \right)^n \left(\tilde{h}_{n,1}^4 Y + \tilde{h}_{n,1}^5 (x_m)^2 Y \right) \right]}{(n-1)^{\alpha-1} \left((f_{n,1}^3 + f_{n,1}^4) x_m \right) \left((f_{n,1}^1 + f_{n,1}^2) (x_m)^2 \right)^{\alpha-1}} \\ &\leq C \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[\tilde{h}_{n,1}^2 Y + \left(1 - \tilde{h}_{n,1}^3 x_1 (x_m)^2 Y \right)^n \left(\tilde{h}_{n,1}^4 Y + \tilde{h}_{n,1}^5 (x_m)^2 Y \right) \right]}{(n-1)^{\alpha-1} \left(f_{n,1}^3 x_1 + f_{n,1}^4 x_m \right) \left((f_{n,1}^1 x_1 + f_{n,1}^2 x_m) x_m \right)^{\alpha-1}} \\ &= C \lim_{n \rightarrow \infty} \frac{\mathbf{E}_Y \left[\tilde{h}_{n,1}^2 Y + \left(1 - \tilde{h}_{n,1}^3 x_1 (x_m)^2 Y \right)^n \left(\tilde{h}_{n,1}^4 Y + \tilde{h}_{n,1}^5 (x_m)^2 Y \right) \right]}{(n-1)^{\alpha-1} \left(f_{n,1}^3 x_1 + f_{n,1}^4 x_m \right) \left(f_{n,1}^1 (x_1)^2 + f_{n,1}^2 x_1 x_m \right)^{\alpha-1}} \lim_{n \rightarrow \infty} \left(\frac{x_1}{x_m} \right)^{\alpha-1} \\ &= C \cdot 1 \cdot 0 \\ &= 0 \end{aligned}$$

where the inequalities are due to $x_1 \leq x_m$, the equality between the two inequalities, for an appropriate choice of a finite, positive constant C , follows from the fact that the (random) sequences $f_{n,k}^i, \tilde{h}_{n,k}^i$ are bounded and bounded away from 0, and the one before last equality follows from (7) for $k = 1$. This is a contradiction, which ends the proof of the second part.

For the third part, assume that $f(x, y) = f(x)$, so an agent's costs are only dependent on his or hers own effort. This allows us to write the asymptotic FOC from (3) as:

$$c_k = \lim_{n \rightarrow \infty} \frac{b_{11}(v_1 - v_2) \mathbf{E}[Y] + v_2 \left(1 - q b_{11}^2 x_k \mathbf{E}[W^2] Y \right)^n \left(b_{11} \mathbf{E}[Y] + n q b_{11}^2 \mathbf{E}[Y] \mathbf{E}[Y^2] \right)}{n^{\alpha-1} 2 a_{20}^{\alpha-1} x_k^{2\alpha-1}} \quad (8)$$

For a fixed n , we can think of the RHS of the above expression as a function of x_k (holding the expectations fixed), and as such it is a strictly decreasing function of x_k . Since the LHS is simply the cost coefficient c_k , we must have that for large enough n lower cost coefficients go with higher equilibrium efforts. \square

LEMMA 3. Assume that $f(x, y) = f(x)$, and that $\alpha \neq 2$. Let $x(n)$ be a sequence of m -vectors of equilibrium intensities, then for every k :

1. it holds that:

$$\liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{2}} > 0$$

and if also

$$\liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{2}} = \infty$$

then

$$\liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{4}} > 0$$

2. For every $\epsilon > 0$:

$$\limsup_{n \rightarrow \infty} x_k(n) n^{\frac{1}{4} - \epsilon} = 0$$

and if

$$\limsup_{n \rightarrow \infty} x_k(n) n^{\frac{1}{4}} = 0$$

then

$$0 < \liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{2}} < \limsup_{n \rightarrow \infty} x_k(n) n^{\frac{1}{2}} < \infty$$

