Modeling self-organization of communication and topology in social networks

M. Rosvall[1,2,*] and K. Sneppen[2,†]

[1]*Department of Theoretical Physics, Umeå University, 901 87 Umeå, Sweden*
[2]*Niels Bohr Institute, Blegdamsvej 17, Dk 2100 Copenhagen, Denmark*
(Received 1 December 2005; published 17 July 2006)

This paper introduces a model of self-organization between communication and topology in social networks, with a feedback between different communication habits and the topology. To study this feedback, we let agents communicate to build a perception of a network and use this information to create strategic links. We observe a narrow distribution of links when the communication is low and a system with a broad distribution of links when the communication is high. We also analyze the outcome of chatting, cheating, and lying, as strategies to get better access to information in the network. Chatting, although only adopted by a few agents, gives a global gain in the system. Contrary, a global loss is inevitable in a system with too many liars.

　　　　PACS number(s): 89.75.Fb, 89.65.Lm, 89.70.+c

I. INTRODUCTION

Who communicates with whom and the social structure of a society are strongly entangled. The social network reflects the access to information that different parts of the system experience, and social mobility may be seen as a quest for better information access. A reliable global perception of the network, often achieved by informal communication with acquaintances [1–8], makes the social mobility meaningful [9,10]. The small talk consists in its simplest form of identifying who to get the information from, and whom to transfer it to [11–16]. To understand the feedback between different communication habits and the topology, we in this paper introduce an agent-based model that self-organizes the social network. That is, we allow agents to create new links to get easier access to some parts of the system, based on interest and the information they obtained through communication with already established acquaintances [17–19].

After defining the model in the next section, we show that organized structures, that can make use of the small-world properties of the network [20–22], emerge when the communication is sufficiently high. This is followed by an investigation of consequences of manipulating information. What are the gains or costs when the agents adopt individual strategies to get better access to the system on, respectively, the local and global level? We investigate consequences of chatting, cheating, and lying, and find that communication, although not equally distributed, is a benefit for everyone. Finally we explore a few variants of the model and, for example, show how separation of interests naturally leads to modular networks in the model.

II. MODEL

Let us now define the model in detail, formulated in the two independent events.

1. *Communication:* Select a random link and let the two agents that it connects communicate about a random third agent [23]. The two agents also update their information about each other.

2. *Rewiring:* Select a random agent and let it use the local information to ask an acquaintance about whom to establish a link to, to shorten its distance to a randomly chosen other agent. Subsequently a random agent loses one of its links.

The communication event is typically repeated of the order of the number of links in the system for each rewiring event. Figures 1 and 2 illustrate the two elements in the model. The basic variables in the network model are nodes represented by N agents and L links that correspond to the available communication channels in the system. We let the agents communicate and in that way build their own perception of where they are relative to other agents in the network. Each agent, in Fig. 1 exemplified by agent **A**, has a list of previously obtained information with entries for each agent $i = \mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$. For each entry i, the agent has a pointer to the agent that provided the most recent information about i. This pointer is updated if someone else comes with newer information about i [22]. Therefore we also keep the age of all obtained information in **A**'s memory (see clocks in Fig. 1). The age of an agent's information about itself is always 0. The age of any other information increases proportionately to the number of ongoing communication events in the system. When two agents communicate about a third agent the agent with the older information disregards this and adopts the viewpoint of the agent with the newer information by copying the age and changing the pointer. In Fig. 1 agent **A** communicates with **B** about agent **H**, and adopts the viewpoint of **B** because **B**'s information about **H** is newer. **A** sets its clock for **H** to the same time as **B**, and changes its pointer for **H** to **B**. The age of the information serves as a qualifier that allows two communicating agents to estimate which of them has the most reliable information.

