Chapter 1

Rumor-like information dissemination in complex computer networks

Maziar Nekovee
Complexity Research Group, BT Exact, Martlesham, Suffolk IP5 3RE, UK
maziar.nekovee@bt.com
Yamir Moreno
Instituto de Biocomputación y Física de Sistemas Complejos, Universidad de Zaragoza, Zaragoza 50009, Spain
yamir@.posta.unizar.es

We investigate the dynamics of a rumor-like process for information dissemination in complex computer and communication networks. We perform large-scale Monte Carlo simulations of these process on top of a scale-free network topology, as a prototype model of networks with strongly heterogenous degree distributions, and compare the results with simulations performed for random graphs, which have a homogeneous degree distribution. Our study provides new insights on how the dissemination dynamics is affected by the complex interplay between network structure and the spreading process. Our results are relevant to other complex systems where rumor-like information dissemination takes place.

1.1 Introduction

Complex network-like structures appear in a wide variety of technological, social and biological systems. Some important examples include the Internet, the

World Wide Web, social and organizational networks, and neural networks. In recent years a substantial amount of research has been devoted to empirical and theoretical investigations of complex networks[1]. A large body of this work has been focused on finding statistical properties, such as path lengths, clustering coefficients and degree distributions that characterize the structure and behavior of networked systems, and to suggest appropriate ways to measure these properties. Another active area of research is on creating realistic models of complex networks that can help us to understand the organizational principles behind their emergence in natural and man-made systems. A third area, which is currently still in its infancy, studies the impacts of network architecture on dynamic processes that take place in networked systems.

An interesting dynamic processes in complex networks is the spontaneous spreading of information via rumor-like mechanisms. In addition to its relevance to social sciences[13], such mechanism also form the basis of an important class of data dissemination protocols in computer and communication networks. These protocols, generally known as epidemic-style or gossip protocols, have recently gained prominence as a methodology for robust and scalable group communications in large distributed systems[11]. Gossip protocols are inspired by the way that infectious diseases spread in populations and their dynamics show interesting parallels with that of epidemic disease spreading[4] (although the goal of these protocols is "infecting" the system as quickly as possible, instead of preventing the spreading). An important result in the modern mathematical theory of epidemics is that the dynamics of infection spreading is greatly affected by the structure of the contact network among individuals in populations[9]. This result suggest that the architecture of the underlying computer and communication networks should also greatly impact the dynamics of gossip protocols in these systems.

In this paper we investigate the impact of network architecture on gossip protocols through large scale Monte Carlo simulations. We focus in particular on the effect of degree distribution (the probability distribution that a randomly chosen node has k links), $P(k)$, on the dynamics. We perform our simulations for two models of complex networks: the so-called scale-free networks (SF)and random graphs (RG). Scale-free networks are models of networks in which the distribution of links (degrees) among nodes is strongly heterogenous, with the probability of a node to be connected to k other nodes obeying a power law distribution. The degree distribution of the Internet, at both the router level and the so-called Autonomous System level, is scale-free[10] Scale-free architecture has also been observed in the popular peer-to-peer file sharing systems, such as GNUTELLA and FREENET, which are formed as virtual networks on top of the Internet[8]. A random graph, on the other hand, is a prototypical example of networks with a homogeneous degree distribution. Our studies provide new insights on how the dynamics of the process unfolds in these systems, and the complex way that the process interacts with the underlying network structure. We find that gossip-style information spreading is more successful in random graphs than in scale-free networks, a result which is very different from the behavior seen in the case of disease spreading in these networks[9]. We show that this is due to the conflicting role that highly connected nodes (the so-called hubs) of SF networks play in the spreading process. In addition to computer systems, our findings are relevant to other scenarios where rumor-like information spreading takes place, such as chain emails and viral marketing.

The rest of this paper is organized as follows. In section 2 we describe the network architectures used in our simulations and introduce a generic model of epidemic protocols. In section 3 we report on our Monte Carlo simulation studies of rumor spreading in these networks, and examine the impact of network architecture on properties such as reliability and efficiency of the process. We conclude in section 4 with a summary and outlook.

1.2 Models

In this section we describe the network models used in our simulations, and present our model of epidemic protocols.