Proof. For the first part, begin by assuming that $\liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{2}} = 0$. Consider the RHS of (8) along the subsequence where the lim sup holds, still indexing it with n . By the assumption and the results of Lemma 2, we have that $(1 - qb_{11}^2 x_k \mathbf{E}[W^2] Y)^n \rightarrow 1$, so (8) simplifies to:

$$c_k = \lim_{n \rightarrow \infty} \frac{b_{11} v_1 \mathbf{E}[Y] + v_2 n q b_{11}^2 \mathbf{E}[Y] \mathbf{E}[Y^2]}{n^{\alpha-1} 2a_{20}^\alpha x_k^{2\alpha-1}} = \lim_{n \rightarrow \infty} \frac{b_{11} v_1 \frac{\mathbf{E}[Y]}{x_k} + v_2 n q b_{11}^2 \frac{\mathbf{E}[Y]}{x_k} \mathbf{E}[Y^2]}{2a_{20}^\alpha (nx_k^2)^{\alpha-1}} \quad (9)$$

Since $\alpha > 1$ and $\lim_{n \rightarrow \infty} nx_k^2 = 0$, the denominator of the RHS in the above goes to 0. However, by Lemma 2, the first term in the numerator is bounded from below, while the second term is nonnegative. Thus, the entire fraction goes to ∞ , which is a contradiction. This establish that $\liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{2}} > 0$.

Assume now that $\liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{2}} = \infty$, while $\liminf_{n \rightarrow \infty} x_k(n) n^{\frac{1}{4}} = 0$. Rewriting (8), we get:

$$c_k = \lim_{n \rightarrow \infty} \frac{b_{11}(v_1 - v_2) \frac{\mathbf{E}[Y]}{x_k} + v_2 (1 - qb_{11}^2 x_k \mathbf{E}[W^2] Y)^n \left(b_{11} \frac{\mathbf{E}[Y]}{x_k} + n q b_{11}^2 \frac{\mathbf{E}[Y]}{x_k} \mathbf{E}[Y^2] \right)}{2a_{20}^\alpha (nx_k^2)^{\alpha-1}}$$

By the first assumption on the asymptotic behavior of x_k , and since $\alpha > 1$ we have that the denominator goes to ∞ , at least when switching to a partial limit. As before, the first term in the numerator is bounded. For the second term:

$$\begin{aligned} & \lim_{n \rightarrow \infty} v_2 (1 - qb_{11}^2 x_k \mathbf{E}[W^2] Y)^n \left(b_{11} \frac{\mathbf{E}[Y]}{x_k} + n q b_{11}^2 \frac{\mathbf{E}[Y]}{x_k} \mathbf{E}[Y^2] \right) \\ &= \lim_{n \rightarrow \infty} v_2 \left(b_{11} \frac{\mathbf{E}[Y]}{x_k} + n q b_{11}^2 \frac{\mathbf{E}[Y]}{x_k} \mathbf{E}[Y^2] \right) \\ &= v_2 b_{11} \lim_{n \rightarrow \infty} \frac{\mathbf{E}[Y]}{x_k} n \mathbf{E}[Y^2] \end{aligned}$$

where the first equality follows from the second assumption on the asymptotic behavior of x_k . Thus, the

following limit holds:

$$c_k = \lim_{n \rightarrow \infty} \frac{v_2 b_{11} \frac{\mathbf{E}[Y]}{x_k} n \mathbf{E}[Y^2]}{2a_{20}^\alpha (nx_k^2)^{\alpha-1}}$$

which since $\alpha \neq 2$, is a contradiction, as the numerator and denominator have different rates of divergence. This completes the first part of the proof.

For the second part of the proof, assume that there exists some $\epsilon > 0$ such that $\limsup_{n \rightarrow \infty} x_k(n) n^{\frac{1}{4}-\epsilon} > 0$. We again switch to a subsequence where this occurs, and there this assumption implies that:

$$\lim (1 - qb_{11}^2 x_k \mathbf{E}[W^2] Y)^n = 0$$

and this term decays exponentially fast in n . Thus, the FOC in (8) reduces to:

$$c_k = \lim_{n \rightarrow \infty} \frac{b_{11}(v_1 - v_2) \mathbf{E}[Y]}{n^{\alpha-1} 2a_{20}^\alpha x_k^{2\alpha-1}} = \lim_{n \rightarrow \infty} \frac{b_{11}(v_1 - v_2) \frac{\mathbf{E}[Y]}{x_k}}{2a_{20}^\alpha (nx_k^2)^{\alpha-1}}$$

The numerator in this last term is bounded above and below, while the denominator goes to infinity. This is a contradiction, and thus $\limsup_{n \rightarrow \infty} x_k(n) n^{\frac{1}{4}-\epsilon} = 0$ for every $\epsilon > 0$.

