

The Law of the Few

Andrea Galeotti and Sanjeev Goyal*

This version: January 8, 2009

First Draft: August 2007.

Abstract

The *law of the few* refers to the following social phenomenon: a large majority of individuals get most of the information needed for their decisions from a very small subset of the group, viz., *the influencers*. Empirical work suggests that there are no significant differences in economic and demographic characteristics between the influencers and the rest. We are led to ask: is this phenomenon a consequence of strategic interaction among identical individuals?

This paper develops a model where players personally acquire information and form connections with others to access their information. We show that the ‘law of the few’ is a robust equilibrium outcome of this model. (JEL C72, D00, D83, H41, O31)

*Galeotti: Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK (e-mail: agaleo@essex.ac.uk); Goyal: Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DD, UK (e-mail: sg472@econ.cam.ac.uk). We thank the editor and three anonymous referees for useful suggestions. We are grateful to Bhaskar Dutta, Francis Bloch, Yann Bramoulle, Myeonghwan Cho, Matthew Jackson, Marco van der Leij, Brian Rogers and a number of seminar audiences for suggestions which have significantly improved the paper.

The determination of information people have when making decisions is a central question in the social sciences. Empirical research across a wide range of subjects including political science, sociology, management, marketing, economics and computer science, has brought out the important role of personal contacts in transmitting information across people. Creating and maintaining personal contacts however takes time and resources. Individuals therefore compare these costs with the costs of direct information acquisition. This paper explores the consequences of these individual choices for the structure of social communication and information available in a society.

Our point of departure is a series of empirical studies about information acquisition and social communication in social groups. The early studies of Paul F. Lazarsfeld, Bernard Berelson, and Hazel Gaudet 1948 and Elihu Katz and Paul F. Lazarsfeld 1955 investigate the impact of personal contacts and mass media on voting and consumer choice with regard to brands, films and fashion changes.¹ They find that personal contacts play a dominant role in disseminating information which in turn shapes individuals' decisions. In particular, they identify 20 percent of their sample of 4000 individuals as the primary source of information for the rest. Similarly, Lawrence F. Feick and Linda L. Price 1987 in their study of information acquisition and sharing about food, household products, nonprescription drugs and beauty products, find that 25 percent of their sample of 1400 individuals acquired a great deal of information which was widely accessed by the rest.²

Personal contacts also play a major role within firms and virtual communities. Rob Cross et al. 2001 study 40 managers working in Fortune 500 firms, and find that about 85 percent of them obtain information useful via personal communication with a few other colleagues (and that personal contacts are significantly more important than other sources of information).³ Jun Zhang, Mark S. Ackerman and Lada Adamic 2007 study the Java Forum: an on-line community of users who ask and respond to queries concerning the programme Java. They identify around 14000 users and find that about 55 percent only ask questions,

¹Recent work in political science arrives at similar conclusions, see e.g., Paul Beck et al. 2002, and Robert Huckfeldt, Paul E. Johnson and John Sprague 2004.

²For recent work on mavens and market collection, see Klaus P. Wiedman, Gianfranco Walsh, and Vincent W. Mitchell 2001, Terrell G. Williams and Mark E. Slama 1995, and Gary L. Geissler and Steve W. Edison 2005.

³We refer to Rob Cross and Andrew Parker 2004 for a survey and references.

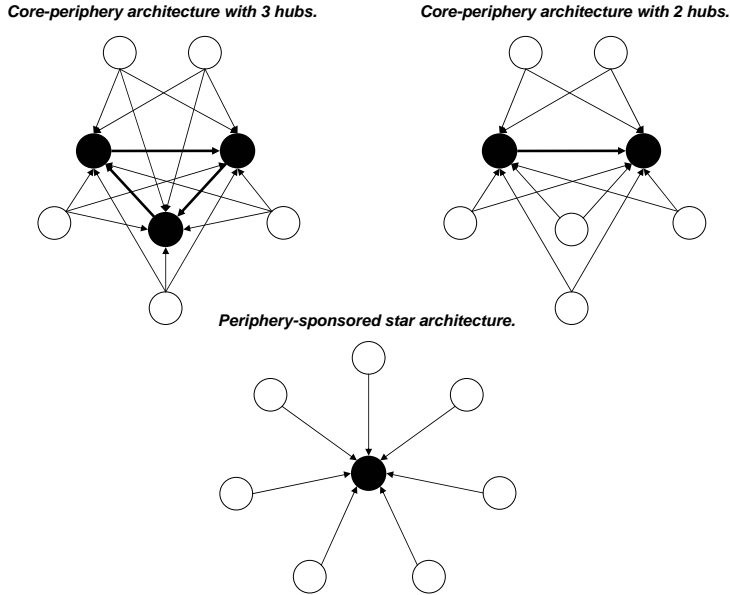


Figure 1: Core-Periphery Networks.

12 percent both ask and answer queries, and about 13 percent only provide answers. Eytan Adar and Bernardo A. Huberman 2000 report similar findings with regard to provision of files in the peer-to-peer network, Gnutella.

The Law of the Few subsumes these different empirical examples: in social groups, a majority of individuals get most of their information from a very small subset of the group, viz., *the influencers*. Moreover, research suggests that there are no significant differences in the observable economic and demographic characteristics between the social influencers and the rest. We are led to ask: is this phenomenon a consequence of strategic interaction among identical individuals?

We propose a game in which individuals choose to personally acquire information and to form connections with others to access the information which they acquire. In this basic setting, we show that every (strict) equilibrium of the game exhibits the law of the few (Propositions 2-3). Figure 1 illustrates equilibrium configurations, the network has the *core-periphery architecture*, the players in the core acquire information personally while the peripheral players acquire no information personally but form links and get all their information from the core players—information hubs. As the number of players gets large, the information hubs constitute a small fraction of the group.

We informally outline the ideas underlying this result. In our model, returns from infor-

mation are increasing and concave while the costs of personally acquiring information are linear. This implies that on his own an individual would acquire an amount of information, which we denote by 1. The second element is the substitutability of information acquired by different individuals. This implies that if A acquires information on his own and receives information from player B then in the aggregate he must have access to 1 unit of information (else he could strictly increase his payoff by modifying personal information acquisition). The third and perhaps key element is that links are costly and rationally chosen by individuals. The implication of this is that if A finds it optimal to maintain a link with B then so must every other player. So the group of individuals who acquire information must be completely linked and the aggregate information acquired in the society must equal exactly 1. Moreover, since linking is costly, A will only link with B if B acquires a certain minimum amount of information. Since total information acquired is 1, it follows that there is an upper bound to the number of people who will acquire information and so the proportion of individuals who acquire information personally in a large group is very small. Finally, we observe that since the aggregate information acquired in the group is 1, everyone who does not personally acquire information must be linked to all those who acquire information, yielding the core-periphery network.

The results mentioned above pertain to a setting in which individuals are ex-ante identical. A recurring theme in the empirical work is that information hubs have similar demographic characteristics as the others. But this work also finds that they have distinctive attitudes which include higher attention of general market information and enjoyment of collecting information, e.g., Feick and Price 1987. Using these empirical considerations as our point of departure, we examine the consequences of minor heterogeneities. We show that a slight cost advantage in acquiring information implies that there is a unique equilibrium in which the low cost individual agent becomes the only information hub: he acquires all the information, while everyone else simply connects with this person (Proposition 3). Small heterogeneities thus help in selecting individuals who will play dramatically different roles in social organization.

In actual practice, we receive information from friends and colleagues which they have themselves received from their friends. This motivates an extension of the basic model to al-

low for indirect information transmission. Our main insight here is that indirect information transmission gives rise to a new type of influential agent: *the connector* (Proposition 5). A connector is someone whose primary role is that of an ‘informediary’: he typically has many more links than the average and connects individuals who personally acquire information.

The principal contribution of our paper is a model of strategic investments in information acquisition and link formation. Our model builds on the link formation formulation introduced in Sanjeev Goyal 1993 and Venkatesh Bala and Sanjeev Goyal 2000 and the model of public goods in networks developed in Yann Bramouille and Rachel Kranton 2007. We show that two economic ideas, one, the substitutability of information acquired personally and that acquired by others, and two, the possibility of forming costly links with others who acquire information, explain the specialization in information acquisition and social communication, reflected in the law of the few. Moreover, in line with the findings of empirical work, we show that a small difference – with regard to the costs of acquiring information or with regard to personal sociability – is sufficient to perfectly identify influencers in a social group.

The theory of network formation as well as the theory of games played on fixed networks are currently active fields of research; for surveys, see Sanjeev Goyal 2007 and Matthew Jackson 2008. We observe that existing models cannot explain the Law of the Few: the pure link formation model cannot account for the specialization in information acquisition, while the public goods model cannot account for the specific patterns of social communication.

Our results also resolve some important theoretical questions in the public goods model. Introducing link formation yields clear cut predictions on individual as well as aggregate information acquisition. By contrast, with exogenous networks, Bramouille and Kranton 2007 show that multiple equilibria typically exist and these equilibria exhibit very different individual and aggregate information acquisition.

We next relate our paper to the literature on games of network formation. In a setting with pure local information sharing, link formation among individuals with exogenously given information yields either the complete or the empty network, depending on whether the cost of linking is lower or higher than the value of individual information. Our main result is that the unique equilibrium is a core-periphery network with the hubs acquiring all

the information. Thus endogenous information acquisition provides an alternative theoretical foundation for core-periphery (and star) networks. Second, our result generates an interesting cost of link effect: an increase in this cost leads to a fall in the number of players who acquire information, an increase in the amount of information acquired by each of the player who acquires information and a decrease in the total number of links. Clearly, changes in costs of linking have no effect on information acquisition in earlier models of pure link formation, where everyone has exogenously provided information (see e.g., Bala and Goyal 2000, Andrea Galeotti, Sanjeev Goyal and Jurjen Kamphorst 2006 and Daniel Hojman and Adam Siezdl 2008).⁴

Non-rival information acquisition is an instance of the private provision of public goods. For global public goods, Theodore Bergstrom, Lawrence Blume and Hal R. Varian 1986 showed that the contributors will be those with higher endowments. The key difference in our model is that access to public good is a matter of individual choice, it is costly and takes the form of bilateral connections. The findings with regard to the existence of information hubs and connectors and the network structure are, to the best of our knowledge, novel in this literature.

The rest of the paper is organized as follows. Section I develops the basic model with local information flow and section II analyzes this model. Section III considers a model with indirect information transmission. Section IV discusses two important aspects of our model, linear costs of acquiring information and forming links and the link formation protocol, respectively. Section V concludes. Appendix A contains the proofs of the results. Appendix B presents and solves three alternative formulations of the model, involving one-way flow of information, discrete information acquisition and investment in strength of links, respectively.

⁴There is a small body of work which combines network formation and play in games, see e.g., Sanjeev Goyal and Fernando Vega-Redondo 2005, Matthew O. Jackson and Alison Watts 2002, Yann Bramoulle et. al. 2004 and Antoni Calvo-Armengol and Yves Zenou 2004. The game studied and the questions addressed in the present paper are quite different from this literature and so a detailed discussion of the relation with these papers is omitted.

I. Model

Let $N = \{1, 2, \dots, n\}$ with $n \geq 3$ be the set of players and let i and j be typical members of this set. Each player i chooses a level of personal information acquisition $x_i \in X$ and a set of links with others to access their information, which is represented as a (row) vector $\mathbf{g}_i = (g_{i1}, \dots, g_{ii-1}, g_{ii+1}, \dots, g_{in})$, where $g_{ij} \in \{0, 1\}$, for each $j \in N \setminus \{i\}$. We will suppose that $X = [0, +\infty)$ and that $\mathbf{g}_i \in G_i = \{0, 1\}^{n-1}$.⁵ We say that player i has a link with player j if $g_{ij} = 1$. A link between player i and j allows both players to access the information personally acquired by the other player. This assumption that exchange of information is mutual appears to be natural in the context of inter-personal communication.⁶ The set of strategies of player i is denoted by $S_i = X \times G_i$. Define $S = S_1 \times \dots \times S_n$ as the set of strategies of all players. A strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g}) \in S$ specifies the personal information acquired by each player, $\mathbf{x} = (x_1, x_2, \dots, x_n)$, and the network of relations $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n)$.

The network of relations \mathbf{g} is a directed graph; let G be the set of all possible directed graphs on n vertices. Define $N_i(\mathbf{g}) = \{j \in N : g_{ij} = 1\}$ as the set of players with whom i has formed a link. Let $\eta_i(\mathbf{g}) = |N_i(\mathbf{g})|$. The closure of \mathbf{g} is an undirected network denoted $\bar{\mathbf{g}} = cl(\mathbf{g})$, where $\bar{g}_{ij} = \max\{g_{ij}, g_{ji}\}$ for each i and j in N . In words, the closure of a directed network involves replacing every directed edge of g by an undirected one. Define $N_i(\bar{\mathbf{g}}) = \{j \in N : \bar{g}_{ij} = 1\}$ as the set of players directly connected to i . The undirected link between two players reflects the bilateral nature of information exchange between them.

The payoffs to player i under strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ are:

$$\Pi_i(\mathbf{s}) = f \left(x_i + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j \right) - cx_i - \eta_i(\mathbf{g})k, \quad (1)$$

where $c > 0$ reflects the cost of information and $k > 0$ is the cost of linking with one other person. We assume pure local externalities: player i only accesses the information personally acquired by his immediate neighbors. Section III studies a model with indirect information transmission.

⁵A model in which information is a discrete variable is presented and completely analyzed appendix B. Our main results on equilibrium information acquisition and networks are robust to a change in the nature of the action sets.

⁶Appendix B presents and analyzes a model in which link formation and information flow are both one-way.

We will assume that $f(y)$ is twice continuously differentiable, increasing, and strictly concave in y . To focus on interesting cases we will assume that $f(0) = 0$, $f'(0) > c$ and $\lim_{y \rightarrow \infty} f'(y) = z < c$. Under these assumptions there exists a number $\hat{y} > 0$ such that $\hat{y} = \arg \max_{y \in X} f(y) - cy$.

We now discuss the key elements of our model.

First, consider the returns from information. We may think of the action x as reflecting draws from a distribution, e.g., the price distribution for a product. If the different draws are independent across individuals and players are interested in the lowest price, then the value of an additional draw, which is the change in the average value of the lowest order statistic, is positive but declining in the number of draws. Another possible interpretation is in terms of individuals choosing an action whose payoffs are unknown. Every individual has access to a costly sample of observations – which may reflect personal experience with a product or a technology. A link with another player then allows access to his personal experience. Under reasonable conditions, the returns from accessing more samples of information – own and others – are increasing but concave.⁷

Second, consider the protocol of link formation. We assume a person can form a binary link (it takes value one or zero) and that this link is formed once a cost is incurred. One possible interpretation of unilateral link formation and two-way exchange of information is that one player pays for a telephone call and the call involves an exchange of information. Alternatively, we may interpret the cost incurred in the formation of a link as a ‘gift’ or a social favor that the player forming the link makes to the person receiving the link. In this case, k becomes a transfer and the payoffs to player i in a strategy profile $\mathbf{s} = (\mathbf{g}, \mathbf{s})$ are given by:

$$\hat{\Pi}_i(\mathbf{s}) = f \left(x_i + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j \right) - cx_i - \eta_i(\mathbf{g})k + \sum_{j \in N_i(\mathbf{g})} g_{ji}k.$$

The last term involving transfers is independent of the strategy of i . It then follows that for all $\mathbf{s}_{-i} \in S_{-i}$, and $\mathbf{s}_i, \mathbf{s}'_i \in S_i$, $\hat{\Pi}_i(\mathbf{s}_i, \mathbf{s}_{-i}) \geq \hat{\Pi}_i(\mathbf{s}'_i, \mathbf{s}_{-i})$, if and only if $\Pi_i(\mathbf{s}_i, \mathbf{s}_{-i}) \geq \Pi_i(\mathbf{s}'_i, \mathbf{s}_{-i})$. Therefore our methods of analysis and our findings with payoffs (1) carry over

⁷For a detailed discussion of these examples on information sharing, see Bramoulle and Kranton 2007 and Hojman and Szeidl 2008.

to the alternative model, where link formation costs are transfers from one individual to another.