Figure 2 describes the second main feature of the model, the social mobility. We implement the social constraints of who can connect to whom by only allowing new links from an agent to acquaintances of its acquaintances [24]. Independent of the communication event, a randomly chosen agent, here **A**, is interested in shortening its distance to another randomly chosen agent in the system, here **H**. **A** therefore asks **B**, the agent that provided **A** with the newest information about **H**, about where the information came from. **B**

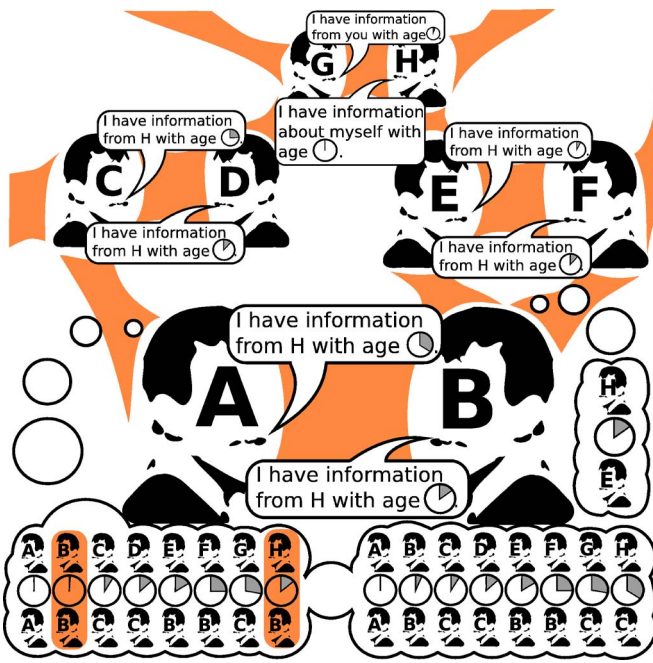*Electronic address: rosvall@tp.umu.se

†URL: http//cmol.nbi.dk

FIG. 1. (Color online) *Communication* as modeled in this paper: Agents communicate with their neighbors in the network about a third target agent and estimate the quality of the information by its age. The agent with the oldest information adopts the viewpoint of the agent with the newest information. Here, agent **A** learns that **B** has newer information about **H**, disregards its old information, and changes its pointer towards **H** to **B**. The information in the bottom bubbles are **A**'s knowledge about the network based on communication with its neighbors before and after the communication event with **B**: For each agent (top row) the time of the most recent information is stored (middle row) together with the acquaintance that provided the information (bottom row).

answers **E** and **A** builds a link to **E** (if there is no link between **A** and **B**, **A** builds a link to **B** and stops after that). The creation of new links is balanced by random removal of links. This is illustrated in Fig. 2, where **C**, chosen randomly, loses its connection to **D**.

In the model, we thus have an interplay between the communication backbone network and the agents' perception of this network. Moreover, this perception can be interpreted as a network where the agents are the nodes and the pointers of all agents are the links. The links are of two types, updated and outdated. The updated links are the pointers that point in the direction of the real links of the communication network, and the outdated links are the pointers that do not point in the direction of a real link. In Fig. 3 we illustrate the concept of a communication backbone and the perception network at low and high communication in a small network with 25 nodes and 38 links. For relatively few communication events per rewiring (much less than the number of links in the system), the communication and perception network diverge since the rewirings that the agents perform based on mainly outdated pointers have little to do with the real topology of the network. As a consequence, any rewiring of the network will be random and the network's overall topology disorganizes into a structure with a narrow degree distribution [25] (see the two networks to the left in Fig. 3). In contrast, a high

communication implies that new links are introduced as a direct function of the present topology (all pointers are updated). They are typically directed towards highly connected nodes since they provide new information. With a tendency of building new links towards the majority of the system, a reliable perception opens for positive feedback and self-organization into a network with broad degree distribution (see the two networks to the right in Fig. 3).
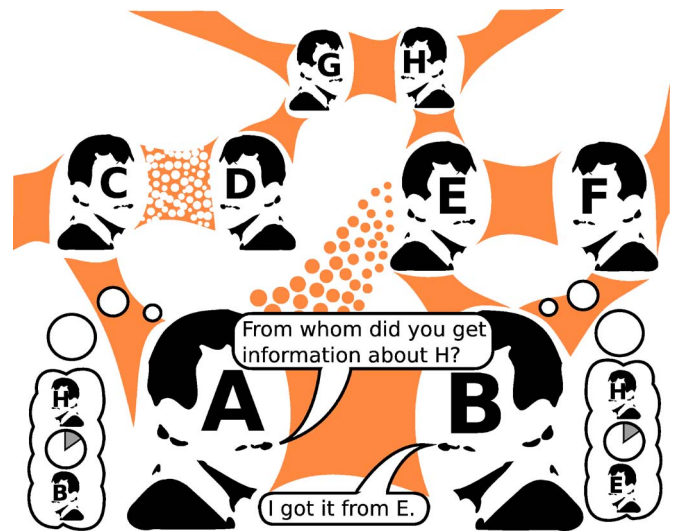