Finally, assume that $\limsup_{n \rightarrow \infty} x_k(n) n^{\frac{1}{4}} = 0$. As in the first part of this proof, we have that the FOC (8) reduces to (9). In (9), the first term in the numerator of the RHS is bounded above and below, while the rates of growth or decay of both the other term in the numerator and the denominator are determined by the behavior of $g(n) = nx_k^2$. If $g(n)$ diverges, then since $\alpha \neq 2$, numerator and denominator diverge at different rates, which is an immediate contradiction. If $g(n)$ converges to 0, then the denominator converges to 0 while the numerator is bounded below, which is again a contradiction. This completes the proof. \square

LEMMA 4. *Let $x(n)$ be a sequence of m -vectors of equilibrium efforts.*

1. *If for some k*

$$\limsup_{n \rightarrow \infty} x_k n^{\frac{1}{2}} < \infty, \tag{10}$$

and along some subsequence it is the case that for each k $\lim_{n \rightarrow \infty} x_k n^{\frac{1}{2}} = d_k$, then $d_k > 0$:

$$\frac{d_k}{d_j} = \left(\frac{c_j}{c_k} \right)^{\frac{1}{2\alpha-1}}.$$

Equation (10) always holds for $\alpha > 2$.

2. *If $1 < \alpha < 2$ then there exists $\tau_{eq}(\alpha, b_{11}, a_{20}, v_1, \mathbf{c}) > 0$ such that if $qv_2 > \tau_{eq}$ then*

$$\liminf_{n \rightarrow \infty} x(n) n^{\frac{1}{4}} > 0.$$

Proof. For the first part, assume that $\alpha > 2$, and that for some k , it holds that

$$\liminf_{n \rightarrow \infty} x_k n^{\frac{1}{2}} = \infty.$$

Then, using (9), we have:

$$c_k \leq \lim_{n \rightarrow \infty} \frac{b_{11}v_1 \mathbf{E}[Y] + v_2 n q b_{11}^2 \mathbf{E}[Y] \mathbf{E}[Y^2]}{n^{\alpha-1} 2a_{20}^\alpha x_k^{2\alpha-1}} = \lim_{n \rightarrow \infty} \frac{b_{11}v_1 \frac{\mathbf{E}[Y]}{x_k} + v_2 n q b_{11}^2 \frac{\mathbf{E}[Y]}{x_k} \mathbf{E}[Y^2]}{2a_{20}^\alpha (nx_k^2)^{\alpha-1}}$$

Now the numerator of the RHS diverges to infinity at a rate of nx_k^2 , while the denominator diverges at the higher rate of $(nx_k^2)^{\alpha-1}$, since $\alpha > 2$. This is a contradiction, and so we have that $\liminf_{n \rightarrow \infty} x_k n^{\frac{1}{2}} < \infty$.

Now assume that for some $1 < \alpha < \infty$, for each k we have that $\lim x_k n^{\frac{1}{2}} = d_k$. Plugging this into (9), gives:

$$c_k = \lim_{n \rightarrow \infty} \frac{b_{11}v_1 \mathbf{E}[Y] + v_2 n q b_{11}^2 \mathbf{E}[Y] \mathbf{E}[Y^2]}{n^{\alpha-1} 2a_{20}^\alpha x_k^{2\alpha-1}} = \frac{b_{11}v_1 \mathbf{E}[D] + v_2 q b_{11}^2 \mathbf{E}[D] \mathbf{E}[D^2]}{2a_{20}^\alpha d_k^{2\alpha-1}}$$

where D is a random variable taking on the value d_k with probability π_k .