Third, we assume that a player has no interest in misleading others and that there are no incentives for refusing to share information. This assumption appears to be reasonable in a number of contexts such as e.g., informal social communication between individual consumers.

Finally, we assume that there is no bargaining and no prices for information. We are aware that this is a strong assumption. As a first step toward incorporating prices, in section IV, we study the case where players can ask for transfers in exchange for sharing information.⁸

A Nash equilibrium is a strategy profile $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$ such that:

$$\Pi_i(\mathbf{s}_i^*, \mathbf{s}_{-i}^*) \geq \Pi_i(\mathbf{s}_i, \mathbf{s}_{-i}^*), \forall \mathbf{s}_i \in S_i, \forall i \in N, \quad (2)$$

where we have used the convention that for any profile \mathbf{s} , \mathbf{s}_{-i} refers to the strategies of all players other than player i . An equilibrium is said to be *strict* if the inequalities in the above definition are strict for every player.

We define social welfare to be the sum of individual payoffs. For any profile \mathbf{s} social welfare is given by:

$$W(\mathbf{s}) = \sum_{i \in N} \Pi_i(\mathbf{s}). \quad (3)$$

A profile \mathbf{s}^* is socially efficient if $W(\mathbf{s}^*) \geq W(\mathbf{s})$, $\forall \mathbf{s} \in S$.

In a *core-periphery* network there are two groups of players, $\hat{N}_1(\bar{\mathbf{g}})$ and $\hat{N}_2(\bar{\mathbf{g}})$, with the feature that $N_i(\bar{\mathbf{g}}) = \hat{N}_2(\bar{\mathbf{g}})$ for all $i \in \hat{N}_1(\bar{\mathbf{g}})$, and $N_j(\bar{\mathbf{g}}) = N \setminus \{j\}$ for all $j \in \hat{N}_2(\bar{\mathbf{g}})$. Nodes which have $n - 1$ links are referred to as central nodes or as hubs, while the complementary set of nodes are referred to as peripheral nodes or as spokes. A core-periphery network with a single hub is referred to as a periphery-sponsored star. Figure 1 illustrates core-periphery networks. There are $n = 8$ players; in each architecture the black nodes are the hubs (the set $\hat{N}_2(\bar{\mathbf{g}})$), the white nodes are the spokes (the set $\hat{N}_1(\bar{\mathbf{g}})$) and an edge starting at i with the arrowhead pointing at j indicates that i sponsors a link to j . We say that $\bar{\mathbf{g}}$ is a regular

⁸For other recent studies of models in which players can charge prices or bargain for their information, see Antonio Cabrales and Piero Gottardi 2007 and Myeonghwan Cho 2007.

network of degree v if each player has v connections in $\bar{\mathbf{g}}$. A complete network is a regular network with $v = n - 1$.

II. Analysis

The main result of this section is that *every* (strict) equilibrium exhibits the law of the few: a small subsets of individuals collect all the information, while the rest of individuals form connections with this small set of information collectors. This differentiation in turn generates a simple and elegant structure of social communication: information collectors constitute hubs while the rest of the individuals are spokes in a core-periphery social network.

We start by noting that in every equilibrium each player must access at least \hat{y} information (where \hat{y} is the solution of the optimization problem of the isolated player, i.e., $\hat{y} = \arg \max_{y \in X} f(y) - cy$). Moreover, the perfect substitutability of own and neighbor's information and the linearity in the costs of acquiring information imply that if a player personally acquires information then the sum of the information he acquires and the information acquired by his neighbors must equal \hat{y} . We next observe that if some player collects \hat{y} , and if $k < c\hat{y}$, then it is optimal for all other players to acquire no information personally and to form a link with this player. Lemma 1 summarizes these observations. For a strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$, let us define $I(\mathbf{s}) = \{i \in N | x_i > 0\}$ as the set of players who acquire information personally, and let $y_i = \sum_{j \in N_i(\bar{\mathbf{g}})} x_j$ be the information that i accesses from his neighbors.⁹

LEMMA 1: *In any equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$, $x_i^* + y_i^* \geq \hat{y}$, for all $i \in N$, and if $x_i^* > 0$ then $x_i^* + y_i^* = \hat{y}$. Moreover, if $k < c\hat{y}$ and $x_i^* = \hat{y}$ then $x_j^* = 0$, for all $j \neq i$.*

This lemma tells us a great deal about the information accessed by individuals, but relatively little about the distribution of personal information acquisition, the aggregate information acquired in a social group and the structure of social communication. Our next result addresses these concerns.

⁹A similar result on equilibrium efforts in the context of exogenously fixed networks is derived in Bramoulle and Kranton 2007.

PROPOSITION 1: Suppose payoffs are given by (1). If $k > c\hat{y}$ then there exists a unique equilibrium in which every player acquires information \hat{y} and no one forms any links. Suppose $k < c\hat{y}$ and let $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$ be an equilibrium.

1. If $\sum_{i \in N} x_i^* = \hat{y}$ then \mathbf{g}^* is a core-periphery network, hubs acquire information personally and spokes acquire no information personally.
2. If $\sum_{i \in N} x_i^* > \hat{y}$ then:
 - 2.1 Every player $i \in I(\mathbf{s}^*)$ chooses $x_i^* = \frac{\hat{y}}{\Delta+1} = \frac{k}{c}$ and has $\Delta \in \{1, \dots, n-2\}$ links within $I(\mathbf{s}^*)$, while every player $j \notin I(\mathbf{s}^*)$ forms $\Delta+1$ links with players in $I(\mathbf{s}^*)$.
 - 2.2 High information level players choose $\bar{x}^* = \frac{k}{c}$, low information level players have η links with high information level players, they are not neighbors of each other and choose information $\underline{x}^* = \hat{y} - \eta \frac{k}{c}$, where $\frac{\hat{y}c}{k} - 1 < \eta < \frac{\hat{y}c}{k}$.

If the cost of a link exceeds the cost of acquiring the threshold information, $k > c\hat{y}$, then no one forms any links. Otherwise every equilibrium is characterized by linking activity, i.e., the network is non-empty. Figure 2 illustrates equilibrium outcomes for $n = 8$, $\hat{y} = 1$ and $k < c$. There are two types of equilibria: one, where aggregate information is equal to \hat{y} (see Figure 2A) and two, where it exceeds \hat{y} (see Figure 2B). When aggregate information equals \hat{y} , equilibrium networks are connected, they have the core-periphery structure, and players in the core collect all information—information hubs. Moreover, as the relative cost of linking k/c grows the number of information hubs decreases, each information hub collects more information and the total number of links decreases. When $k/c \in (\hat{y}/2, \hat{y})$ there is only one information hub and the social communication structure takes the form of a periphery sponsored star.

When aggregate information exceeds \hat{y} , equilibrium networks may not be connected and players acquire at most two levels of information. A key feature of these equilibria is that there exist players who are accessed by some but not by other players. This means that the cost of linking is *exactly* equal to the cost of information acquired by such a player. But then other players are indifferent between forming a link with this agent or acquiring information personally. In other words, the strategies of the players are not a *strict* best

Figure 2A: Nash equilibria in which aggregate information acquisition is 1.

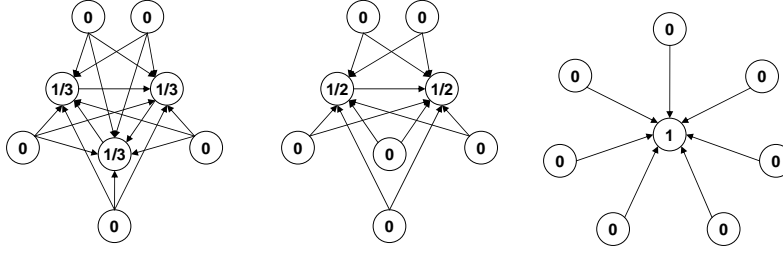


Figure 2B: Nash equilibria in which aggregate information acquisition exceeds 1

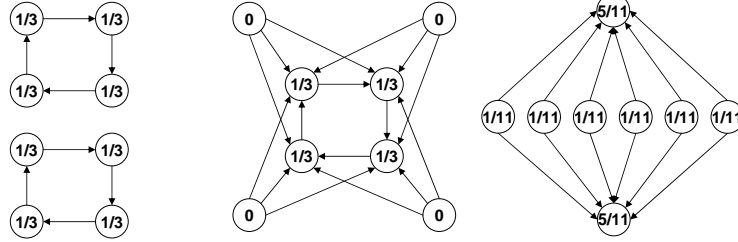


Figure 2: Example of Nash Equilibria, $n = 8$, $\hat{y} = 1$.

response to the strategies of others. The following result brings out the general implications of this observation.

PROPOSITION 2: Suppose payoffs are given by (1) and $k < c\hat{y}$. In every strict equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$: (1.) $\sum_{i \in N} x_i^* = \hat{y}$, (2.) the network has a core-periphery architecture, hubs acquire information personally and spokes acquire no information personally, and (3.) for given c and k , the ratio $|I(\mathbf{s}^*)|/n \rightarrow 0$ as $n \rightarrow \infty$.

If a positive information player strictly prefers to retain a link with another player j , it must be the case that the costs of linking with j are strictly lower than the cost of the information that player j acquires. But then all other players find it optimal to access j 's information and so all positive information players are linked among themselves. Lemma 1 tells us that any players who acquire information personally must access exactly \hat{y} information. Hence the desired conclusion: aggregate information acquisition must equal \hat{y} . Part 2 of the Proposition 2 then follows from Proposition 1.

The last part of the result derives bounds on the number of information hubs. If player i links with j then the cost of the link must be less than the cost of personally acquiring the information accessed from j , i.e., $cx_j > k$, and so player j must be acquiring at least k/c information. Since aggregate information is \hat{y} , it follows that there is an upper bound on

the total number of information hubs, and this number is independent of n . It then follows that the ratio $|I(\mathbf{s}^*)|/n$ can be made arbitrarily small by suitably raising n . Proposition 2 shows that these properties arise in *every* equilibrium of our game.¹⁰ Thus, the “law of the few” obtains as a consequence of strategic interaction among ex-ante identical and rational individuals.

Proposition 2 tells us that the number of information hubs is small relative to the number of players in large societies. But for fixed n , it does not determine the number of information hubs nor does it tell us who the hub players are. As discussed in the introduction, a recurring theme in the empirical literature is that even though information hubs seem to have similar demographic characteristics than others, they have distinctive attitudes that include higher attention to general market information and greater enjoyment of collecting information. A natural way to model this is to suppose that some players have slightly lower costs of collecting information. We consider a situation where $c_i = c$ for all $i \neq 1$, while $c_1 = c - \epsilon > 0$, where $\epsilon > 0$ is a small number. Let $\hat{y}_1 = \arg \max_y f(y) - c_1 y$. Clearly, as long as $\epsilon > 0$, $\hat{y}_1 > \hat{y}$, and $\hat{y}_1 \rightarrow \hat{y}$ as $\epsilon \rightarrow 0$. We focus on strict Nash equilibria.

PROPOSITION 3: *Suppose payoffs are given by (1), $c_i = c$ for all $i \neq 1$ and $c_1 = c - \epsilon$, $\epsilon > 0$. If $k < f(\hat{y}_1) - f(\hat{y}) + c\hat{y}$ then in a strict equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$: (1.) $\sum_{i \in N} x_i^* = \hat{y}_1$, (2.) the network is a periphery-sponsored star and player 1 is the hub, and (3.) either $x_1^* = \hat{y}_1$ and spokes choose $x^* = 0$, OR, $x_1^* = [(n - 1)\hat{y} - \hat{y}_1]/[n - 2]$ and $x^* = [\hat{y}_1 - \hat{y}]/[n - 2]$.*

Proposition 3 shows that a very small difference in the cost of collecting information is sufficient to separate the player who will act as information hub from those who will acquire little or not information personally and will only form connections.¹¹

This result builds on the ideas developed in Proposition 2. First, observe that for the low cost player the optimal information level is greater than the optimal information level for other players, i.e., $\hat{y}_1 > \hat{y}$. From earlier arguments in Proposition 2 we know that aggregate information collected by all players other than 1 will be at most \hat{y} . This implies that in

¹⁰Indeed, part 3 of Proposition 2 can be strengthened to read: for every strict equilibrium, $\lim_{n \rightarrow \infty} |I(\mathbf{s}^*)|/n^\alpha \rightarrow 0$, for all $\alpha > 0$. We thank a referee for this observation.

¹¹We have focused on slight differences in costs of acquiring information; analogous arguments show that if one player derives greater marginal benefits from acquiring information, as compared to others, then he will constitute the hub of the social network and acquire more information than the others.

equilibrium player 1 must acquire information personally, $x_1 > 0$. If $x_1 = \hat{y}_1$, the best reply of every other player is to acquire no information and to form a link with player 1. In case that $x_1 < \hat{y}_1$, we know, from Lemma 1, that $x_1 + y_1 = \hat{y}_1$ and so there is a player $i \neq 1$ with $x_i > 0$ and $x_i + y_i = \hat{y}$. If some player wants to link with i then so must everyone else. But then player i accesses all information \hat{y}_1 , which contradicts Lemma 1. Thus *no* player must have a link with player $i \neq 1$ in equilibrium. Hence, $i \neq 1$ must form a link to player 1, and, from optimality of linking, so must every other player. Finally, since every player is choosing positive effort, the equilibrium values of x_1 and x_i can be derived from the two equations $x_1 + (n - 1)x_i = \hat{y}_1$ and $x_1 + x_i = \hat{y}$.¹²

A. Efficient Outcomes

Given their salience it is important to understand the welfare properties of specialization in information acquisition and social communication. Proposition 1 tells us that in a Nash equilibrium for every player $i \in N$, $x_i + y_i = \hat{y}$. Thus, in every equilibrium, aggregate gross returns are $nf(\hat{y})$. If $k < c\hat{y}$, given the linearity in costs of information and linking, the efficient equilibrium minimizes the total costs of information and links. This immediately implies that the efficient equilibrium is a periphery-sponsored star network in which the hub collects information \hat{y} and every spoke chooses 0.

However, the public good nature of individual information acquisition implies that so long as equilibrium entails any links, there will be under-provision of information acquisition relative to the social optimum. To see this, note that in the star the hub player chooses \hat{y} , and at this point $f'(\hat{y}) = c$. But marginal social returns are given by $nf'(\hat{y})$, which are larger than c , for $n \geq 2$. The following proposition characterizes efficient outcomes.

PROPOSITION 4: *Suppose payoffs are given by (1). For every c , there exists a $\bar{k} > c\hat{y}$ such that if $k < \bar{k}$ then the socially optimal outcome is a star network in which the hub chooses \tilde{y}*

¹²In a recent paper on network formation with *exogenous* information levels, Hojman and Szeidl 2008 obtain a result on how small differences between players can help select the identity of the hub players. Their result relies on stochastic stability type arguments. Our result shows that heterogeneity has powerful effects in the selection of hubs when we endogenize information collection. In particular, with endogenous information acquisition, the selection of hubs obtains within a *one-shot static* model.