FIG. 2. (Color online) *Rewiring* as modeled in this paper: Agents create new contacts based on their available information (see Fig. 1). In this example **A** uses **B**, the agent that **A** got the newest information about **H** from, to get a better position relative to **H**. **A** then creates a link to **E**, the agent that **B** got its information about **H** from. Thereafter a random agent loses a random link, in this example the connection between **C** and **D**.
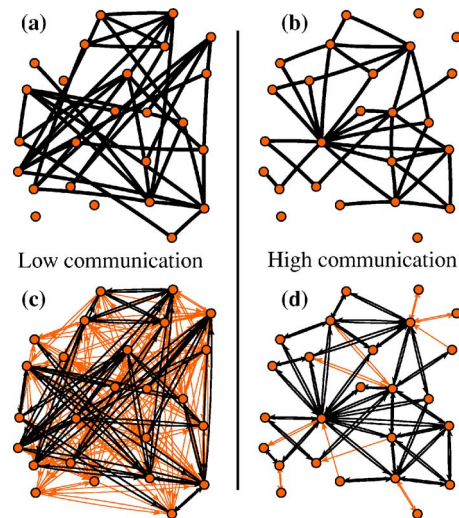

FIG. 3. (Color online) Illustration of the two types of social bonds in the network at two different levels of communication. (a) and (b) show the communication backbone over which agents communicate. (c) and (d) show the perception network associated to the agents' directions to other agents in the network. The pointers are colored black when they are updated and coincide with active connections. In the network with high communication (right panel), almost all pointers overlap with the communication backbone.
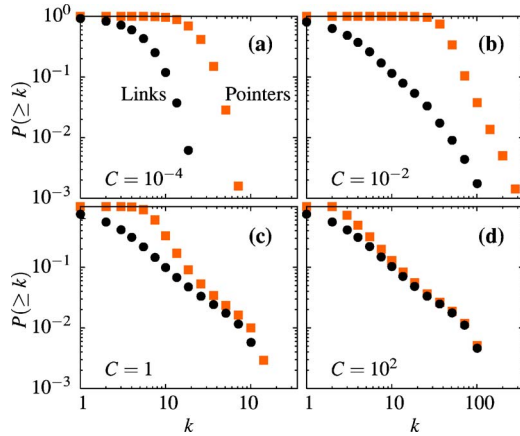
FIG. 4. (Color online) Illustration of the feedback of communication on the topology of both communication backbone and perception network at four different levels of communication C. $C=1$ corresponds to on average one communication event per link and rewiring event. Network size is $N=1000$ agents connected by $L=2500$ links in the communication backbone.