If d_k is 0 this gives a contradiction, so $d_k > 0$ for every k . Since the above equation holds for every k , we have that:

$$\frac{d_k}{d_j} = \left(\frac{c_j}{c_k} \right)^{\frac{1}{2\alpha-1}}.$$

Plugging this result into the equation for c_1 , gives:

$$c_1 = \frac{b_{11}v_1 \mathbf{E}\left[S^{\frac{1}{2\alpha-1}}\right] d_1 c_1^{\frac{1}{2\alpha-1}} + v_2 q b_{11}^2 d_1^3 c_1^{\frac{3}{2\alpha-1}} \mathbf{E}\left[S^{\frac{1}{2\alpha-1}}\right] \mathbf{E}\left[S^{\frac{2}{2\alpha-1}}\right]}{2a_{20}^\alpha d_1^{2\alpha-1}} \quad (11)$$

where S is a random variable taking on the value $s_k = 1/c_k$ with probability π_k . This translates to the following equation:

$$2a_{20}^\alpha d_1^{2(\alpha-1)} = \mathbf{E}\left[S^{\frac{1}{2\alpha-1}}\right] \left(b_{11}v_1 s_1^{\frac{2(\alpha-1)}{2\alpha-1}} + v_2 q b_{11}^2 d_1^2 s_1^{\frac{2(\alpha-2)}{2\alpha-1}} \mathbf{E}\left[S^{\frac{2}{2\alpha-1}}\right] \right)$$

Rearranging this as an equation for $v_2 q$ gives:

$$v_2 q = \frac{s_1^{\frac{1}{2\alpha-1}} \left(2a_{20}^\alpha s_1^{\frac{2-2\alpha}{2\alpha-1}} d_1^{2(\alpha-1)} - b_{11}v_1 \mathbf{E}\left[S^{\frac{1}{2\alpha-1}}\right] \right)}{b_{11}^2 \mathbf{E}\left[S^{\frac{1}{2\alpha-1}}\right] \mathbf{E}\left[S^{\frac{2}{2\alpha-1}}\right] d_1^2} \quad (12)$$

Viewing $v_2 q$ as a function of d_1 , we can maximize via elementary analysis with respect to d_1 to obtain a largest possible value of $v_2 q$ for which the equation is solvable. Let this value (which is clearly positive) be called τ_{eq} .

This completes the proof. \square

LEMMA 5. Assuming $v_2 q \leq \tau_{eq}$, there is some $\beta > 0$ so that for large enough n , there exists an equilibrium of $\Gamma(n)$ with $x_k n^{1/2} < \beta$ for every k .

Proof sketch. Consider the first-order conditions of the agents in $\Gamma(n)$ and perform the change of variables $x_k = w_k n^{-1/2}$. Write that system as $G_n(\mathbf{w}) = 0$. Consider also the asymptotic system of equations

$$c_k = \frac{b_{11}v_1 \mathbf{E}[D] + v_2 q b_{11}^2 \mathbf{E}[D] \mathbf{E}[D^2]}{2a_{20}^\alpha d_k^{2\alpha-1}},$$

in the notation of the previous lemma. Define G_∞ so that this system can be written $G_\infty(\mathbf{w}) = 0$. This asymptotic system of equations, as seen above, boils down to just one equation, (11), which has at least one solution \mathbf{d} under the assumption $v_2 q \leq \tau_{\text{eq}}$. Moreover, it can be checked that the derivative of that equation at one of those solutions is nonzero. Thus, the degree of G_∞ in a punctured neighborhood of \mathbf{d} , say $\Omega \setminus \{\mathbf{d}\}$, is nonzero. In the closure of Ω , it can be shown that G_n converges uniformly to G_∞ . Thus the degree of G_n for high enough n is nonzero in $\Omega \setminus \{\mathbf{d}\}$, and in particular G_n has a zero there.⁶ \square

LEMMA 6. *Given any $\gamma > 0$, for large enough n , there exists an equilibrium of $\Gamma(n)$ with $x_k > \gamma n^{-1/4}$ for every k .*

Proof sketch. Fix $\epsilon > 0$. By Lemma 2, all equilibria eventually lie in an ϵ -neighborhood of 0. Since they are interior, and the utility function is concave in each type's action, equilibrium behavior is characterized by a solution of the first-order condition. If this ϵ is chosen small enough, it can be checked by taking derivatives and using asymptotic analysis as above that any agent's best-response is increasing in others' actions.