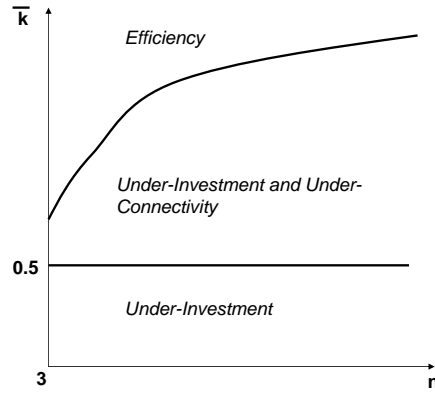


Figure 3: Equilibrium and Efficient Outcomes

(where $nf'(\tilde{y}) = c$), while all other players choose 0. If $k > \bar{k}$, then in the socially optimal outcome every player chooses \hat{y} and no one forms links.

The key point to note is that given any profile of information acquisition and linking, there is a corresponding star network in which the hub does all the information acquisition, which is strictly better. This is a consequence of the linear costs of information collection and the positive costs of linking. So, if the optimal social organization is a non-empty network then it must be a star where the hub acquires all the information. The value of \bar{k} is obtained by equating the best social welfare attained in the empty network and the star network. The following example illustrates the relation between equilibrium and socially efficient outcomes.

Example 1: Suppose $c = 1/2$ and $f(y) = \ln(1 + y)$. In this case $\hat{y} = 1$, while $\tilde{y} = 2n - 1$. In Figure 3 we plot \bar{k} as a function of the number of players. For a given n there are three regions. For low costs of linking, $k < 1/2$, the efficient equilibrium is a star where the hub acquires 1 and the spokes choose 0. As compared to socially optimal outcomes, in equilibrium there is under investment in information. For moderate costs of linking, $k \in (1/2, \bar{k})$, in equilibrium we have under investment and under connectivity relative to socially optimal outcomes (noting that $\bar{k} > 1/2$). In the remaining region, equilibrium outcomes coincide with socially optimal outcomes. ■

III. Indirect flow of information

In the basic model, a person can either acquire information personally or get it from another person who has directly acquired it herself. In this context, Propositions 2 and 3

show that equilibrium leads to core-periphery social communication structures where players in the core are information hubs. In actual practice, we often receive information from friends and colleagues which they have themselves received from their friends. The aim of this section is to examine the implications of this form of indirect information transmission for personal information acquisition and the social structure of communication. We find that information spillovers give rise to a new sort of influential agent: *the connector*. Similarly to an information hub, a connector occupies a central position in the social communication structure. However, differently from an information hub, a connector does not collect information personally but acts as an intermediary between other people who collect information.

We say that there is a path in $\bar{\mathbf{g}}$ between i and j if either $\bar{g}_{ij} = 1$ or there exist players j_1, \dots, j_m distinct from each other and i and j such that $\{\bar{g}_{ij_1} = \bar{g}_{j_1j_2} = \dots = \bar{g}_{j_mj} = 1\}$. A network $\bar{\mathbf{g}}$ is connected if there exists a path between every pair of players; we say that a network $\bar{\mathbf{g}}$ is minimally connected if it is connected and there exists only one path between every pair of players. Given two players i and j in \mathbf{g} , the geodesic distance, $d(i, j; \bar{\mathbf{g}})$, is defined as the length of the shortest path between i and j in $\bar{\mathbf{g}}$. If no such path exists, the distance is set equal to infinity. Let $N_i^l(\bar{\mathbf{g}}) = \{j \in N : d(i, j; \bar{\mathbf{g}}) = l\}$, that is $N_i^l(\bar{\mathbf{g}})$ is the set of players who are at distance l from i in $\bar{\mathbf{g}}$. We measure the level of spillovers by a vector $\mathbf{a} = \{a_1, a_2, \dots, a_{n-1}\}$, where $a_1 \geq a_2, \dots, \geq a_{n-1}$ and $a_l \in [0, 1]$ for all $l \in \{1, \dots, n-1\}$. It is assumed that if $j \in N_i^l(\bar{\mathbf{g}})$, then agent's j information accessed by i is $a_l x_j$. Observe that the case $a_1 = 1$ and $a_2 = 0$ corresponds to the pure local spillovers model analyzed in Section II. To bring out the role of indirect information transmission in the simplest form we start by considering the polar case of no decay or delay in flow across links, i.e., $a_{n-1} = 1$.¹³

Given a network $\bar{\mathbf{g}}$, let $y_i(\bar{\mathbf{g}})$ be the information which i derives from others, and define $y_{ij}(\bar{\mathbf{g}}) = y_i(\bar{\mathbf{g}}) - y_i(\bar{\mathbf{g}} - \bar{g}_{ij})$, i.e., the information which i accesses exclusively via j . The payoffs to player i under strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ can be written as:

$$\Pi_i(\mathbf{s}) = f \left(x_i + \sum_{l=1}^{n-1} \sum_{j \in N_i^l(\bar{\mathbf{g}})} x_j \right) - cx_i - \eta_i(\mathbf{g})k. \quad (4)$$

¹³See Hojman and Szeidl 2008 for an elegant model of inter-personal communication which leads to declining value of information with respect to distance in a social network.

Our next result obtains some properties of equilibria in the presence of frictionless information transmission.

PROPOSITION 5: *Suppose payoffs are given by (4). If $k > c\hat{y}$, there exists a unique equilibrium: every player personally acquires information \hat{y} and no one forms links. If $k < c\hat{y}$, then $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$ is an equilibrium if and only if: (1.) $\sum_{i \in N} x_i^* = \hat{y}$, (2.) $\bar{\mathbf{g}}^*$ is minimally connected and (3.) $k \leq cy_{ij}(\bar{\mathbf{g}}^*)$ for all $g_{ij}^* = 1, i, j \in N$.*

Frictionless information flow implies that equilibrium networks are minimal.¹⁴ From Lemma 1 we know that in equilibrium every individual must access at least \hat{y} information. If the costs of linking k are smaller than the costs of acquiring the threshold level of information \hat{y} , then standard considerations imply that the network is connected. Finally note that the costs of a link that player i forms with player j must be lower than the value of information that player i accesses exclusively via the link with j , i.e., $k \leq cy_{ij}(\bar{\mathbf{g}}^*)$. This implies that either player j collects enough information on his own, or that player j is a *connector*, and accesses others who have enough information.

We explore the distribution of information acquisition and the architecture of social communication via an examination of different classes of equilibria.

Hubs acquire information: Here the hubs are the players who personally collect information. Figure 4A illustrates these equilibria. At a superficial level these equilibria are similar to the core-periphery equilibria of the basic model. However, there is a key difference: in the present context, the information hubs acquire as well as aggregate information. This can be seen in the equilibrium on the right Figure 4A: each hub personally acquires information but also passes on information from the other hub.

Hub as Active Connector: The hub is an active connector in the sense that he acquires no information himself but forms links with all players who are acquiring information. All players who acquire no information in turn link with the hub. Figure 4B provides examples of such equilibria. An equilibrium with $|I(\mathbf{s}^*)|$ peripheral players collecting information exists whenever the costs of linking are smaller than the benefits of accessing a single information

¹⁴This builds on a result of Bala and Goyal 2000, which establishes minimality of equilibrium networks in a model of pure network formation and frictionless information flow.

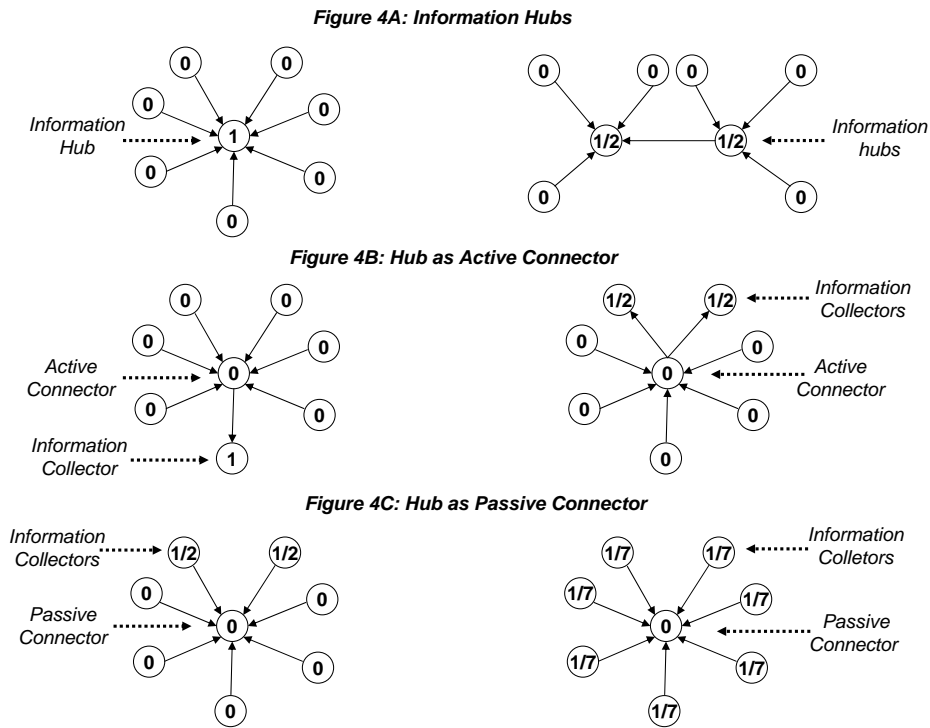


Figure 4: Information Hubs, Connectors and Others, $n = 8$.

collector, i.e., $k < \hat{c}y/|I(\mathbf{s}^*)|$. As k increases the number of information collectors falls and each positive information agent must acquire more information. This is in line with one of the results in the basic model: the number of players acquiring information declines with the cost of linking.

Hub as Passive Connector. Figure 4C illustrates this type of equilibrium. All players link with a single player who himself acquires no information but serves as a gateway to those who do. As costs of linking increase, initially, the number of players collecting information increases and each of them acquires less information. That is, an increase in costs of linking necessitates a greater number of active information sources. Once every peripheral player collects information, and costs of linking increase further, each information collector must acquire less information, which is only possible if the hub starts acquiring some information. In other words, an equilibrium where the hub acts as a passive connector is no longer sustainable.

We conclude this section with three remarks. *First*, the equilibria illustrated in Figure 4 exhibit patterns which are consistent with the empirical work on social communication. In their study of personal influence, Katz and Lazarsfeld 1955 emphasize that social influencers typically have more social ties and also acquire more information (via radio, newspapers and

television). We interpret this as a situation in which hubs are information hubs. In other instances, hubs acquire some – possibly small amount of – information personally but their numerous contacts provide new information which they then communicate to their neighbors and friends. Here the highly connected individual functions primarily as a connector. See Malcolm Gladwell 2000 for a discussion of such patterns of information acquisition and social communication and Cross et al. 2001 for a description of social connectors within firms.

Second, we note that the equilibria presented in Figure 4 are the only outcomes in the presence of small heterogeneities across players. In particular, if there is a player with slightly lower costs of acquiring information and slightly lower costs of being accessed by others, then such player becomes the only information hub and everyone else forms a link with him. If, instead, the player with the lowest cost of information acquisition and the most sociable player are different, then in the unique (strict) equilibrium the most sociable player is a hub who acts as a connector: he links with the low cost information player, who collects all the information, while all other players form a link with the connector. These claims are stated formally and proved in the appendix.

Our *final* remark is about a general model of information transmission with gradual decay. We note that perfect substitutability is a *necessary* condition for equilibrium where hubs collect information and spokes acquire no information. Indeed, in every equilibrium with information hubs the aggregate information acquisition is \hat{y} and there is always a player who relies at least partly on information collected by others. If $a_1 < 1$, then information available for this player is strictly less than \hat{y} , which contradicts Lemma 1.¹⁵ Next, suppose $a_1 < 1$ and $a_2 > a_3 \geq 0$. A periphery sponsored star network in which each of the peripheral players chooses $x = \hat{y}/[1 + a_2(n - 2)]$ while the central hub player collects *no* information personally is an equilibrium so long as n is sufficiently large. Hence, under imperfect substitutability of information we should expect hubs to function as connectors.

IV. Discussion

In this section we discuss two aspects of the model which play a prominent role in our

¹⁵Note also that perfect substitutability is *sufficient* for equilibria with information hubs: an equilibrium where the network is a periphery sponsored star and the hub acquires all the information exists so long as $a_1 = 1$ and $k < c\hat{y}$. We thank a referee for drawing our attention to this fact.

analysis: one, the linear costs of acquiring information and forming links and two, the link formation protocol.

A. Convex costs

The linear costs of acquiring information and forming links has the following implication: for any player acquiring information the total information accessed is \hat{y} and this level is independent of the amount of information acquired by the neighbors. However, if the costs of personally acquiring information are increasing and convex, this is no longer true. This section explores equilibrium outcomes under convex costs.

Define $z_i = x_i + \eta_i(\mathbf{g})k$, and let $C(z_i)$ satisfy the following properties: $C(0) = 0$, $C'(0) = C''(0) = 0$, $C'(z_i) > 0$, for $z_i > 0$, and $C''(z_i) > 0$, for $z_i > 0$. We focus on the case of frictionless information transmission, i.e., $a_{n-1} = 1$. Recall that $y_i \geq 0$ is the information accessed by player i from the others. The payoffs to a player from strategy \mathbf{s}_i faced with strategy profile \mathbf{s}_{-i} are then given by:

$$\Pi_i(\mathbf{s}_i, \mathbf{s}_{-i}) = f(x_i + y_i) - C(x_i + \eta_i(\mathbf{g})k). \quad (5)$$

We first note that in equilibrium a network is either empty or minimally connected.¹⁶ Moreover, the lack of decay in information transmission implies that in a minimally connected network every player accesses the information acquired by all players. Let y be the aggregate information acquired in equilibrium. For any player i who acquires information personally, the following first order condition must hold:

$$f'(y) = C'(x_i + \eta_i(\mathbf{g})k). \quad (6)$$

It then follows that players sponsoring an equal number of links must acquire an equal amount of information and this personal information acquisition is declining in the number of links sponsored. In any minimally connected network there are $n - 1$ links and so at least one player forms no links. Thus specialization in information acquisition remains an essential aspect of equilibrium behavior even under convex costs.

¹⁶This follows from standard arguments which rely on network externalities and the lack of decay in information transmission through the network, see e.g., Bala and Goyal 2000.

We now turn to aggregate information acquisition in equilibrium. Recall that in the basic model with linear costs, Propositions 2, 3, and 5 prove that aggregate information is invariant with regard to the number of players and the costs of forming links (so long as $k < c\hat{y}$). A greater number of players allows for smaller per capita acquisition of information; under convex costs this leads to lower marginal costs. So as we raise the number of players aggregate information should increase. Similarly, we expect that link formation costs will affect aggregate information, as these costs now enter the first order conditions of individual optimization (for all positive information players).

We now develop these intuitions formally. We focus on non-empty networks and suppose all players acquire information. Define \tilde{x} as the information acquired by the zero link player. From the optimality of equilibrium actions and equation (6) it follows that for the zero link player $f'(y) = C'(\tilde{x})$ (where y is aggregate information acquired), while for a player i with η_i links we have that $f'(y) = C'(x_i + \eta_i k)$. So we infer that for any player i , $x_i + \eta_i k = \tilde{x}$. Summing across all players we obtain an expression for aggregate information acquisition $y = n\tilde{x} - (n - 1)k$.

PROPOSITION 6: *Suppose that payoffs are given by (5) and that $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$ is a non-empty network equilibrium in which all players acquire information. Aggregate information is decreasing in the costs of linking and increasing in the number of players.*

Figure 5 illustrates these findings. Figure 5A presents periphery-sponsored star equilibria for $n = 4$ and different values of k . Figure 5B illustrates periphery-sponsored stars equilibria for $k = 0.1$ and different values of n .