1.2.1 Network models

Our scale free networks are generated using the Barabási-Albert (BA) algorithm [3]. In this algorithm, one starts from a small number, m_0, nodes in the network. At every timestep, one new node is added to the network, and is attached with probability

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{1.1}$$

to $m(\leq m_0)$ randomly chosen nodes of the network. After sufficient iterations of this algorithm a scale-free network builds up with the following power law degree distribution $P(k) \sim k^{-\gamma}$ (for $k \geq m$), with $\gamma = 3$ and average connectivity $\langle k \rangle = 2m$. This distribution reflects the existence of a few nodes with a very high number of links and many with only a few links.

As a model of random graphs we considered the Watts-Strogatz (WS) networks[12] in the limit of complete random rewiring. In this case, one starts from a ring with N nodes, each of them connected symmetrically to $2m$ neighbors. With probability p each link connected to a clockwise neighbor is rewired to a randomly chosen node; otherwise it is preserved. After enough iterations, this algorithm produces a network which, in the random graph limit of the model ($p = 1$), has the following degree distribution [1]:

$$P(k) = \frac{m^{k-m}}{(k-m)!} e^{-m}, \tag{1.2}$$

which gives an average connectivity $\langle k \rangle = 2m$. Henceforth, we will use $m_0 = m = 3$ for the SF network and $p = 1$ and $m = 3$ for the RG network, giving $\langle k \rangle = 6$ for the average connectivity of both networks[1].

[1] An interesting feature of this choice of our model networks is that, despite their very

1.2.2 Model of rumor propagation

Gossip protocols are being used in computer and communication networks in a wide range of applications such as file discovery in peer-to-peer[8] and Grid networks[5] and reliable group communications on the Internet[6]. The general structure of these protocols is as follows. New messages are initially injected (or created) in one or more nodes in the network. These nodes then forward the message to either all or a randomly selected subset of other nodes to which they are linked. Upon receiving a forwarded message a node examines the message and decides whether or not to forward it to its neighbors [2].

The gossip protocol we consider here was introduced in the pioneering work of Demers and co-workers[4] in the context of file replication in large distributed databases. The algorithm of this protocol is very generic and its parameters could be tuned to mimic several existing protocols. We consider a system consisting of N nodes and connections between them, which form a network (these connections could be either physical links, as in the Internet, or logical links, as in peer-to-peer networks). Each of the N nodes of the network can be in three possible states. We call a node holding an update and willing to transmit it a *spreader*. Nodes that are unaware of the update will be called *ignorant* while those that already know it but are not willing to spread the update anymore are called *stiflers*. At each communication round, each spreader contacts all its neighbors. When a spreader contacts an ignorant, the last one turns into a new spreader with a probability λ. On the other hand, the spreader becomes a stifler with a probability α if it contacts another spreader or a stifler. This latter mechanism ensures the algorithm terminates in a finite time; the algorithm terminates when there are no spreaders left.

1.3 Simulation studies

We performed large-scale Monte Carlo simulations of the gossip protocol running on top of SF and RG networks. Each simulation starts from an initial state in which only one node holds the message (we call such a node the initial infected node (IIN)), and is willing to spread it; the rest of the nodes are in the ignorant state. At each timestep each spreader contacts all the nodes to which it is connected and the dynamic rules of the protocol are applied in parallel. The size of the network used in the simulations was $N = 10,000$. For a given initial infected node, and a given realization of the networks, we performed 100 Monte Carlo runs of the epidemic protocol, and averaged the results. Furthermore, unless stated otherwise, the final results were obtained by performing these Monte

different degree distributions, both networks have the well-known small-world property, i.e. the mean shortest distance between vertex pairs in the network scale logarithmically or slower with N.

[2] The forwarding decision is made based on a combination of criteria which are built into the protocol, such as the maximum number of times a node forwards a message, and the so-called time-to-live (ttl) of a message.

Carlo runs for 100 different randomly chosen initial infected nodes, and 10 different network realizations, per IIN. In the current study we fix $\lambda = 1$ in the simulations and investigate the dynamics for a range of values of α.