Assume that n is large enough so that all equilibria both of the game with the original costs and with costs all equal to the maximum lie have actions of less than ϵ . Now restrict the strategy space to forbid playing effort levels greater than ϵ . The game is now one of strategic complements.

It is also clear that agents' response to increases in costs (holding all else fixed) is to decrease actions, and vice versa. So let us consider the following exercise: change each type's cost to the maximum cost, and then iterate best responses until an equilibrium is reached. After this is done for all types, the monotonicity discussed above shows that all actions are strictly lower than in the original equilibrium.

Now all types have the same first-order condition. It can be checked that if everyone is playing $x = \gamma n^{-1/4}$, then marginal benefits exceed marginal costs for large enough n and agents will want to adjust upwards. The structure of games with strategic complements implies the existence of an equilibrium with effort levels exceeding $\gamma n^{-1/4}$. \square

Proof of Proposition 1 These observations follow from standard results in random graph theory. See Jackson (2008), Theorem 4.1 for the results on connectedness. As for the diameter: Corollary 10.12(i) in Bollobás (2001) gives that the diameter of \mathbf{G}^E is 3, so this is an upper bound on the diameter of the final network. Moreover, since, a.a.s., there is a pair of agents in \mathbf{G}^E at distance 3 from each other, they cannot end up at distance 1 due to the addition of links in \mathbf{G}^L . Thus, the diameter of the final network is at least 2. \blacksquare

Mingling as an Equilibrium

Proof of Theorem 5 Fix $i \in N$ and $k \in \{1, \dots, m\}$. Let \mathbf{x} be the strategy profile corresponding to the equilibrium considered in the problem statement. Observe that all agents $j \neq i$ are, by assumption, playing symmetric mingling strategies. We claim that for enough n , playing the same mingling strategy is the unique best response for agent i . Let $\bar{x} = \min_k x_k^{(j,\ell)}$ and $\underline{x} = \min_k x_k^{(j,\ell)}$ for any fixed $j, \ell \in N \setminus \{i\}$. This definition makes sense because $x^{(j,\ell)}$ does not depend on j or ℓ , given that \mathbf{x} is a symmetric mingling equilibrium.

⁶We are grateful to Pietro Majer and André Henriques for their suggestions, which were crucial to the proof of this result.

Let $Y_j = Z^{(j,\ell)}$ for any ℓ . This is the random variable describing the level of j 's effort. It is random because of the randomness of j 's type. Given that the equilibrium is mingling, the choice of ℓ does not matter. The symmetry of the equilibrium implies that the Y_j are identically distributed across j , and the assumption on how types are drawn implies that they are independent. We will drop references to the cost type of agent i , taking that as known. Write

$$u_i(\mathbf{z}^i; \mathbf{x}^{(-i,\cdot)}) = u_i^E(\mathbf{z}^i; \mathbf{x}^{(-i,\cdot)}) + u_i^L(\mathbf{z}^i; \mathbf{x}^{(-i,\cdot)}) - u_i^C(\mathbf{z}^i; \mathbf{x}^{(-i,\cdot)}) - \Delta(\mathbf{z}^i).$$

where:

$$u_i^E = v_1 \mathbf{E} \left[\sum_{j \neq i} p(z^{(i,j)}, Y_j) \right];$$

$$u_i^L = v_2 \mathbf{E} \left[\sum_{j \neq i} [1 - p(z^{ij}, Y_j)] \sum_{L \subset N \setminus \{i,j\}} [1 - (1-q)^{|L|}] \prod_{\ell \in L} [p(z^{i\ell}, Y_\ell) p(Y_\ell, Y_j)] \prod_{\substack{\ell \notin L \\ \ell \neq i,j}} [1 - p(z^{i\ell}, Y_\ell) p(Y_\ell, Y_j)] \right];$$

and

$$u_i^C = \mathbf{E} \left[\frac{c_i}{\alpha} \left(\sum_{j \neq i} f(z^{ij}, Y_j) \right)^\alpha \right].$$

Here we have split the expected utility of agent i into three pieces: the benefits coming from early-stage friends, the benefits coming from late-stage friends, and the costs.