Figure 5 also helps us explore role differentiation and social communication structures under convex costs in greater detail. To fix ideas, suppose that every player acquires information and the network is a periphery-sponsored star. From above arguments we know that in such an equilibrium the hub acquires information \tilde{x} , each peripheral player acquires information $x_p + k = \tilde{x}$ and $f'(n\tilde{x} - (n - 1)k) = C'(\tilde{x})$. Assume that $\lim_{x \rightarrow \infty} f'(x) \rightarrow 0$; since $C'(0) = 0$ and strictly increasing thereafter, it now follows that $x_p \rightarrow 0$ and that $\tilde{x} \rightarrow k$, as $n \rightarrow \infty$. The example in Figure 5B satisfies these hypotheses. We have thus shown that in large societies, there is an equilibrium which displays sharp role differentiation and

Figure 5A: Periphery sponsored star equilibrium, $n=4$, $k=0.1, 0.15, 0.2$.

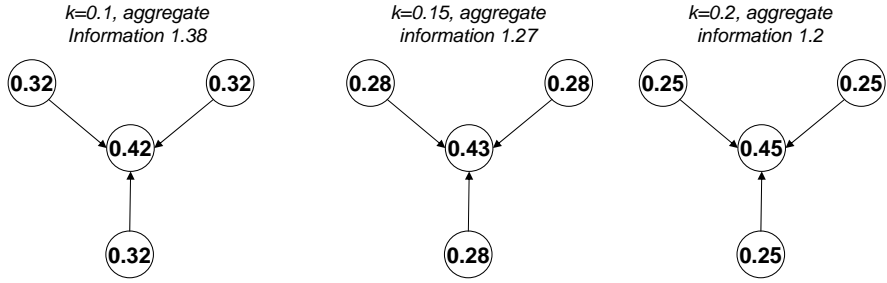


Figure 5B: Periphery sponsored star equilibrium, $k=0.1$, $n=4, 5, 6$.

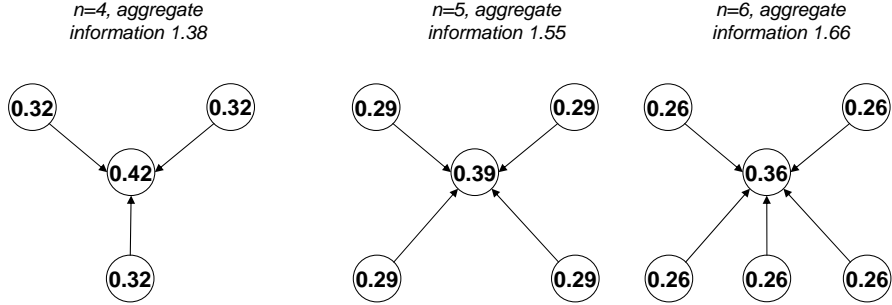


Figure 5: Aggregate Information under Convex Costs, $f(y) = \ln(1 + y)$, $C(z) = z^2/2$.

a core-periphery network of communication.

B. The link formation protocol

We have so far assumed that a player can unilaterally form a link with another player. This is a convenient and simple way to model link formation and the research on network formation over the last decade has shown that it offers a tractable framework to address a range of interesting questions. However, from a descriptive point of view, in social contexts it is more natural to view link formation as requiring the active participation of the two players involved in a link. This section examines role differentiation and core-periphery social communication in a model with two-sided network formation.

Suppose a link is formed only when both players agree and the costs of the link are shared equally between linked players. Formally, $\bar{g}_{ij} = 1$ if and only if $g_{ij}g_{ji} = 1$, and for each link $\bar{g}_{ij} = 1$ both players i and j pay $k/2$. The payoffs to player i facing a strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ are given by:

$$\Pi_i(\mathbf{s}) = f\left(x_i + \sum_{j \in N(i; \bar{\mathbf{g}})} x_j\right) - cx_i - \frac{k}{2} \sum_{j \in N} \bar{g}_{ij}. \quad (7)$$

Following the convention in this literature, we allow for players to delete links unilat-

erally. In addition, we allow for coordinated changes in information acquisition by players contemplating a new link. These considerations are summarized in the following solution concept.

DEFINITION 1. *A strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ is a pairwise-equilibrium if (1.) \mathbf{s} is a Nash Equilibrium and (2.) for all $\bar{g}_{ij} = 0$, if $\Pi_i(\bar{\mathbf{g}} + \bar{g}_{ij}, x'_i, x'_j, \mathbf{x}_{-ij}) > \Pi_i(\mathbf{s})$ then $\Pi_j(\bar{\mathbf{g}} + \bar{g}_{ij}, x'_i, x'_j, \mathbf{x}_{-ij}) < \Pi_j(\mathbf{s})$, $\forall x'_i, x'_j \in X$.*

We now provide a partial characterization of pairwise equilibrium.

PROPOSITION 7: *Suppose payoffs are given by (7) and $k < c\hat{y}$. In a pairwise equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$, $x_i^* \geq k/(2c)$ for all $i \in N$. Moreover, $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$ with a regular network $v \in \{1, \dots, n-2\}$ and each player choosing effort $x^* = \hat{y}/[v+1]$ is pairwise equilibrium if and only if $k \in [\hat{y}/(v+1), 2\hat{y}/(v+1)]$. The complete network with $x_i^* = x^* = \hat{y}/n$ is a pairwise equilibrium if and only if $k \leq 2c\hat{y}/n$.*

If link formation requires mutual acceptance *and* if costs of linking are equally shared between two players then every individual must acquire some information personally. Indeed, given costs of linking k and costs of acquiring information c , there is a lower bound to information acquired by each player and this is independent of the number of players. Moreover, regular social communication structures in which everyone collects the same amount of information is an equilibrium. These equilibria are in sharp contrast to our main findings with one-sided link formation. We now examine the sources of this difference in the predictions of the one-sided and two-sided models more closely.

Since social communication is costly a player who acquires information will agree to form a link with someone else who has no (or much less) information only if he is offered some compensation. More generally, an informed person may well ask for some compensation for his efforts at acquiring the information in the first place. This compensation may take the form of social favors or direct transfers (monetary or in kind). These ideas motivate a model of two-sided link formation with transfers.

Suppose each player chooses information acquisition and also proposes a set of transfers to every other player. The transfers may be positive (contributing to the costs of communi-

ation) or they may be negative (asking compensation for information acquired). Transfers proposed by player i are denoted by $\tau_i = \{\tau_{ij}\}_{j \in N}$ where $\tau_{ij} \in \mathcal{R}$. We assume that $\bar{g}_{ij} = 1$ if and only if $\tau_{ij} + \tau_{ji} \geq k$. So, a strategy profile $\mathbf{s} = (\mathbf{x}, \tau)$ specifies efforts \mathbf{x} and transfers τ . A strategy profile $\mathbf{s} = \{\mathbf{x}, \tau\}$ supports a core-periphery network $\bar{\mathbf{g}}(\tau)$ if there are two groups of players $N_1(\bar{\mathbf{g}})$ and $N_2(\bar{\mathbf{g}})$ such that (1.) $\tau_{ij} + \tau_{ji} < k$ for all $i, j \in N_1(\bar{\mathbf{g}})$, (2.) for all $i, j \in N_2(\bar{\mathbf{g}})$, $\tau_{ij} + \tau_{ji} \geq k$, and (3.) for all $i \in N_2(\bar{\mathbf{g}})$ and $j \in N_1(\bar{\mathbf{g}})$, $\tau_{ij} + \tau_{ji} \geq k$.

The payoffs to i given a profile $\mathbf{s} = (\mathbf{x}, \tau)$ are

$$\Pi_i(\mathbf{s}) = f \left(x_i + \sum_{j \in N(i; \bar{\mathbf{g}}(\tau))} x_j \right) - cx_i - \sum_{j \in N} \bar{g}_{ij}(\tau) \tau_{ij}. \quad (8)$$

Building on the work of Francis Bloch and Matthew Jackson (2007) we propose the following solution concept for our game with transfers.

DEFINITION 2. *A strategy profile $\mathbf{s} = (\mathbf{x}, \tau)$ is pairwise-equilibrium if and only if (1.) \mathbf{s} is a Nash Equilibrium and (2.) for all $\tau_{ij} + \tau_{ji} < k$, if $\Pi_i(\mathbf{s}_{-ij}, \tau'_{ij}, \tau'_{ji}, x'_i, x'_j) > \Pi_i(\mathbf{s})$ then $\Pi_j(\mathbf{s}_{-ij}, \tau'_{ij}, \tau'_{ji}, x'_i, x'_j) < \Pi_j(\mathbf{s})$, $\forall x'_i, x'_j \in X$ and $\forall \tau'_{ij}, \tau'_{ji}$.*

The following proposition shows that even when individuals ask for compensation for communicating the information they acquire, the core-periphery communication network with information hubs is an equilibrium. We assume $\epsilon > 0$.

PROPOSITION 8: *Consider the model with transfer. Suppose that payoffs are given by (8) and assume that $k < c\hat{y}$. The profile $\mathbf{s}^* = (\mathbf{x}, \tau)$ such that: (1.) $\forall i \in I(\mathbf{s}^*)$, $x_i^* = \hat{y}/|I(\mathbf{s}^*)|$, $\tau_{ij} = k/2$ for $j \in I(\mathbf{s}^*)$ and $\tau_{il} = -\epsilon cx^*$ for $l \notin I(\mathbf{s}^*)$ and (2.) $\forall l \notin I(\mathbf{s}^*)$, $x_l^* = 0$, $\tau_{li} = k + \epsilon cx_i^*$ for $j \in I(\mathbf{s}^*)$ and $\tau_{lk} = 0$, for $k \notin I(\mathbf{s}^*)$ is an equilibrium so long as $k + \epsilon c\hat{y}/I(\mathbf{s}^*) < c\hat{y}/I(\mathbf{s}^*)$.*

While core-periphery structures with information hubs are sustainable in equilibrium, regular networks with homogenous information acquisition are also equilibrium outcomes. For example, the following profile constitutes a pairwise equilibrium in the game with transfers if $k/c \in [\hat{y}/(v+1), 2\hat{y}/(v+1)]$: For each i , $x_i^* = c\hat{y}/[v+1]$ and there is a set of players $\{i_1, i_2, \dots, i_v\}$, such that $\tau_{ij} = k/2$ and $\tau_{ji} = k/2$ for exactly $j \in \{i_1, i_2, \dots, i_v\}$ players. These

different possibilities naturally raise a question about relative robustness of equilibria. This is an interesting question for future research.¹⁷

V. Conclusions

The determination of information people have when making decisions is a central problem in the social sciences. Our paper makes two contributions to the study of this problem. One, we develop a model in which players choose investments in personal information acquisition as well as in forming links with others to access their information. Two, we show that the *Law of the Few* – the phenomenon where a large majority of individuals get most of the information needed for their decisions from a very small subset of the group – is a robust equilibrium phenomenon in such a model.

There are two directions in which the analysis of this paper can be extended which appear to us to be especially promising. In actual practice individuals decide on information acquisition and links with others over time, and it is important to understand the dynamics. A second line of investigation concerns the relation between personally acquired information and information acquired by others. We have focused on the case where they are substitutes; it would be interesting to study the case where they are complements.

¹⁷We note that the incentives to acquire information are not affected by transfers as transfers are not conditional on information acquisition. This also limits the ability of transfers to facilitate efficient information acquisition.

Appendix A.

Given a network $\bar{\mathbf{g}}$, we define a component as a set $C(\bar{\mathbf{g}})$ of players such that $\forall i, j \in C(\bar{\mathbf{g}})$ there exists a path between them and there does not exist a path between $\forall i \in C(\bar{\mathbf{g}})$ and a player $j \in N \setminus C(\bar{\mathbf{g}})$. A component $C(\bar{\mathbf{g}})$ is non-singleton if $|C(\bar{\mathbf{g}})| > 1$. A player i is isolated if $\bar{g}_{ij} = 0, \forall j \in N \setminus \{i\}$.

PROOF OF LEMMA 1:

We first prove statement 1 in the lemma. Suppose not and $x_i + y_i < \hat{y}$ for some i in equilibrium. Under the maintained assumptions $f'(x_i + y_i) > c$ and so player i can strictly increase his payoffs by increasing personal information acquisition. Next suppose that $x_i > 0$ and $x_i + y_i > \hat{y}$. Under our assumptions on $f(\cdot)$ and c , if $x_i + y_i > \hat{y}$ then $f'(x_i + y_i) < c$; but then i can strictly increase payoffs by lowering personal information acquisition. This completes the proof of statement 1 in the lemma.

We now prove statement 2 in the lemma. Suppose that $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ is an equilibrium in which $x_i = \hat{y}$ and there is $x_j > 0$, for some $j \neq i$. First, since $x_i > 0$, it follows from the first part of Lemma 1 that $x_i + y_i = \hat{y}$. This also implies that every player in the neighborhood of i must collect no information personally. Now consider j , with $x_j > 0$. This means that $\bar{g}_{ij} = 0$. It follows from Lemma 1 that $x_j + y_j = \hat{y}$. If $x_j = \hat{y}$ then this player must get payoff $f(\hat{y}) - c\hat{y}$. If he switched to a link with i and reduced personal information acquisition to 0, his payoff is $f(\hat{y}) - k$. Since $k < c\hat{y}$, $x_j = \hat{y}$ is clearly not an optimal strategy for player j . So \mathbf{s} is not an equilibrium. Next suppose that $x_j < \hat{y}$. In equilibrium $x_j + y_j = \hat{y}$ and so there is some player $l \neq i$ such that $\bar{g}_{jl} = 1$ and $x_l \in (0, \hat{y})$. It is clear that if $g_{jl} = 1$ then player j can strictly increase his payoffs by switching the link from l to i . Similarly, if $g_{lj} = 1$, then player l gains strictly by switching link from j to i . So \mathbf{s} cannot be an equilibrium. A contradiction which completes the proof.

PROOF OF PROPOSITION 1:

The proof for the case $k > c\hat{y}$ is trivial and therefore omitted; we focus on $k < c\hat{y}$. First suppose that $\sum_{i \in N} x_i = \hat{y}$. In this case it follows from Lemma 1 that $I(\mathbf{s})$ must be a clique. Furthermore, $\bar{g}_{ij} = 0$ for all $i, j \notin I(\mathbf{s})$ and $g_{ij} = 0$ for all $i \in I(\mathbf{s}), j \notin I(\mathbf{s})$. Therefore,

each player choosing 0 must sponsor a link with every positive information player. This also shows that the network must be a core-periphery network.

Hereafter, let $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ be an equilibrium where $\sum_{i \in N} x_i > \hat{y}$. The proof for this case is developed in three steps. In the first step, we consider the case in which positive information players choose the same level of information. In the second step we consider situations in which positive information players choose different level of information. The third step uses the observations derived in the previous two steps to conclude the proof.

Step 1. We prove that if all players who acquire information personally choose the same level then \mathbf{s} satisfies Part 1 of Proposition 1. Suppose $x_i = x, \forall i \in I(\mathbf{s})$. If $x = \hat{y}$, Lemma 1 implies that $|I(\mathbf{s})| = 1$ and therefore aggregate information is \hat{y} , a contradiction. Assume $x \in (0, \hat{y})$; from Lemma 1 it follows that $x_i + y_i = \hat{y}, \forall i \in I(\mathbf{s})$. Since, by assumption, $x_i = x, \forall i \in I(\mathbf{s})$, it follows that every player who accesses information personally also gets an equal amount of information from his neighbors, which immediately implies that every positive information player has the same number of links with positive information players; let Δ be this number. Note that for all $i \in I(\mathbf{s}), x_i + y_i = x + \Delta x = \hat{y}$, which implies that $x = \frac{\hat{y}}{\Delta+1}$. Since aggregate information is strictly higher than \hat{y} it follows that $\Delta + 1 < |I(\mathbf{s})|$. Also, from Lemma 1 we know that $x < \hat{y}$, which implies that $\Delta \geq 1$. Thus, there exist two positive information players who are neighbors, implying that $k \leq cx$. Also, since, by assumption, $\sum_{i \in N} x_i > \hat{y}$, there exist two positive information players who are not neighbors, implying that $k \geq cx$. Hence, $k = cx$. Finally, if $I(\mathbf{s}) = N$, the proof follows. If not, select $j \notin I(\mathbf{s})$. Clearly, in equilibrium no player forms a link with j . So, in equilibrium j must sponsor $\Delta + 1$ links with positive information players. This concludes the proof of part 1 of the proposition.