III. RESULTS

To quantify the interplay between the self-organization of the network topology and the overall communication level we show in Figs. 4 and 5 degree distributions for simulations of a system with $N=1000$ agents, $L=2500$ links. We started with random Erdős-Rényi networks [25] (the results are independent of initial conditions) and let the system evolve at different communication levels C. $C \cdot L$ is the number of
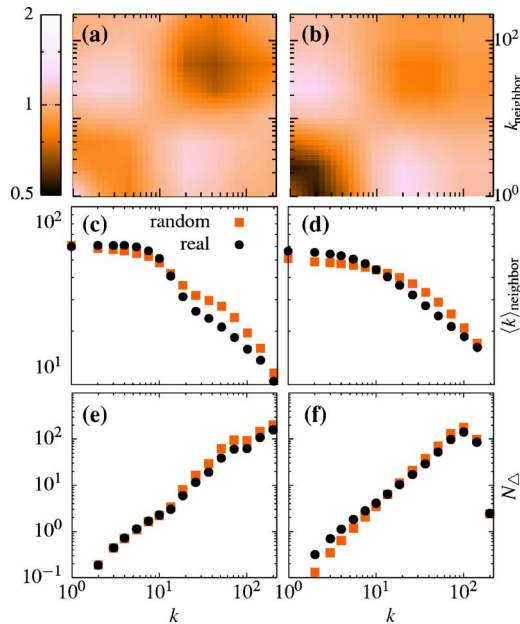

FIG. 5. (Color online) The topology of the generated networks at two different levels of communication C. $C=0.01$ in left panel and $C=1$ in the right panel. The first row shows the correlation profile, the second row the average neighbor degree, as a function of degree, and the third row the clustering measured as the number of triangles. All measures are compared with randomized counterparts of the networks with unchanged degree sequence.

communication events per rewiring event in the network, and the degree k of a node is its number of links. We have also simulated networks with a different number of links and found similar results with a tendency towards more pronounced nonrandom features with fewer links. Figure 4 shows the degree distribution when we varied the number of communication events per rewiring and link between $C=10^{-4}$ and $C=100$. At low communication level, $C<1$, the perception network has many more links than the backbone network. As C approaches $C \sim 1$ the perception network prunes its links whereas the backbone network develops nodes with high degrees [the distribution can be approximated by a power law $P(k) \propto k^{-2.2}$]. At even higher values of C the two networks converge toward the same broad degree distribution.

Beyond the degree distribution, we show in Fig. 5 the correlation profile (top), the average neighbor degree (middle), and the number of triangles (bottom) as a function of degree for low (left) and high (right) communication. In all cases we compare with a randomized network where the degree sequence is identical to the model generated, but all other features are reshuffled [26]. We chose $C=10^{-2}$ as the low and $C=1$ as the high communication level. The overrepresentation of links between nodes of high and low degree gives extended community structures: Triangles are overrepresented around low-degree nodes and under-represented around high-degree nodes [20].

All the presented results until now are based on agents that are all the same. However, at any time their social position will be different because their sequence of communication and rewirings are strongly influenced by the history of the system. The presented model describes a social game where the aim is to be central, and a winner is an agent with many connections that provide short and reliable communication to other agents. The fact that we observe agents with a wide range of degrees reflects the diversity of the possible outcomes of the game, and raises the questions about whether there are some particular strategies with which agents can improve their standing in the network. Can acting like a winner make you more likely to become a winner?

A highly connected agent became highly connected because it attracted new links by providing new information about other agents, and to be able to provide new information is essential to win the game. Accordingly we investigate a number of individual strategies, associated to different local costs, where agents attempt to convince other agents about their attractiveness as an acquaintance.

1. *Chatting* represents an increased communication rate. We let the *chatters* communicate twice as much as other agents by increasing the probability that their links are chosen for a communication event by a factor of 2. Note that this also affects their acquaintances because they share links with the chatters.