Note that u_i^E is concave and u_i^C is convex, so it suffices to show that $u_i^L - s$ is concave in \mathbf{z}^i . Define \mathbf{H}^g to be the Hessian of a function g in the variables z^{ij} for $j \neq i$. That is, let the (s, t) entry⁷ of \mathbf{H}^g be

$$H_{st}^g = \frac{\partial^2 g}{\partial z^{(i,t)} \partial z^{(i,s)}}.$$

LEMMA 7. $\max_{s,t \in N \setminus \{i\}} H_{st}^{u_i^L} \in O(n^{-1/2} \log^6 n)$.

Once this is established, the assumed concavity of s and the rate of growth of the entries of its Hessian guarantees that $\mathbf{H}^{u_i^L}$ becomes arbitrarily small, entry by entry, relative to \mathbf{H}^Δ . Since the Hessian of s is positive definite, that implies that the Hessian of $u_i^L - \Delta$ is eventually negative definite, yielding that this function is concave. It remains to prove the lemma.

Proof of Lemma 7. Using Taylor expansion of p and Lagrange's error bound as in [[TODO:cite]] we have, for some absolute constants $R_1, R_2 \in (0, \infty)$,

$$\frac{H_{st}^{u_i^L}}{R_1} = -\mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,s\} \\ t \in L}} Y_s^2 Y_t^2 [1 - (1-q)^{|L|}] \prod_{\substack{\ell \in L \\ \ell \neq t}} [R_2^2 z^{i\ell} Y_\ell^2 Y_s] \prod_{\substack{\ell \notin L \\ \ell \neq i,s}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_s] \right]$$

⁷Note that we are not allowing the index j to take the value i , so the rows and columns of \mathbf{H}^g are, alas, not numbered consecutively.

$$\begin{aligned}
& + \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,s\} \\ t \notin L}} Y_s^2 Y_t^2 \left[1 - (1-q)^{|L|} \right] \prod_{\ell \in L} [R_2^2 z^{i\ell} Y_\ell^2 Y_s] \prod_{\substack{\ell \notin L \\ \ell \neq i,s,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_s] \right] \\
& + \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,t\} \\ s \notin L}} Y_s^2 Y_t^2 \left[1 - (1-q)^{|L|} \right] \prod_{\ell \in L} [R_2^2 z^{i\ell} Y_\ell^2 Y_t] \prod_{\substack{\ell \notin L \\ \ell \neq i,s,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_t] \right] \\
& - \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,t\} \\ s \in L}} Y_s^2 Y_t^2 \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq s}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right] \\
& + \mathbf{E} \left[\sum_{j \neq i} \sum_{\substack{L \subset N \setminus \{i,j\} \\ s,t \in L}} Y_s^2 Y_t^2 Y_j^2 \left[1 - R_2 z^{ij} Y_j \right] \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq t,s}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right] \\
& - \mathbf{E} \left[\sum_{j \neq i} \sum_{\substack{L \subset N \setminus \{i,j\} \\ s \in L, t \notin L}} Y_s^2 Y_t^2 Y_j^2 \left[1 - R_2 z^{ij} Y_j \right] \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq s}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right] \\
& - \mathbf{E} \left[\sum_{j \neq i} \sum_{\substack{L \subset N \setminus \{i,j\} \\ s \notin L, t \in L}} Y_s^2 Y_t^2 Y_j^2 \left[1 - R_2 z^{ij} Y_j \right] \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq t}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j,s}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right] \\
& + \mathbf{E} \left[\sum_{j \neq i} \sum_{\substack{L \subset N \setminus \{i,j\} \\ s,t \notin L}} Y_s^2 Y_t^2 Y_j^2 \left[1 - R_2 z^{ij} Y_j \right] \left[1 - (1-q)^{|L|} \right] \prod_{\ell \in L} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j,t,s}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right].
\end{aligned}$$