Step 2. Let \mathbf{g}' be the subgraph of \mathbf{g} defined on $I(\mathbf{s})$. Let $C(\bar{\mathbf{g}}')$ be a component of $\bar{\mathbf{g}}'$. By construction each player in $C(\bar{\mathbf{g}}')$ chooses positive information. Suppose that (A1) total sum of information in $C(\bar{\mathbf{g}}')$ is strictly higher than \hat{y} and (A2) there exists at least a pair of players in $C(\bar{\mathbf{g}}')$ who choose a different level of information. The following Lemma is key.

LEMMA 2: Suppose that (A1) and (A2) hold in $C(\bar{\mathbf{g}}')$. Then there are two types of players in $C(\bar{\mathbf{g}}')$: high information players choose \bar{x} and low information players choose $\underline{x} < \bar{x}$. Moreover, every low information player forms η links with high information players, there

are no links between low information players, $k = c\bar{x}$, $\underline{x} = \hat{y} - \eta\bar{x}$ and $\frac{\hat{y}c}{k} - 1 < \eta < \frac{\hat{y}c}{k}$.

PROOF OF LEMMA 2:

Without loss of generality label players in $C(\bar{\mathbf{g}}')$, so that $\hat{y} > x_1 \geq x_2 \geq \dots \geq x_m$. (A2) implies that there exists $l \in C(\bar{\mathbf{g}}')$, $l \neq m$, such that $x_j = x_l = \bar{x}$, for all $j \leq l$, and $\bar{x} > x_{l+1}$. We start by proving two claims.

CLAIM 1: For all $j > l$, $g_{ji} = 1$ for some $i \leq l$.

PROOF OF CLAIM 1: Suppose that there exists a $j > l$ such that $g_{ji} = 0$, $\forall i \leq l$. This implies that j does not sponsor links. If, on the contrary, player j sponsors links, then these links are directed to players $j' > l$, but then player j could strictly gain by switching a link from j' to some $i \leq l$. Note that, it must also be the case that j does not receive any links. Suppose j receives a link from a player j' . Then it must be the case that player l is also a neighbor of j' , otherwise j' strictly gains by switching the link from j to l . But this says that every player who sponsors a link to j is l 's neighbor and since player j only receives links, this means that player j accesses from his neighbors at most as much information as player l does. Since $x_j + y_j = x_l + y_l = \hat{y}$, this implies that $x_j \geq x_l$, contradicting our hypothesis that $x_j < \bar{x} = x_l$. Thus j does not receive links. But then $x_j + y_j = x_j < \hat{y}$, which contradicts Lemma 1. Hence, claim 1 follows.

CLAIM 2: There exists some $i, i' \leq l$ such that $\bar{g}_{ii'} = 0$.

PROOF OF CLAIM 2: Suppose $\{1, \dots, l\}$ is a clique. Since aggregate information in component $C(\bar{\mathbf{g}})$ exceeds \hat{y} , it must be the case then that for every $i \leq l$, there is one player $j > l$ such that $\bar{g}_{ij} = 0$. Select such a player j . Clearly, $g_{jj'} = 0$ for all $j' > l$, otherwise j strictly gains by switching the link from j' to i . Analogously, if j receives a link from some $j' > l$, then i must also be a neighbor of j' . Therefore, since $\{1, \dots, l\}$ is a clique, it follows that every neighbor of j is also i 's neighbor, and this contradicts the hypothesis that $x_j < x_l = \bar{x}$. So $\bar{g}_{jj'} = 0$, $\forall j' > l$. Finally note that this implies that i accesses a superset of the players accessed by j , i.e., $y_i \geq y_j$. We know that $x_i + y_i = \hat{y} = x_j + y_j$ and so $x_j \geq x_i$, which contradicts the hypothesis that $j > l$. Claim 2 follows.

We can now conclude the proof of Lemma 2. From claim 1, there exists a player $j > l$ who sponsors a link to a player $i \leq l$, so $k \leq c\bar{x}$. Similarly, claim 2 implies that there exists $i, i' \leq l$ such that $\bar{g}_{ii'} = 0$; this implies $k \geq c\bar{x}$. Hence, we have $k = c\bar{x}$. Next, since $k = c\bar{x}$ and $x_j < \bar{x}$ for all $j > l$, it follows that $\bar{g}_{j'j} = 0$ for all $j, j' > l$. Therefore, every player $j > l$ forms only links with players in $\{1, \dots, l\}$. We now show that $x_j = x_{j+1}$ for all $j > l$. Select $j > l$ and assume that $x_j > x_{j+1} > 0$. Then, $x_j + y_j = x_j + \eta_j(\mathbf{g})\bar{x}$ and $x_{j+1} + y_{j+1} = x_{j+1} + \eta_{j+1}(\mathbf{g})\bar{x}$. Lemma 1 implies that $x_j + y_j = x_{j+1} + y_{j+1} = \hat{y}$, which holds whenever $x_j - x_{j+1} = (\eta_{j+1}(\mathbf{g}) - \eta_j(\mathbf{g}))\bar{x}$. Since $x_j > x_{j+1}$, then $\eta_{j+1}(\mathbf{g}) - \eta_j(\mathbf{g}) \geq 1$, but then $(\eta_{j+1}(\mathbf{g}) - \eta_j(\mathbf{g}))\bar{x} \geq \bar{x} > x_j - x_{j+1}$, where the last inequality follows because, by assumption, $x_j < \bar{x}$. Thus, all players $j > l$ choose the same information, say \underline{x} , and from Lemma 1 it follows that $\underline{x} + \eta_j(\mathbf{g})\bar{x} = \hat{y}$. Thus, every low information player sponsors the same number of links with high information players, say η , and $\underline{x} + \eta\bar{x} = \hat{y}$. This concludes the proof of Lemma 2.

Step 3: We now conclude the proof of Proposition 1. Recall that \mathbf{g}' is the subgraph of \mathbf{g} defined on $I(\mathbf{s})$. We need to consider two cases: one, $\bar{\mathbf{g}}'$ is connected, and two $\bar{\mathbf{g}}'$ is not connected.

$\bar{\mathbf{g}}'$ is connected: first observe that (A1) holds by assumption. If all positive information players choose same action then step 1 applies and the proof follows. If (A2) holds then Lemma 2 applies. We next observe that in this case every player $i \in N$ must choose positive information. To see this note that since $k = c\bar{x}$ every player $j \notin I(\mathbf{s})$ will only sponsor links to high information players. Then, by symmetry, low information players must obtain the same payoffs as players $j \notin I(\mathbf{s})$. It is easy to check that this is possible if and only if $\underline{x} = \bar{x}$, which contradicts (A2).

$\bar{\mathbf{g}}'$ is not connected; Let C_1 and C_2 be two components in $\bar{\mathbf{g}}'$. We observe that $x_i < \hat{y}$, and from Lemma 1 it follows that the components must contain at least two players each. Here, note that for every $i, i' \in C_1$ and $j, j' \in C_2$ such that $g_{ii'} = g_{jj'} = 1$, $x_{i'} = x_{j'} = x \geq x_i, x_j$ and $k = cx$. Indeed, $x_{i'} = x_{j'} = x$ follows because, if $x_{i'} < x_{j'}$ then player i would strictly gain by switching a link from i' to j' ; for analogous reasons it follows that $x_i, x_j \leq x$; Since i sponsors a link to i' , $k \leq cx$, while i' does not sponsor a link to j' , and so $k \geq cx$. Thus

$k = cx$. Together, these observations imply that every player who receives a link in C_1 and every player who receives a link in C_2 chooses information x . Thus, if in C_1 and C_2 every positive player receives at least one link, every player chooses the same information and the proof follows from step 1.

Suppose next that there is some player in C_1 who does not receive a link *and* information acquisition is not equal across players. If the aggregate information in C_1 equals \hat{y} , then C_1 is a clique and therefore there is at most one player who only sponsors links and receives no links. Since C_1 is a clique and aggregate information is \hat{y} , this player will choose $\underline{x} = \hat{y} - (|C_1| - 1)x$.

Finally consider the case where aggregate information in C_1 exceeds \hat{y} , and personal information acquisition is not equal. Then Lemma 2 applies and there are two positive information acquisition levels, x and x' , with $x' < x$. We observe that as in the case of connected network above, it is possible to rule out j such that $x_j = 0$. Since C_1 was arbitrary, this completes the proof of Proposition 1.

PROOF OF PROPOSITION 2:

We first prove that in every strict Nash equilibrium $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ the aggregate information is equal to \hat{y} , i.e., $\sum_{i \in N} x_i = \hat{y}$. Let $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ be a Nash equilibrium in which aggregate information exceeds \hat{y} . First, suppose s satisfies part 1 of Proposition 1. We know that a positive information player i chooses $x_i = k/c < \hat{y}$ and forms Δ links such that $x_i[\Delta + 1] = \hat{y}$, and that $\Delta + 1 < I(\mathbf{s})$. Moreover $cx = k$. Then it is immediate that this player is indifferent between a link and acquiring additional information k/c . This means that equilibria in part 1 are not strict. Second, suppose s satisfies part 2 of Proposition 1. Again a player with positive information acquisition is indifferent between forming a link and acquiring extra information himself, since $c\bar{x} = k$. Hence, equilibria in part 2 of the proposition are not strict. Taken together with Proposition 1 this implies that in every strict equilibrium aggregate information acquisition is equal to \hat{y} . The core-periphery architecture of equilibrium networks follows directly from Proposition 1.

We now consider the proportion $|I(\mathbf{s})|/n$ in every strict equilibrium \mathbf{s} . Recall that in a strict equilibrium $\sum_{i \in N} x_i = \hat{y}$ and that $x_i + y_i \geq \hat{y}$, for all $i \in N$. This means that every

player who acquires information personally is accessed by every player in equilibrium. This implies that there is at most one player i with no incoming links, i.e., $g_{ji} = 0$, for all $j \in N$. For all other players $l \in I(\mathbf{s})$, it must be the case there is at least one player j such that $g_{jl} = 1$; but this implies that $x_l > k/c$. So the number of accessed players who acquire information personally and have incoming links, $I(\mathbf{s}) - 1$, is bounded above by $(\hat{y}c)/k$. It follows that $I(\mathbf{s})/n \leq ([\hat{y}c/k] + 1)/n$, which can be made arbitrarily small by raising n . This completes the proof.

PROOF OF PROPOSITION 3:

It is immediate to see that if $x_1 = \hat{y}_1$ then the proposition follows. Next, if $x_1 = 0$ then we can use Proposition 2 to show that in a strict equilibrium aggregate information equals \hat{y} . Note however that if $x_1 = 0$, player 1 must access at least $\hat{y}_1 > \hat{y}$ from his neighbors, a contradiction. We now take up the case of $x_1 \in (0, \hat{y}_1)$.

CLAIM 3: $\forall i, j \in I(\mathbf{s}) \setminus \{1\}$, if $\bar{g}_{ij} = 1$ then i and j share the same neighbors, i.e., for every $l \in I(\mathbf{s}) \setminus \{i, j\}$, $l \in N_i(\bar{\mathbf{g}})$ if and only if $l \in N_j(\bar{\mathbf{g}})$.

PROOF OF CLAIM 3: Let $\bar{g}_{ij} = 1$, $i, j \in I(\mathbf{s}) \setminus \{1\}$, and suppose, without loss of generality, that $x_i \leq x_j$. We first prove that for every $l \in I(\mathbf{s}) \setminus \{i, j\}$, if $l \in N_i(\bar{\mathbf{g}})$ then $l \in N_j(\bar{\mathbf{g}})$. Suppose not and there exists a player $l \in I(\mathbf{s})$, with $l \in N_i(\bar{\mathbf{g}})$ and $l \notin N_j(\bar{\mathbf{g}})$. If $g_{li} = 1$, then, since $x_i \leq x_j$, l (weakly) gains by switching the link from i to j . Hence, let $g_{il} = 1$. Since $x_i > 0$, it follows from Lemma 1 that $x_i + y_i = \hat{y}$ and the payoffs to i in equilibrium \mathbf{s} are $f(\hat{y}) - cx_i - \eta_i(\mathbf{g})k$. Suppose that i deletes the link with player l and chooses an information level $\tilde{x}_i = x_i + x_l$, then he obtains payoffs $f(\hat{y}) - cx_i - cx_l - (\eta_i(\mathbf{g}) - 1)k$. Since \mathbf{s} is a strict equilibrium this deviation strictly decreases i 's payoffs, which requires that $k < cx_l$. Let $k < cx_l$ and consider the following two possibilities.

(1:) $x_j \geq x_l$. In this case, since $\bar{g}_{jl} = 0$, and since \mathbf{s} is a strict equilibrium, player j must strictly loose if he forms an additional link with l and chooses information level $\tilde{x}_j = x_j - x_l$. That is, $f(\hat{y}) - cx_j - \eta_j(\mathbf{g})k > f(\hat{y}) - c(x_j - x_l) - (\eta_j(\mathbf{g}) + 1)k$, which holds if and only if $k > cx_l$; but this contradicts that $k < cx_l$.

(2:) $x_j < x_l$. Here we have two sub-cases. (2a:) Suppose $g_{ij} = 1$; this implies that the costs

for i to link with j are strictly lower than the costs of information that i accesses from j , i.e., $k < cx_j$. Since $k < cx_j$, $\bar{g}_{lj} = 0$ and, by assumption, $x_l > x_j$, then l strictly gains if he links with j and chooses information level $\tilde{x}_l = x_l - x_j$. So \mathbf{s} is not a strict equilibrium.

(2b:) Suppose $g_{ji} = 1$. Since j does not access l but he sponsors a link to i , it follows that $x_i > x_l$. Since, by assumption, $x_j < x_l$, it follows that $x_i > x_l > x_j$, which contradicts that $x_i \leq x_j$. We have then shown that for every $l \in I(\mathbf{s}) \setminus \{i, j\}$, if $l \in N_i(\bar{\mathbf{g}})$ then $l \in N_j(\bar{\mathbf{g}})$.

We now show that if $l \in I(\mathbf{s}) \setminus \{i, j\}$ and $l \in N_j(\bar{\mathbf{g}})$ then $l \in N_i(\bar{\mathbf{g}})$. Suppose not; then player j accesses all positive information players that i accesses plus some other positive information players. But this would contradict that $x_i \leq x_j$. This concludes the proof of Claim 3.