2. *Cheating* represents a decreased clock-speed, and therefore a slower aging of the information. We let the *cheaters* use clocks that run at half the speed of the other agents' clocks. In practice they cheat by pretending that they have newer information than they really have. Cheating might be either deterministic (every time unit is of half-length) or stochastic (a time unit is counted with probability 1/2).
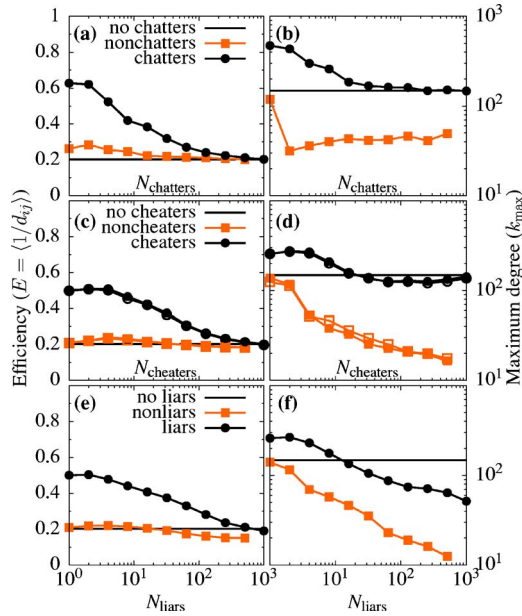
FIG. 6. (Color online) Topological consequences of introducing strategic agents, quantified through efficiency (defined as $E = \langle 1/d_{ij} \rangle$) respectively maximum degree of both the strategic agents (circles) and the remaining agents (squares). (a) and (b) Effect of having $N_{chatters}$ agents which communicate twice as much as the remaining $N-N_{chatters}$ agents in network. (c) and (d) Effect of $N_{cheaters}$ agents that cheat by running their internal clock at half the speed of the other $N-N_{cheaters}$ agents' clocks. (e) and (f) Effect of a more brutal strategy where N_{liars} agents always pretend that their information about all other agents is very new (of the order of what the remaining $N-N_{liars}$ agents have for their nearest or next nearest neighbors). The communication rate is $C=1$ in a system with $N=1000$ nodes and $L=2500$ links. In (c) and (d) open squares and circles correspond to a simulation with stochastic cheaters.

3. *Lying* represents a pure lie about the age of the information in a communication event. Instead of updating the clock, the *liars* replace the time by a random number. Here we choose a random number between 1 and 100 that represents the typical age of information about an agent within the second nearest-neighbor radius in a system with 1000 agents.

In all three strategies the information is manipulated to gain a local advantage to reach the global goal of being central. The chatters work hard to continuously get new and updated information. The cheaters instead pretend that they have newer information than they really have by referring to half the real age of their information upon request. Contrary, the liars do not care about keeping track of the age of their information, but pretend that they are always updated. The local cost of playing the different strategies—high for communication and low for lying—has different effects on the global scale. This is what we examine in Figs. 6 and 7. Figure 6 shows the topological consequences on the communication backbone and Fig. 7 the effect on the perception network, as we vary the number of strategic agents between 1 and the system size at communication level $C=1$. We investigate the communication backbone to understand how the different strategies affect the possibilities to have better access to the system. The perception network instead reflects
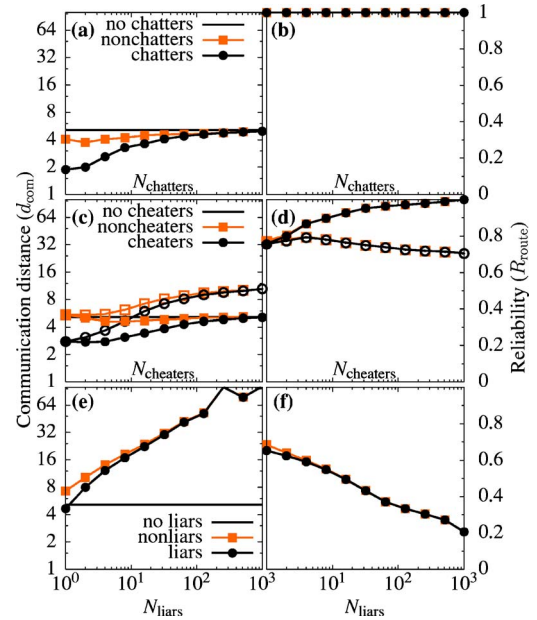

FIG. 7. (Color online) Perception consequences of introducing strategic agents, quantified through communication distance, d_{com}, and reliability of routed messages, R_{route}. (a) and (b), (c) and (d), and (e) and (f) In a society the information horizon is set by each individual's social contacts, which in turn is a part of the global network of human communication and corresponds to the same strategies as in Fig. 6. In (c) and (d) open squares and circles correspond to a simulation with stochastic cheaters.