By the triangle inequality, and recalling that \bar{x} is an upper bound for every Y_j , we have:

$$\begin{aligned}
\left| \frac{H_{st}^{u_i^L}}{R_1} \right| & \leq \bar{x}^4 \cdot \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,s\} \\ t \in L}} \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq t}} [R_2^2 z^{i\ell} Y_\ell^2 Y_s] \prod_{\substack{\ell \notin L \\ \ell \neq i,s}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_s] \right] \\
& + \bar{x}^4 \cdot \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,s\} \\ t \notin L}} \left[1 - (1-q)^{|L|} \right] \prod_{\ell \in L} [R_2^2 z^{i\ell} Y_\ell^2 Y_s] \prod_{\substack{\ell \notin L \\ \ell \neq i,s,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_s] \right] \\
& + \bar{x}^4 \cdot \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,t\} \\ s \notin L}} \left[1 - (1-q)^{|L|} \right] \prod_{\ell \in L} [R_2^2 z^{i\ell} Y_\ell^2 Y_t] \prod_{\substack{\ell \notin L \\ \ell \neq i,s,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_t] \right] \\
& + \bar{x}^4 \cdot \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,t\} \\ s \in L}} \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq s}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right]
\end{aligned}$$

$$\begin{aligned}
& + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,j\} \\ s,t \in L}} [1 - R_2 z^{ij} Y_j] \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq t,s}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right] \\
& + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,j\} \\ s \in L, t \notin L}} [1 - R_2 z^{ij} Y_j] \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq s}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j,t}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right] \\
& + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,j\} \\ s \notin L, t \in L}} [1 - R_2 z^{ij} Y_j] \left[1 - (1-q)^{|L|} \right] \prod_{\substack{\ell \in L \\ \ell \neq t}} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j,s}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right] \\
& + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{E} \left[\sum_{\substack{L \subset N \setminus \{i,j\} \\ s,t \notin L}} [1 - R_2 z^{ij} Y_j] \left[1 - (1-q)^{|L|} \right] \prod_{\ell \in L} [R_2^2 z^{i\ell} Y_\ell^2 Y_j] \prod_{\substack{\ell \notin L \\ \ell \neq i,j,t,s}} [1 - R_2^2 z^{i\ell} Y_\ell^2 Y_j] \right].
\end{aligned}$$

We will now bound (modulo a universal multiplicative constant) each expectation by a probability of i and j having a late connection, under some strategy profile, and then bound each of these probabilities by 1. In particular, define $\tilde{\mathbf{x}}(i; s, t; j)$ to be the strategy profile in which everyone behaves as in \mathbf{x} , except that, for all types of i, j, s , and t :

- i directs effort 1 at s and t ;
- each of s and t directs effort 1 at i ;
- j directs effort 1 at s and t ;
- each of s and t directs effort 1 at j .

When there is only one index between the semicolons, the definition is analogous. That is, define $\tilde{\mathbf{x}}(i; s; j)$ to be the strategy profile in which everyone behaves as in \mathbf{x} , except that, for all types of i, j , and s ,

- i directs effort 1 at s ;
- s directs effort 1 at i and at j ;
- j directs effort 1 at s .

Then, using that $\bar{x} \rightarrow 0$, we deduce that for some constant R_3 we have:

$$\begin{aligned}
\left| \frac{H_{st}^{u_i^L}}{R_3} \right| & \leq \bar{x}^4 \cdot \mathbf{P}^{\tilde{\mathbf{x}}(i;t;s)} (G_{is}^L = 1) + \bar{x}^4 \cdot \mathbf{P}^{\mathbf{x}} (G_{is}^L = 1) + \bar{x}^4 \cdot \mathbf{P}^{\mathbf{x}} (G_{it}^L = 1) + \bar{x}^4 \cdot \mathbf{P}^{\tilde{\mathbf{x}}(i;s;t)} (G_{is}^L = 1) \\
& + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{P}^{\tilde{\mathbf{x}}(i;s,t;j)} (G_{ij}^L = 1) + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{P}^{\tilde{\mathbf{x}}(i;s;j)} (G_{ij}^L = 1) + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{P}^{\tilde{\mathbf{x}}(i;t;j)} (G_{ij}^L = 1) \\
& + \bar{x}^6 \cdot \sum_{j \neq i} \mathbf{P}^{\mathbf{x}} (G_{ij}^L = 1).
\end{aligned}$$

It follows that for some constant R_4 we have

$$\left| H_{st}^{u_i^L} \right| \leq R_4 (\bar{x}^4 + n \bar{x}^6).$$

□

This completes the proof of the theorem.

■