CLAIM 4: *Suppose $i, j \in N_1(\bar{\mathbf{g}})$, $i, j \in I(\mathbf{s})$, and $\bar{g}_{ij} = 0$, then $\bar{g}_{li} = \bar{g}_{lj} = 0$ for all $l \neq 1$, $l \in I(\mathbf{s})$.*

PROOF OF CLAIM 4: Suppose, without loss of generality, $x_i \leq x_j$. We first show that $\bar{g}_{li} = 0$ for all $l \neq 1$, $l \in I(\mathbf{s})$. Suppose, on the contrary, that $\bar{g}_{li} = 1$, for some $l \neq 1$, $l \in I(\mathbf{s})$. In view of claim 3, since $\bar{g}_{ij} = 0$, then $\bar{g}_{lj} = 0$; this fact and $x_i \leq x_j$ implies that $g_{il} = 1$, and since $\bar{g}_{ij} = 0$, then from the strictness of equilibrium, it follows that $x_l > x_j$. Since $x_l > x_j$, $g_{il} = 1$, and $x_j + y_j = \hat{y} = x_i + y_i$, it must be the case that there exists some $l' \in I(\mathbf{s})$, $l' \in N_j(\bar{\mathbf{g}})$ and $l' \notin N_i(\bar{\mathbf{g}})$. Claim 3 implies that $l' \notin N_l(\bar{\mathbf{g}})$. Since $x_l > x_j$ and $l' \notin N_l(\bar{\mathbf{g}})$, then $g_{jl'} = 1$. But $g_{jl'} = 1$ and $l' \notin N_j(\bar{\mathbf{g}})$ implies that $x_{l'} > x_l$; similarly, $g_{il} = 1$ and $l' \notin N_i(\bar{\mathbf{g}})$ implies that $x_l > x_{l'}$, a contradiction. Thus, the only neighbor of i is player 1. It is easy to see that the same holds for player j . Indeed, if $l \in I(\mathbf{s})$ and $l \in N_j(\bar{\mathbf{g}})$, then claim 3 implies that $l \notin N_i(\bar{\mathbf{g}})$, but then player j accesses strictly higher information than player i , which contradicts our initial hypothesis that $x_j \geq x_i$. Claim 4 follows.

Final step in proof of Proposition 3: We are concerned with the case $x_1 \in (0, \hat{y}_1)$. Since $x_1 + y_1 = \hat{y}_1$, there exists some $i \in I(\mathbf{s})$ such that $i \in N_1(\bar{\mathbf{g}})$. Observe that given such an i , there exists a $j \in N_1(\bar{\mathbf{g}})$ such that $j \notin N_i(\bar{\mathbf{g}})$ and $x_j > 0$. This is because otherwise $x_i + y_i \geq \hat{y}_1 > \hat{y}$, and this contradicts $x_i > 0$ and Lemma 1. From Claim 4 above we know that players i and j do not have any links with players in $I(\mathbf{s})$. This means that

$x_i + x_1 = \hat{y} = x_j + x_1 = \hat{y}$, and so $x_i = x_j$. Given that we are in a strict equilibrium, it follows that $k < cx_i$. This implies that i and j choose the same information and form a link with 1. We observe that this also means that $\hat{y} > x_1 > x_i = x_j > 0$.

We now show that the player 1 constitutes the hub of his component and that all other players behave as players i and j identified above. In a path of length of two or more starting at 1, there are four possible patterns: two players choosing 0, two players choosing positive information, and two mixed cases. Clearly it is not possible to have two players choosing 0 as the costs of linking are strictly positive. Next consider a positive information player followed by a zero information player. Suppose player $l \notin I(\mathbf{s})$ and suppose there is a link with player m such that $\bar{g}_{m1} = 1$. Since $l \notin I(\mathbf{s})$, it must be the case that $g_{lm} = 1$ and so $m \in I(\mathbf{s})$. However, it is profitable for l to form this link only if $k < cx_m$. Moreover, claim 3 implies that $\bar{g}_{mi} = 0$; if $x_i \geq x_m$, then, since $k < cx_m$, player i strictly gains by forming a link with m . So $x_i < x_m$ and since $\bar{g}_{mi} = 0$ it follows that $k > cx_i$. Putting together these facts we get that $x_m > x_i$, and so $x_m + y_m \geq x_m + x_1 > x_i + x_1 = \hat{y}$. This contradicts Lemma 1. Thus we have ruled out case 2. The case of two positive levels of information acquisition is ruled out by noting that in that case there is a sequence of players 1, l and l' such that $x_1 + x_l + x_{l'} \leq \hat{y}$, but this means that $x_l, x_{l'} < x_i$, and so a link $\bar{g}_{ll'} = 1$ is not profitable for either l or l' . The last case to consider has a sequence 1, l and l' , with $x_l = 0$ and $x_{l'} > 0$. Clearly then $g_{ll'} = 1$ and so $cx_{l'} > k$. If $x_{l'} > x_1$, then it is strictly profitable for l' to lower information acquisition and form a link with 1, while if $x_{l'} < x_1$ then it is strictly profitable for 1 to lower information acquisition and form a link with l' . We have thus shown there cannot exist a path of length of two or more starting at player 1. So player 1 constitutes a hub. Now we can exploit the fact that there exists a player i such that $g_{i1} = 1$ and $x_i + x_1 = \hat{y}$ to conclude that there cannot exist any links between the neighbors of player 1. This proves that player 1 is a hub of his component and that all other players choose information level x and form a link with player 1.

The above argument is done for a single component. The connectedness of non-empty strict equilibrium networks follows from standard arguments; the details are omitted. Finally, note that if $x_1 = \hat{y}_1$, then $x_i = 0$ for all $i \neq 1$ and therefore property (3) follows. Suppose $x_1 \in (0, \hat{y}_1)$, then we know that each player $i \neq 1$ chooses $x_i = x$. Since spokes sponsor only

a link to the hub, in a strict equilibrium it must be the case that $x_1 > x$. Furthermore, for a player i to play x is optimal only if $x = \hat{y} - x_1$ and, similarly, for player 1 to play x_1 is optimal only if $x_1 + (n - 1)x = \hat{y}_1$. It is now easy to see that as $\epsilon \rightarrow 0$, then $x \rightarrow 0$.

PROOF OF PROPOSITION 4:

Suppose $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ corresponds to an efficient profile. We first show that if \mathbf{g} is not empty, then \mathbf{g} is a star. Let \mathbf{g} be a non-empty network and suppose that C is a component in \mathbf{g} . Let $|C| \geq 3$ be the number of players in C . Suppose that y is the total information acquired in component C . Then it follows that the total payoff of all players in component C is at most $|C|f(y) - cy - (|C| - 1)k$. Consider a star network with $|C|$ players in which the hub player alone chooses y . It then follows that this configuration attains the maximum possible aggregate payoff given effort y . Moreover, note that aggregate payoff in any profile \mathbf{s} , in which two or more players acquire positive information is strictly less than this, since it will entail the same total costs of information acquisition and a strictly higher cost of linking or a strictly lower payoff to at least one of the players. So the star network with the hub acquiring all the information personally is the optimal profile for each component.

Next consider two or more components in an efficient profile \mathbf{s} . It is easy to see that in a component of size m , efficiency dictates that information y satisfy $mf'(y) = c$. If the components are of unequal size then information acquisition efforts will be unequal and a simple switching of spoke players across components raises social welfare. So in any efficient profile with two or more components, the components must be of equal size. Let m be the size and let the effort y satisfy $mf'(y) = c$. Suppose now that the network contains two components C_1 and C_2 of size m . Consider the network in which the spoke players in component 2 are all switched to component 1. This yields a network \mathbf{g}' with components C'_1 and C'_2 with the former containing $2m - 1$ players while the latter contains 1 player. Then the payoff remains unchanged. However, the information level y is no longer optimal in either of the components. So, for instance, information can be lowered in component 2 and the aggregate payoff thereby strictly increased, under the assumptions on $f(\cdot)$. A similar argument also applies to networks with three or more components, and so we have proved that no profile with two or more components can be efficient. Thus, if \mathbf{g} is not empty then \mathbf{g}

is a star and the information of the central player is $\tilde{y} = \arg \max_{y \in X} nf(y) - cy$. The social welfare associated to such profile is: $SW = nf(\tilde{y}) - c\tilde{y} - (n-1)k$.

Finally, note that if \mathbf{s} is socially efficient and \mathbf{g} is not a star, then \mathbf{g} must be empty and every player will choose information \hat{y} . The social welfare is then $SW = n[f(\hat{y}) - c\hat{y}]$. The expression for \bar{k} is obtained by equating the social welfare in these two configurations, i.e. $(n-1)\bar{k} = n[f(\tilde{y}) - f(\hat{y})] + c[(n-1)\hat{y} - \tilde{y}] + c\hat{y}$. To see that $\bar{k} > c\hat{y}$, note that if $\bar{k} \leq c\hat{y}$, then $n[f(\tilde{y}) - f(\hat{y})] + c[(n-1)\hat{y} - \tilde{y}] + c\hat{y} \leq (n-1)c\hat{y}$, which holds if and only if $nf(\tilde{y}) - c\tilde{y} \leq nf(\hat{y}) - c\hat{y}$. Given that $\tilde{y} = \arg \max_{y \in X} nf(y) - cy$, $\hat{y} = \arg \max_{y \in X} f(y) - cy$ and that $f(\cdot)$ is strictly concave, the above inequality cannot hold. This concludes the proof of Proposition 4.

PROOF OF PROPOSITION 5:

The proof of Part 2 of Proposition 5 is straightforward and it is omitted. Hereafter, we focus on Part 1. We first prove that if \mathbf{s} satisfies properties (i.)-(iii.) in the proposition then \mathbf{s} is a Nash equilibrium. Take a player i ; since $\sum_{j \in N} x_j = \hat{y}$, $\bar{\mathbf{g}}$ is minimally connected, and there is no decay, then $x_i + y_i = \hat{y}$, so player i does not want to change his own information level and also he does not want to form an additional link. The payoffs to i at equilibrium \mathbf{s} are $f(\hat{y}) - cx_i - \eta_i(\mathbf{g})k$. If $\eta_i(\mathbf{g}) = 0$, then player i plays a best reply. Suppose $\eta_i(\mathbf{g}) > 0$, then $g_{ij} = 1$ for some j . Note that player i is indifferent between keeping the link with j and switching the link from j to a player that i accessed via j . Also, property (iii.) says that $k \leq cy_{ij}(\bar{\mathbf{g}})$ and therefore player i does not gain by deleting the link with j . Hence, \mathbf{s} is a Nash equilibrium.

We now prove the converse. Let $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ be an equilibrium. Frictionless information flow implies that every component of $\bar{\mathbf{g}}$ must be minimal. Also, frictionless information flow together with Lemma 1 imply that in every component the aggregate information is \hat{y} . Next, suppose $\bar{\mathbf{g}}$ is not connected. Let C_1 be a component of $\bar{\mathbf{g}}$. If $x_i = \hat{y}$ for some $i \in C_1$, then all i 's neighbors choose information 0 and sponsors a link to i , so i 's payoffs are $f(\hat{y}) - c\hat{y}$ and $k < c\hat{y}$. But then player i strictly gains if he chooses 0 and forms a link with a player $j \in C_2$. Thus, in C_1 there are at least two players choosing positive level of personal information acquisition; since C_1 is minimal it must be the case that there is a link $g_{i'j'} = 1$ such that

player i' accesses via the link with j' strictly less information than \hat{y} , say $z < \hat{y}$. It is then clear that if player i' deletes the link with j' and forms a new link with a player in C_2 , he will incur the same costs but he will access strictly higher information. Therefore player i' can strictly improve his payoffs, a contradiction. Thus, $\bar{\mathbf{g}}$ is connected. Finally, it is readily seen that if $g_{ij} = 1$ and \mathbf{s} is equilibrium, then $k \leq cy_{ij}(\bar{\mathbf{g}})$. This concludes the proof.

PROOF OF PROPOSITION 6:

Let $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ be a non-empty network equilibrium in which $x_i > 0$ for all i . Let $n(l)$ be the number of players who sponsor l links, $l = 0, \dots, n-1$; also let $x(l)$ be the information acquired by a player who sponsors l links. Note that in a minimally connected network there are $n-1$ links and n nodes and therefore $n(0) > 0$. Since every player acquires information, we have that $x(0) = x(l) + kl$. Thus, aggregate information is $\sum_{l=0}^{n-1} n(l)x(l) = \sum_{l=0}^{n-1} n(l)(x(0) - kl) = nx(0) - k(n-1)$. Since $x(0)$ is part of equilibrium, it is the solution to $f'(nx(0) - k(n-1)) = C'(x(0))$.

For given k , let $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ be a non-empty network equilibrium in which $x_i > 0$ for all $i \in N$; similarly, for given $k' > k$, let $\mathbf{s}' = (\mathbf{x}', \mathbf{g}')$ be a non-empty network equilibrium in which $x'_i > 0$ for all $i \in N$. Let y and y' be the aggregate information under \mathbf{s} and \mathbf{s}' , respectively. We claim that $y > y'$. Suppose $y \leq y'$, then $f'(y) = C'(x(0)) \geq f'(y') = C'(x'(0))$ (where recall that $x(0)$ is the information of a player who does not sponsor any links.); so, $x(0) \geq x'(0)$. But then, since $k' > k$, $y = nx(0) - (n-1)k \geq nx'(0) - (n-1)k > nx'(0) - (n-1)k' = y'$, a contradiction.

Next, fix k ; for a given n , let $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ be a non-empty network equilibrium with $x_i > 0$ for all $i \in N$; similarly, for given $n' > n$, let $\mathbf{s}' = (\mathbf{x}', \mathbf{g}')$ be a non-empty network equilibrium with $x'_i > 0$ for all $i \in N$. We claim that $y' > y$. Suppose that $y' \leq y$, then $f'(y) = C'(x(0)) \leq f'(y') = C'(x'(0))$; so, $x(0) \leq x'(0)$. But then, $y = nx(0) - (n-1)k \leq nx'(0) - (n-1)k < n'x'(0) - (n'-1)k = y'$, where the last inequality follows because $nx'(0) - (n-1)k$ is increasing in n whenever $x'(0) > k$, which holds because $x'(l) = x'(0) - kl > 0$ for all $l \geq 1$. This concludes the proof.

PROOF OF PROPOSITION 7:

First, suppose that \mathbf{s}^* is pairwise equilibrium and that $x_i^* = 0$ for some $i \in N$. Since \mathbf{s}

is Nash, it follows that $\bar{g}_{ij}^* = 0$ for all $j \in N$, which, in view of Lemma 3.1, it implies that $x_i^* = \hat{y}$. This contradicts our initial hypothesis that $x_i^* = 0$.

Second, let \mathbf{s}^* be such that $\bar{\mathbf{g}}^*$ is a regular network of degree $v = \{1, \dots, n-2\}$ and each player exerts effort $x^* = \hat{y}/[v+1]$. Note that each player accesses \hat{y} and that the payoffs to a player i are $\Pi(\mathbf{s}^*) = f(\hat{y}) - cx^* - vk/2$. If player i deletes some of $t \leq v$ of his links, his best new effort would be $x_i = (t+1)x^*$. Hence, player i 's payoffs are $f(\hat{y}) - (t+1)x^*c - (v-t)k/2$. Since \mathbf{s} is pairwise equilibrium, condition 2 implies that $f(\hat{y}) - (t+1)x^*c - (v-t)k/2 \leq \Pi(\mathbf{s}^*)$, which is equivalent to $k \leq 2cx^*$. Next, note that since $v < n-1$ there are two players, say i and j , with $\bar{g}_{ij}^* = 0$. Suppose i and j deviates by forming a new link between them. Then the new effort of i and the new effort of j must be such that $x'_i + x'_j = x^*$. Suppose, without loss of generality, that $x'_i \leq x'_j$; it is clear that if j wishes to form the new link with i then also i wishes to form the new link with j . So, we can focus on player j . His payoffs, after adding the new link and revising his effort, are $f(\hat{y}) - cx'_j - (v+1)k/2$. From above we know that $x'_j = x^* - x'_i$ and therefore pairwise equilibrium requires that $f(\hat{y}) - cx^* - vk/2 \geq f(\hat{y}) - cx'_j - (v+1)k/2 = f(\hat{y}) - c(x^* - x'_i) - (v+1)k/2$ or $k/2 > cx'_i$. Since $x'_i \in [0, x^*/2]$, as long as $k > cx^*$ the condition is always satisfied. Hence the profile s is pairwise stable whenever $k \in [cx^*, 2cx^*]$. The proof for the case in which the network is complete follows similar arguments and therefore it is omitted. This concludes the proof. \square

Appendix B.