the actual gain that the agents get from their different strategies. For example, a more central position in the communication network may not always result in shorter communication pathways—which is what really counts—if the communication is inefficient. The right panel in Fig. 6 shows how the maximum degree of, respectively, the strategic agents (circles) and nonstrategic agents (squares) changes with $N_{strategic}$. We use the maximum degree to quantify how attractive it is to connect to the agents. When less than about ten agents adopt any of the three strategies they gain in terms of degree. However, as the number of liars increase, the overall network topology degenerates and it becomes impossible to sustain hubs. Also the liars become losers. A more global examination of the effect of the various strategies is shown in the left panel of Fig. 6. The efficiency $E = \langle 1/d_{ij} \rangle$ is the average value of the reciprocal distance [27] of, respectively, the strategic agents and the nonstrategic agents. We use this measure to estimate the centrality of the agents in the network. With the reciprocal distance, unlike the average distance, it is possible to handle the case with temporarily disconnected agents in the network. In terms of the efficiency, all strategies are successful, and in addition they also seem to favor the other agents by providing short paths.

That the strategic agents become central in the communication backbone network does not directly imply that they can use their centrality. The use of various strategies may influence the reliability of information that the agents have about the system and thereby make long-distance communication more difficult. In Fig. 7 we examine the ability of agents to communicate across the system. d_{com} in the left

panel is the average number of agents that participate in communicating a message from an agent to another agent. It is thereby the perception-network analog of the average distance in the communication network. For chatters, we again see that everybody gains. For cheaters, on the other hand, everybody gains if the cheating is deterministic, but already a few agents with stochastic cheating (faded) make communication across the system less efficient. The seemingly successful strategy of lying from Fig. 6 completely destroys the communication abilities [Fig. 7(e)]. One single liar makes some benefit of its strategy, but two liars are enough to not only destroy for the nonliars, but also for the liars themselves.

To emphasize this result, we show in the right panel of Fig. 7 the reliability, R_{route}, of the perception network. This is the ultimate test to investigate whether the agents' strategies destroy the possibility to send messages across the network. To calculate R_{route}, we send messages between any pairs of nodes and let the intermediate agents route the messages with their pointers. A message fails when it reaches an agent for the second time and the path forms a loop. R_{route} is the fraction of messages that reach the target. The chatters are able to keep perfect reliability, but the cheaters and especially the liars destroy it. When there are 1000 deterministic cheaters the reliability is again 100% [see Fig. 7(d)]. This is because it only corresponds to a rescaling of time when all agents are deterministic cheaters. The liars, examined in Fig. 7(f), systematically destroy the signaling capacity of the network.

The presented model is the simplest in a family of models based on an interplay between communication and dynamical changes of topology. We have investigated a range of variations, including versions where each agent has a biased interest in other agents. For example, we let an agent's target of interest be chosen inversely proportional to the age of the information about the target [28]. Thereby interests are focused around the neighborhood and we observe an increase in the number of triangles in the system. In another variant, we divided the agents into several interest groups. By increasing the probability to communicate and move inside the interest group, the network develops a modular topology.

IV. DISCUSSION

In this work we have introduced a model framework that allows us to investigate the interplay between social structures and communication habits. We have shown that low communication leads to random networks with narrow degree distributions. Increased communication naturally gives nonrandom structures characterized in particular by social networks with broad degree distributions. In addition to developing broad degree distributions, the networks also tend to organize highly connected agents to connect preferably to low connected agents.

With the model, we have investigated how manipulating information influences the social structure, quantified by the topology of the emerging network. First we increased the communication frequency of individual agents. The result was striking, the more an agent chats with its surroundings, the better it performs. Increased chatting requires increased effort, but our model shows that there is both a local and a global gain to this effort.

Second we investigated the effect of cheating with the age of the information a particular agent distributes. If an agent only underestimated the time since it received the information, the agent improved its position but at a cost to the remaining system. As cheating does not cost more communication effort of the agent, it is the cheap way to optimize the social position selfishly. However, already a single cheater decreases the overall reliability to send signals across the network, reflecting a moderate global cost to this strategy.