Heterogeneity in indirect transmission model: Suppose that player 1 has the lowest marginal costs of information acquisition, i.e., $c_1 < c_i = c$ for all $i \in N \setminus \{1\}$. Let k_i be the costs that any player has to pay to link with player i . Assume that there is a player j , with $k_j < k = k_i$, for all $i \neq j$. In words, player j is the socially most available player, an individual attribute which may reflect higher communication skills, higher sociability and innate interest in providing public goods. Recall that $\hat{y} = \arg \max_y f(y) - cy$, while $\hat{y}_1 = \arg \max_y f(y) - c_1y$.

PROPOSITION A: *Suppose payoffs are given by (4) and suppose that $c_1 < c = c_i$, for all $i \neq 1$ and there is some player j such that $k_j < k = k_i$ for all $i \neq j$.*

1. Suppose $j = 1$. If $k_1 < f(\hat{y}_1) - f(\hat{y}) + c\hat{y}$ then there is a unique equilibrium, $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$; with $x_1^* = \hat{y}_1$, $x_i^* = 0$ for all $i \neq 1$, \mathbf{g}^* is a periphery-sponsored star and player 1 is the hub.
2. Suppose $j \neq 1$. If $k < f(\hat{y}_1) - f(\hat{y}) + c\hat{y}$ then in every equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$, $x_1^* = \hat{y}_1$, $x_i^* = 0$ for all $i \neq 1$, \mathbf{g}^* is such that there are $l \in \{0, n - 2\}$ players who form a link with j , and all other players lie in the path between player j and player 1. Moreover, there is a unique strict equilibrium with $l = n - 2$.

PROOF OF PROPOSITION A:

The assumption on frictionless information transmission implies that in every equilibrium the network is minimal. Let $z_{C(\mathbf{g})}$ be the aggregate information in component $C(\mathbf{g})$

CLAIM 5: Suppose $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ is a non-empty equilibrium network and let $C(\mathbf{g})$ be a component. If $1 \in C(\mathbf{g})$ then $z_{C(\mathbf{g})} = \hat{y}_1$. If $1 \notin C(\mathbf{g})$ then $z_{C(\mathbf{g})} = \hat{y}$.

PROOF OF CLAIM 5: First, note that in every component $C(\mathbf{g})$, $x_i > 0$ for some i . Second, let $1 \in C(\mathbf{g})$. Suppose $x_1 = 0$, then from Lemma 1 we know that $x_1 + y_1 \geq \hat{y}_1$; but this contradicts Lemma 1 because $\hat{y}_1 > \hat{y}$ and there exists a player $i \in C(\mathbf{g})$ with $x_i > 0$. So, $x_1 > 0$, and again from Lemma 1 it follows that $x_1 + y_1 = \hat{y}_1 = z_{C(\mathbf{g})}$. The second part of the claim can be proved using analogous arguments, the details of which are omitted. This completes the proof of claim 5.

Proof of Part 1. Suppose $j = 1$, $k_1 < f(\hat{y}_1) - f(\hat{y}) + c\hat{y}$ and $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ is equilibrium. We first show that \mathbf{g} is connected. Suppose \mathbf{g} is not connected. First, assume that $1 \in C(\mathbf{g})$ and $C(\mathbf{g})$ is singleton. Claim 5 implies that $z_{C(\mathbf{g})} = \hat{y}_1$. Select a component $\tilde{C}(\mathbf{g})$, and let $i \in \tilde{C}(\mathbf{g})$, $i \neq 1$; claim 5 implies that $z_{\tilde{C}(\mathbf{g})} = \hat{y}$. If $\tilde{C}(\mathbf{g})$ is singleton, $x_i = \hat{y}$ and $\Pi_i(\mathbf{s}) = f(\hat{y}) - c\hat{y}$. If player i forms a link to 1 and chooses zero information, he gets $f(\hat{y}_1) - k_1 > \Pi_i(\mathbf{s})$, where the inequality follows from $k_1 < f(\hat{y}_1) - f(\hat{y}) + c\hat{y}$. This contradicts the hypothesis that \mathbf{s} is equilibrium. Thus, $i, j \in \tilde{C}(\mathbf{g})$ and assume, without loss of generality, that $g_{ij} = 1$. The payoffs to i are $\Pi_i(\mathbf{s}) = f(\hat{y}) - cx_i - k\eta_i(\mathbf{g})$, where $x_i \geq 0$ and $\eta_i(\mathbf{g}) \geq 1$. If i deletes all his links and forms a link to 1, then he gets at least $f(\hat{y}_1) - cx_i - k_1 > f(\hat{y}) - cx_i - k \geq \Pi_i(\mathbf{s})$,

where the first inequality follows from $k_1 < k$ and $\hat{y}_1 < \hat{y}$, while the second inequality follows because $\eta_i(\mathbf{g}) \geq 1$. This contradicts the hypothesis that \mathbf{s} is equilibrium. The case where 1 belongs to a non-singleton component can be ruled out using analogous arguments, the details of which are omitted. Hence, \mathbf{g} is connected.

Since the network \mathbf{g} is minimally connected, claim 5 implies that aggregate information is \hat{y}_1 . If $x_i > 0$, $i \neq 1$, then Lemma 1 implies that $x_i + y_i = \hat{y} < \hat{y}_1$, which is a contradiction. Thus, $x_i = 0$ for all $i \neq 1$ and therefore $x_1 = \hat{y}_1$ and player 1 does not form any links. This implies that if $g_{ij'} = 1$ and $j' \neq 1$, then i must access 1 via j' ; for otherwise player i would not access any information via j' and therefore i would strictly gain by setting $g_{ij'} = 0$. But then the payoff of player i is $\Pi(\mathbf{s}) = f(\hat{y}_1) - k\eta_i(\mathbf{g}) < f(\hat{y}_1) - k_1$, where the last expression is the payoff to i by deleting all his links and sponsoring a link to player 1. This contradicts the hypothesis that \mathbf{s} is equilibrium. Hence, $g_{ij} = 0$ for all $i, j \neq 1$, which implies that \mathbf{g} is a star, the hub is player 1 and each spoke forms one link with the hub. To conclude the proof of Part 1, note that the payoffs to player $i \neq 1$ are $\Pi_i(\mathbf{s}) = f(\hat{y}_1) - k_1$ and, given the assumption on k_1 , $\Pi_i(\mathbf{s}) > f(\hat{y}) - c\hat{y}$, where the last expression is the best payoff that i can earn should he delete his link and become isolated.

Proof of part 2: Suppose $j \neq 1$, $k < f(\hat{y}_1) - f(\hat{y}) + c\hat{y}$ and $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ is equilibrium. Without loss of generality, set $j = 2$. From standard arguments we can establish that \mathbf{g} is connected and from the assumption of frictionless flow we conclude that \mathbf{g} is minimally connected. Claim 5 implies that aggregate information equals \hat{y}_1 . Therefore, $x_i = 0$ for all $i \neq 1$ and $x_1 = \hat{y}_1$. There are three facts that follow. Fact 1: player 1 does not form any links. Fact 2: for all $i \neq 1$, if $g_{ii'} = 1$, then i accesses player 1 via i' (including the possibility that $i' = 1$); for otherwise, i strictly gains by deleting the link to i' . Fact 3: for all $i \neq 1$, with $g_{ii'} = 1$ and $i' \neq 2$, if i accesses l via i' , then $l \neq 2$; for otherwise player i strictly gains by switching from i' to 2. These three facts immediately imply that player 1 is an end-agent. We then have two possibilities. One, $g_{21} = 1$; in this case, fact 2 and 3 imply that \mathbf{g} is a star, player 2 is the hub, and $g_{i2} = 1$ for all $i \neq 1$. Two, $g_{21} = 0$. Consider the path between 2 and 1: $\bar{g}_{2j_1} = \bar{g}_{j_1j_2} = \dots = \bar{g}_{j_d1} = 1$. Fact 2 implies that $g_{2j_1} = g_{j_1j_2} = \dots = g_{j_d1} = 1$; fact 3 implies that players $2, j_1, \dots, j_d$ do not sponsor additional links. Let $i \neq \{2, j_1, \dots, j_d, 1\}$, fact 3 implies

that player i only links with player 2, i.e., $g_{i2} = 1$. This concludes the characterization of equilibria. Finally, it is easy to verify that every non-star equilibrium network involves a path such as $g_{2j_1} = g_{j_1j_2} = \dots = g_{j_d1} = 1$; and player 2 is indifferent between a link with j_1 and 1; so such paths are not sustainable in a strict equilibrium. This concludes the proof of the proposition.

Alternative formulations of model

We analyze three alternatives formulation of the model, involving discrete information acquisition, one-way flow of information in communication and investment in strength of communication links.

The Best Shot Game. We now study a model in which players have a choice between two actions 0 and 1, where we interpret 1 as acquire information and 0 as not acquire information. The main point we wish to bring out is that every equilibrium exhibits the law of the few and that equilibria are efficient. We focus on the case where players only access the information personally acquired by their direct neighbors.

Formally, a player either acquires information at a cost c or he does not acquire any information personally, i.e. $X = \{0, 1\}$. The returns to a player from acquiring information are $f(x_i + y_i) = 1$ if $x_i + y_i \geq 1$, otherwise $f(x_i + y_i) = 0$, where $y_i = \sum_{j \in N_i(\bar{g})} x_j$. We assume that $c < 1$. This specification resembles the best shot game which has been widely studied in economics. The best-shot game is a good metaphor for situations in which there are significant externalities between players' information collection efforts.¹⁸ The following proposition characterizes the equilibria in the best shot game.

PROPOSITION B: *Suppose $X = \{0, 1\}$. If $k < c$ then every equilibrium has a periphery-sponsored star architecture, the hub chooses 1 and every spoke chooses 0. If $k > c$ then there exists a unique equilibrium: every player chooses 1 and no one forms any links.*

PROOF: Suppose $k < c$ and let $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ be an equilibrium. We claim that there exists an $i \in N$ such that $x_i = 1$ and that $x_j = 0, \forall j \neq i$. First, since $k < c$, there must be at least a

¹⁸For a discussion of best-shot games within the context of public good games see, e.g., Jack Hirshleifer 1983 and Glenn W. Harrison and Jack Hirshleifer 1989.

player who acquires information 1. Second, suppose both i and j choose 1. Then, it must be the case that $x_{i'} = 0, \forall i' \in N_i(\bar{\mathbf{g}})$; for if a neighbor of i chooses 1, player i strictly gains by choosing 0. Since $x_{i'} = 0, \forall i' \in N_i(\bar{\mathbf{g}})$, then $g_{il} = 0$ for all l . Hence, player i 's payoffs in equilibrium \mathbf{s} are $1 - c$. If player i chooses 0 and forms a link with j then he obtains $1 - k$. Since $k < c, 1 - k > 1 - c$ and therefore \mathbf{s} cannot be an equilibrium. Next, let $x_i = 1$ and $x_j = 0, \forall j \neq i$. Trivially, $g_{j'i} = 0, \forall j' \in N, j' \neq i$, and, since $k < c$, every player $j \neq i$ has a link with i . This completes the proof for the case $k < c$. The proof for the case $k > c$ is trivial and therefore omitted.

We finally note that in best shot game every equilibrium is efficient. This is in sharp contrast to the case in which information is a continuous variable.

PROPOSITION C: *Suppose $X = \{0, 1\}$. If $k < c$, then the socially optimal outcome is a star network, the hub chooses 1 and every spoke chooses 0. If $k > c$, then in the socially optimal outcome every player chooses 1 and no one forms links.*

PROOF: Suppose $k < c$ and suppose that $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ is efficient. It is easy to see that the only links in \mathbf{g} are between pair of players (i, j) with $x_i \neq x_j$. Also, if player i chooses 0 then player i has only one link with a player choosing 1. Indeed, if player i had two distinct links with two players choosing 1, then welfare can be made strictly higher by deleting one of the link. Hence, the total number of links are $(n - m)$, where m is the number of players choosing 1, and each player gets returns of 1. Then the social welfare is $n - mc - (n - m)k$. If $k > c$, this expression decreases with m and therefore $m = 1$, which implies the result. Suppose now that $k > c$. The above arguments show that if there are $m < n$ players choosing 1, and \mathbf{s} is efficient then the social welfare is $n - mc - (n - m)k$, but then welfare can be increased by setting $m = n$, which implies the result. This concludes the proof.

One-way information transmission: We have assumed so far that if i forms a link with j then i accesses the information of j and *vice-versa*. An alternative model is one where a link $g_{ij} = 1$ allows only player i to access the information acquired by player j , i.e., one-way information flow.¹⁹ This is a reasonable model for links on the web, or for citations. Let

¹⁹For a general analysis of a game of pure network formation with one-way information flow, see Bala and

$N_i(\mathbf{g}) = \{j \in N : g_{ij} = 1\}$; the payoffs to player i in a strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$ are then given by

$$\Pi_i(\mathbf{s}) = f \left(x_i + \sum_{j \in N_i(\mathbf{g})} x_j \right) - cx_i - \eta_i(\mathbf{g})k. \quad (9)$$

We observe that some of our main results carry over to the one-way information flow formulation. To illustrate this, consider the equilibria described in Proposition 2. Recall that at equilibrium $\mathbf{s} = (\mathbf{x}, \mathbf{g})$, aggregate information equals \hat{y} , \mathbf{g} is a core-periphery network, hubs acquire information and spokes do not collect any information. Let \mathbf{g}' be a network in which if $g_{ij} = 1$ then $g'_{ij} = 1$ and if $g_{ij} = 0$ and $x_i, x_j > 0$, then $g'_{ij} = 1$. It is easy to see that $\mathbf{s}' = (\mathbf{x}, \mathbf{g}')$ is an equilibrium under the one-way flow formulation (for appropriately chosen k). In particular, in the one way flow formulation, the periphery-sponsored star, with the hub collecting all the information is an equilibrium whenever $k < c\hat{y}$. An analogous argument can be used to extend the results of Proposition 3. Indeed, it is easy to see that a periphery-sponsored star, where the player with the lowest cost of information acquisition is the hub, he personally acquires information while all other players do not acquire information personally is an equilibrium under the assumption of one-way information flow as well.

Investments in link strength: We consider an alternative model where the investment in linking is also a continuous choice. A strategy of player i is then an investment in information acquisition $x_i \in X$ and an investment in linking summarized by a vector $\mathbf{v}_i = \{v_{i1}, \dots, v_{in}\}$, where $v_{ij} \in [0, 1]$ and $v_{ii} = 0$. A profile \mathbf{v} generates a (undirected) symmetric value network \mathbf{g} , where the strength of an arbitrary link $g_{ij} = g_{ji} = \max[v_{ij} + v_{ji}, 1]$.²⁰ The payoffs to player i in a strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{v})$ are then given by:

$$\Pi_i(\mathbf{s}) = f \left(x_i + \sum_{j \in N} g_{ij}x_j \right) - cx_i - k \sum_{j \in N} v_{ij}. \quad (10)$$

So, the information that i accesses from j is the information that j has acquired personally weighted by the strength of the link between i and j . We can adapt the definition of a core-periphery network as follows. In a core-periphery network there are two groups of players

Goyal (2000) and Andrea Galeotti (2006).

²⁰For a general analysis of a game of pure network formation when link strength is endogenous, see Francis Bloch and Bhaskar Dutta 2008.

$\hat{N}_1(\mathbf{g})$ and $\hat{N}_2(\mathbf{g})$ with the feature that: (1.) for all $i, j \in \hat{N}_1(\mathbf{g})$, $g_{ij} = 1$, (2.) for all $i \in \hat{N}_1(\mathbf{g})$, $v_{ij} = 1$ for all $j \in \hat{N}_2(\mathbf{g})$ and (3.) all other links have zero strength. The following proposition shows that the results presented in our basic model are robust to this extension.

PROPOSITION D: *Suppose payoffs are given by (10). If $k > c\hat{y}$ then there exists a unique equilibrium in which every player acquires information \hat{y} and no one invests in any links. If $k < c\hat{y}$ then in every strict equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{v}^*)$: (1) $\sum_{i \in N} x_i^* = \hat{y}$; (2) every equilibrium network has a core-periphery architecture, hubs acquire information personally and spokes acquire no information personally; (3) for given c and k , with $k < c\hat{y}$ the ratio $|I(\mathbf{s}^*)|/n \rightarrow 0$ as $n \rightarrow \infty$.*

The new insight of the analysis is that if two players who acquire information have a link then it must be a full strength link. To see why this is true, suppose player i acquires information personally and also invests effort $v_{ij} > 0$ in forming a link with another player j but that $v_{ij} + v_{ji} < 1$. This is strictly profitable only if the costs of connecting are lower than the costs of personal information acquisition; in other words, if $k < cx_j$. But then it is strictly profitable to lower personal information acquisition and increase the strength of the link with player j . Thus it is never optimal to have a link of less than full strength. Once we have proved that links between positive effort players are full strength, we can exploit the endogeneity of links to show, as in Propositions 1-2, that aggregate information acquisition in a strict equilibrium must equal \hat{y} ; the rest of the result then follows.

PROOF OF PROPOSITION:

First, the proof for the case in which $k > c\hat{y}$ is straightforward and therefore it is omitted. Second, suppose that $k < c\hat{y}$. The following two claims are useful to prove the second part of the proposition.

CLAIM 6: *Assume that $k < c\hat{y}$. If \mathbf{s}^* is equilibrium then $v_{ij}^* + v_{ji}^* \leq 1$. Moreover, in every strict equilibrium \mathbf{s}^* if $i, j \in I(\mathbf{s}^*)$ and $v_{ij}^* + v_{ji}^* > 0$ then $v_{ij}^* + v_{ji}^* = 1$, i.e. the link between i and j has full strength.*

PROOF OF CLAIM 6: The first part of the claim is straightforward. To see the second part, assume that $i, j \in I(\mathbf{s}^*)$ and $v_{ij}^* + v_{ji}^* \in (0, 1)$; suppose, without loss of generality, that $v_{ij}^* > 0$. Since $x_i^* > 0$, Lemma 1 implies that $x_i^* + y_i^* = \hat{y}$. So the profits of i in \mathbf{s}^* are $\Pi_i(\mathbf{s}^*) = f(\hat{y}) - cx_i^* - \sum_{j' \neq j} v_{ij'}^* k - v_{ij}^* k$. Consider the deviation in which player i chooses $v'_{ij} = 0$ and an effort $x'_i = x_i^* + v_{ij}^* x_j^*$. Note that, under this deviation we have that $x'_i + y'_i = x_i^* + v_{ij}^* x_j^* + y_i^* - v_{ij}^* x_j^* = \hat{y}$ and player i 's payoffs are $\Pi_i(\mathbf{s}'_i, \mathbf{s}^*_{-i}) = f(\hat{y}) - c[x_i^* + v_{ij}^* x_j^*] - \sum_{j' \neq j} v_{ij'}^* k$. Since \mathbf{s}^* is a strict equilibrium $\Pi_i(\mathbf{s}^*) > \Pi_i(\mathbf{s}'_i, \mathbf{s}^*_{-i})$ and so $k < cx_j^*$. However, given that $k < cx_j^*$, player i strictly gains by increasing the strength of the link with j by some small and positive ϵ (and so i increases his costs of linking by $k\epsilon$) and by reducing his own information acquisition by ϵx_j^* (and so i decreases his costs of information acquisition by $c\epsilon x_j^*$). This means that \mathbf{s}^* is not an equilibrium, a contradiction which completes the proof of Claim 6.

CLAIM 7: Assume $k < c\hat{y}$ and suppose that \mathbf{s}^* is a strict equilibrium. If $\sum_{i \in N} x_i^* > \hat{y}$ then $x_i^* = x$ for all $i \in I(\mathbf{s})$.

PROOF OF CLAIM 7: First note that since $\sum_{i \in N} x_i^* > \hat{y}$, Lemma 1 implies that $|I(\mathbf{s}^*)| > 1$. Second, in contradiction with the claim suppose that $x_j^* \neq x_{j'}^*$ for some $j, j' \in I(\mathbf{s}^*)$. In particular, select player $i \in I(\mathbf{s}^*)$ such that $x_i^* = \underline{x} = \min_{j' \in I(\mathbf{s}^*)} \{x_{j'}^*\}$ and select player $j \in I(\mathbf{s}^*)$ such that $x_j^* = \bar{x} = \max_{j' \in I(\mathbf{s}^*)} \{x_{j'}^*\}$; hence, $\underline{x} < \bar{x}$. Lemma 1 implies that $\underline{x} + y_i^* = \bar{x} + y_j^* = \hat{y}$. Since $\underline{x} < \bar{x}$, it follows that $y_i^* > y_j^*$. We need to analyze two cases.

Case 1. There exists a player $l \in I(\mathbf{s}^*)$ such that $g_{lj}^* = 0$ and $g_{li}^* = 1$; note that $\underline{x} \leq x_l^* \leq \bar{x}$. If $v_{li}^* > 0$, then player l strictly gains by setting $v'_{li} = 0$ and choosing $v'_{lj} = v_{li}^*$. In fact, under this deviation player l faces the same costs of linking and the same costs of effort, but he accesses an extra effort $v_{li}^*(\bar{x} - \underline{x}) > 0$. So, suppose $v_{li}^* = 0$; in this case claim 6 implies that $v_{il}^* = 1$. Since \mathbf{s}^* is a strict equilibrium and $v_{il}^* = 1$ we have that $k < cx_l^*$. Since $g_{jl}^* = 0$, $x_l^* \leq \bar{x}$ and $k < cx_l^*$, player j strictly gains by setting $x'_j = \bar{x} - x_l^*$ and $v'_{jl} = 1$.

Case 2. Suppose that for all $l \in I(\mathbf{s}^*)$ if $g_{li}^* = 1$ then $g_{lj}^* = 1$. In this case, since $y_i^* > y_j^*$, for all $l \in I(\mathbf{s}^*)$ such that $g_{li}^* = 1$ it must be that $g_{lj}^* = 1$. That is, i and j share the same positive information neighbors. Furthermore, since $\underline{x} + y_i^* = \bar{x} + y_j^* = \hat{y}$ and $\underline{x} < \bar{x}$, it must be the case that $g_{ij}^* = 1$. These observations together with our initial hypothesis that aggregate

information exceeds \hat{y} imply that there exists a player $l \in I(\mathbf{s}^*)$ such that $g_{il}^* = g_{jl}^* = 0$. In a strict equilibrium, player l must set $v_{ll'}^* = 0$ for all $l' \in I(\mathbf{s}^*)$; in fact, if $v_{ll'}^* > 0$ for some $l' \in I(\mathbf{s}^*) \setminus \{j\}$ then player l will weakly gain by setting $v_{ll'}' = 0$ and $v_{lj}' = v_{ll'}^*$. Thus, for each $g_{ll} = 1$, $v_{ll} = 1$; this implies that in a strict equilibrium $k < cx_l^*$. But then player j strictly gains by setting $v_{jl}^* = 1$ and choosing effort $x_j' = \bar{x} - x_l^* \geq 0$, where the inequality follows because by construction $\bar{x} \geq x_l^*$.

Hence, in a strict equilibrium where aggregate information is strictly higher than \hat{y} , all positive effort players must exert the same effort. This concludes the proof of Claim 7.

Final Step. An immediate consequence of Claim 6 and Claim 7 is that in a strict equilibrium where aggregate information exceeds \hat{y} the network among positive effort players must be regular (with each link having strength 1). We now show that any such strategy profile cannot constitute a strict equilibrium. Since aggregate information exceeds \hat{y} and $k < c\hat{y}$, there exists a player $i \in I(\mathbf{s}^*)$ such that $g_{ij}^* = 0$ for some $j \in I(\mathbf{s}^*)$ and $g_{ij'}^* = 1$ for some $j' \in I(\mathbf{s}^*)$ and $v_{ij'}^* > 0$. Since $x_j^* = x_{j'}^* = x_i^*$, it is easy to check that player i is indifferent between playing his current strategy and deviating by setting $v_{ij'}' = 0$ and $v_{ij}' = v_{ij'}^*$. Hence, in every strict equilibrium aggregate information cannot exceed \hat{y} . This fact and Lemma 1 imply that in equilibrium aggregate information must be exactly \hat{y} . Recall that Claim 6 says that links among positive effort players have full strength. So, since aggregate effort equals \hat{y} positive effort players must form a clique. It is now clear that if $i \notin I(\mathbf{s}^*)$ then $v_{ij}^* = 1$ for all $j \in I(\mathbf{s}^*)$; it is also straightforward to see that for all $i, j \notin I(\mathbf{s}^*)$, $v_{ij}^* = 0$. So the network has a core-periphery architecture, hubs acquire information personally, while spokes acquire no information personally. Part (3) of the proposition now follows by replicating the arguments used in the proof of Proposition 2. This concludes the proof of Proposition 8.

REFERENCES

- Adar, Eytan, and Bernardo A. Huberman.** 2000. "Free Riding on Gnutella." *First-Monday*, 5(10).
- Bala, Venkatesh, and Sanjeev Goyal.** 2000. "A non-cooperative model of network formation." *Econometrica*, 68(5): 1181-1229.
- Beck, Paul, Russel J. Dalton, Steven Green, and Robert Huckfeldt.** 2002. "The Social Calculus of Voting: Interpersonal, Media and Organizational Influences on Presidential Choices." *American Political Science Review*, 96(1): 57-73.
- Bergstrom, Theodore, Lawrence Blume, and Hal R. Varian.** 1986. "On the private provision of public goods." *Journal of Public Economics*, 29(1): 25-49.
- Bloch, Francis, and Bhaskar Dutta.** Forthcoming. "Communication networks with endogenous link strength." *Games and Economic Behavior*.
- Bloch, Francis, and Matthew O. Jackson.** 2007. "The formation of networks with transfers among players." *Journal of Economic Theory*, 133: 83-110.
- Bramoullé, Yann, Dunia Lopez-Pintado, Sanjeev Goyal, and Fernando Vega-Redondo.** 2004. "Social interaction in anti-coordination games." *International Journal of Game Theory*, 33: 1-20.
- Bramoullé, Yann, and Rachel Kranton.** 2007. "Public goods in networks." *Journal of Economic Theory*, 135(1): 478-494.
- Cabrales, Antonio, Antoni Calvó-Armengol, and Yves Zenou.** 2007. "Social interactions and spillovers: incentives, segregation and topology." www.eco.uc3m.es/acabrales/research/WP.
- Cabrales, Antonio, and Piero Gottardi.** 2007. "Markets for Information: Of Inefficient Firewalls and Efficient Monopolies." <http://www.eco.uc3m.es/acabrales/research/WP>.
- Calvó-Armengol, Antoni, and Yves Zenou.** 2004. "Social Networks and Crime Decisions: The Role of Social Structure in Facilitating Delinquent Behavior." *International Economic Review*, 45(3), 939-958.
- Cho, Myeonghwan.** 2007. "Endogenous formation of networks for local public goods." <http://my.win.psu.edu/muc/52/research>.
- Cross, Rob, Andrew Parker, Laurence Prusak, and Stephen P. Borgatti.** 2001.

- “Knowing What We Know: supporting knowledge creation and sharing in social networks.” *Organizational Dynamics*, 30(2):110-120.
- Cross, Rob, and Laurence Prusak.** 2002. “The People Who Make Organizations Stop-or Go.” *Harvard Business Review*, 80(6): 104-112.
- Cross, Rob, and Andrew Parker.** 2004. *The Hidden Power of Social Networks: Understanding How Work Really Gets Done In Organizations*. Harvard Business School Press.
- Feick, Lawrence F., and Linda L. Price.** 1987. “The Market Maven: A Diffuser of Marketplace Information.” *Journal of Marketing*, 51(1): 83-97.
- Galeotti, Andrea.** 2006. “One-way flow networks: the role of heterogeneity.” *Economic Theory*, 29: 163-179.
- Galeotti, Andrea, Sanjeev Goyal, and Jurjen Kamphorst.** 2006. “Network formation with heterogenous players.” *Games and Economic Behavior*, 54: 353-372.
- Geissler, Gary L., and Steve W. Edison.** 2005. “Market Mavens’ Attitudes Towards General Technology: Implications for Marketing Communications.” *Journal of Marketing Communications*, 11(2): 73-94.
- Gladwell, Malcolm.** 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. Little Brown.
- Goyal, Sanjeev.** 1993. “Sustainable communication networks”, *Tinbergen Institute Discussion Paper*, TI 93-250.
- Goyal, Sanjeev.** 2007. *Connections: an introduction to the economics of networks*. Princeton University Press.
- Goyal, Sanjeev, and Fernando Vega-Redondo.** 2005. “Network formation and social coordination.” *Games and Economic Behavior*, 50, 178-207.
- Harrison, Glenn W., and Jack Hirshleifer.** 1989. “An Experimental Evaluation of Weakest Link/Best Shot Models of Public Goods.” *Journal of Political Economy*, 97(1): 201-225.
- Hirshleifer, Jack.** 1983. “From Weakest-Link to Best-Shot: The Voluntary Provision of Public Goods.” *Public Choice*, 41: 371-386.
- Hojman, Daniel, and Adam Szeidl.** 2008. “Core and periphery in networks.” *Journal of Economic Theory*, 139: 295-309.

- Huckfeldt, Robert, Paul E. Johnson, and John Sprague.** 2004. *Political Disagreement: The survival of diverse opinions within Communication networks*. New York: Cambridge University Press.
- Jackson, Matthew O.** 2008. *Social and economic networks*. Princeton University Press, New Jersey.
- Jackson, Matthew O., and Asher Wolinsky.** 1996. "A strategic model of social and economic networks." *Journal of Economic Theory*, 71: 44-74.
- Jackson, Matthew O., and Alison Watts.** 2002. "On the Formation of Interaction Networks in Social Coordination Games." *Games and Economic Behavior*, 41(2): 265-291.
- Jackson, Matthew O., and Brian W. Rogers.** 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks." *American Economic Review*, 97(3): 890-915.
- Katz, Elihu, and Paul F. Lazarsfeld.** 1955. *Personal Influence*. New York: The Free Press.
- Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet.** 1948. *The People's Choice*. New York: Columbia University Press.
- Wiedmann, Klaus P., Gianfranco Walsh, and Vincent W. Mitchell.** 2001. "The Mannmaven: an agent for diffusing market information." *Journal of Marketing Communications*, 7: 185-212.
- Williams, Terrell G., and Mark E. Slama.** 1995. "Market mavens' purchase decision evaluative criteria: implications for brand and store promotion efforts." *Journal of Consumer Marketing*, 12(3): 4-21.
- Zhang, Jun, Mark S. Ackerman, and Lada Adamic.** 2007. "Expertise Networks in Online Communities: Structures and Algorithms" *WWW2007*, Banff, Canada.