Third we investigated the more violent strategy of lying. The lying agents pretend that they have recent information about everybody else. The strategic agents in this way succeed to attract links and thereby become central in the communication backbone network. However, only a single liar in a system with nonliars benefits from the strategy. Lying is so destructive that one liar is enough to break down the reliability of the communication and nobody is in reality a winner.

V. SUMMARY

In a broad perspective the proposed model suggests an information theoretical perspective on social and possibly also economic systems. By introducing an information game based on social links and communication rules we present an approach to the dynamics of human organization. Agents in a network use information, obtained through communication in the network, to form new links for better access to information. The introduced feedback enables us to study the topological consequences of different communication habits. In particular, we saw that communication, even when it is unequally distributed, is a benefit for everyone. Further, our studies of liars and false information illustrated that social possibilities are not solely defined by the position in the network, but also by the quality of the surrounding information.

Our scheme of investigating self-organization of networks under various degrees of limited information opens for a family of socioeconomic models, where gain and loss are quantified through redistribution of links in a network. The models will in their most simple form operate with globally homogeneous strategies, like in this minimalistic two-step model. More realistic scenarios include agents with an assortment of heterogeneous strategies of communication and link redistribution. Overall we found that unreliable communication results in chaotic or Erdős-Rényi like networks, whereas higher communication-reliability leads to structured network topologies with broad degreedistributions.

[1] D. Krackhardt and J. Hanson, Harvard Bus. Rev. **71**, 104 (1993).

[2] F. Bass, Manage. Sci. **15**, 215 (1969).

[3] J. J.Brown and P. H. Reingen, J. Consum. Res. **14**, 350 (1987).

[4] D. Knoke, *Political Networks: The Structural Perspective* (Cambridge University Press, New York, 1990).

[5] E. Rogers and D. Kincaid, *Communication Networks: Toward a New Paradigm for Research* (The Free Press, New York, 1981).

[6] B. A. Huberman and J. O. Ledyard, Inform. Syst. Front. **5,** 7 (2003).

[7] M. Rosvall and K. Sneppen (unpublished).

[8] M. Rosvall and K. Sneppen, Phys. Rev. Lett. **91**, 178701 (2003).

[9] N. E. Friedkin, Social Forces **62**, 54 (1983).

[10] P. S. Dodds, R. Muhamad, and D. J. Watts, Science **301**, 827 (2003).

[11] S. Milgram, Psychol. Today **1**, 61 (1967).

[12] T. Jeffrey and S. Milgram, Sociometry **32**, 425 (1969).

[13] J. M. Kleinberg, Nature (London) **406**, 945 (2000).

[14] J. M. Kleinberg, in *Proceedings of the 2001 Neural Information Processing Systems Conference*, edited by T. G. Dietterich, S. Becker, and Z. Ghahrmani (MIT Press, Cambridge, MA, 2002).

[15] J. M. Kleinberg, in *The small-world pheonmenon: an algorithm perspective, Proceedings of the thirty-second annual ACM Symposium on Theory of computing*, Portland, Oregon, 2000, pp. 163–170.

[16] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).

[17] D. Carpenter, K. Esterling, and D. Lazer, Ration. Soc. **15**, 411 (2003).

[18] N. E. Friedkin, Soc. Networks **3**, 273 (1982).

[19] A. Trusina, M. Rosvall, and K. Sneppen, Phys. Rev. Lett. **94**, 238701 (2005).

[20] D. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).

[21] *The Small World*, edited by M. Kochen (Ablex, Norwood, 1989).

[22] C. E. Perkins and P. Bhagwat, in *Proceedings of the ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM 94, London, 1994, p.234.

[23] J. R. Bergmann, *Discreet Indiscretions: The Social Organization of Gossip* (Aldine De Gruyter, New York, 1993).

[24] J. Davidsen, H. Ebel, and S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).

[25] P. Erdős and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[26] S. Maslov and K. Sneppen, Science **296**, 910 (2002).

[27] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).

[28] Javasimulation available at http://cmol.nbi.dk/models/inforew/inforew.html