1.3.1 Reliability, delivery speed and load

Qualitatively the dynamics of the process can be described as follows. In the first stage of the evolution, the number of spreader nodes increases and, at a lower rate, the population of stiflers grows as well. As a consequence, the spreader-spreader and spreader-stifler contacts become more frequent resulting in an increase in the decay of spreader nodes into the stifler state. Eventually, the spreader population start to decline and vanishes, at which point dissemination stops. In figure Fig. 1.1 (left panel) we show time evolution of the number of stiflers in our networks, as obtained from Monte Carlo simulations. The results are shown for random graphs and scale-free networks, and several different values of α. It can be seen that for $\alpha = 0.1$ the protocol reaches almost the same high level of reliability (we define reliability of the protocol as the final density of stiflers in the system) in both networks. However on scale-free networks the system reaches its final state much faster. As α increases to 0.5 and then to 1.0, the reliability of the protocol dramatically drops in scale-free networks. In random graphs, on the other hand, the dissemination process maintains a rather high level of reliability, even for $\alpha = 1$. In order to further investigate the impact of parameter α on the reliability we show in Fig. 1.1 (right panel) the variations of reliability with α. It can be seen that, indeed, the reliability of the protocols on both networks decreases monotonically with increasing α. But the drop in reliability is much more graceful in random networks. Initially, one

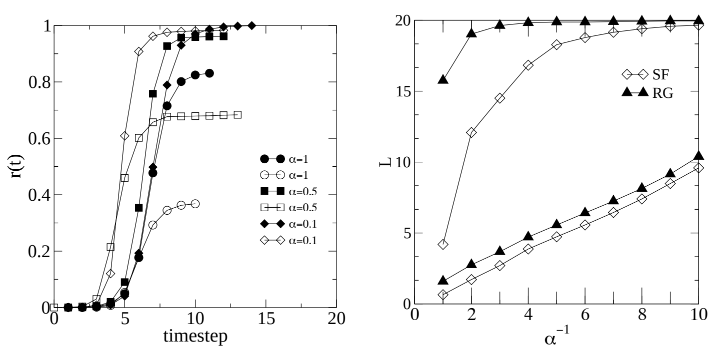

Figure 1.1: Left panel shows time evolution of the density of stifler nodes for different values of α in scale-free (open symbols) and random graphs (filled symbols). Right panel shows variations of reliability (main figure) and load (inset) with α. The system size is $N = 10,000$.

might expect that the existence of nodes with a large number of connections (the so-called hubs) in scale-free networks should help disseminate the rumor to a larger fraction of nodes than in random graphs, as is the case in the spreading of epidemics[9]. However, the presence of hubs introduces conflicting effects in the dynamics, due to spreader-spreader and spreader-stifler interactions. While a hub in the spreader state can in principle reach a very large number of nodes, it also blocks the spreading very effectively once it is switched to the stifler state. Indeed, we have found that the "lifetime" of spreaders decays exponentially with their degree, and so it is very unlikely that a spreader can contact all its ignorant neighbors before turning into a stifler. Once a few hubs are turned into stiflers many of the neighboring nodes could be isolated and never get the update.

In addition to high reliability and delivery speed, an important requirement for gossip protocols is that the network traffic generated in propagating the message remains as low as possible. As a measure of this quantity we consider the average number of messages sent per node in order to propagate one update to the network. We call this quantity load, L. The inset of Fig. 1.1 (right panel) shows the variations of load with α^{-1} for both networks. It can be seen that, for both topologies, L grows linearly as a function of α^{-1}, but the load imposed on SF networks is somewhat smaller than on WS networks. This is due to a more efficient routing of gossip traffic through the hubs in SF networks. Another interesting quantity, which is a measure of the efficiency of the protocol, is the ratio between the final number of stiflers and the total number of messages that are generated in order to reach the final state . We found that because SF networks impose a lower load on the network, the protocol runs somewhat more efficiently on these networks but the difference in efficiency between the SF and RG networks is not larger than $\sim 10\%$ [7].

1.3.2 The impact of the initial infected node

In a random graph the variations in degree distribution of nodes are rather small and so one expects that the dynamics are rather insensitive to the choice of the initial infected node. In scale-free networks, on the other hand, this choice might have a significant impact on the dynamics, and this could be used to maximize the spreading of information. To investigate this possibility, we performed another set of simulations in which we run the simulations for scale-free networks using three different initial infective nodes, which had $k = 3$, $k = k_{av} = 6$ and $k = k_{max} = 280$ links, respectively. The simulation runs were averaged over 10 different realization of the network and 200 Monte Carlo runs, per initial infected node and per network realization.

Figure 1.2 displays time evolution of the density of nodes in the stifler class when the rumor starts propagating from a node of connectivity $k = 3$, $k = 6$, $k = 280$, and for two different values of α: 1.0 (left panel) and 0.1 (right panel). For $\alpha = 1$ it can be seen that the larger the connectivity of the initial infected node, the faster the system reaches its stationary state (i.e. the delivery latency is the lowest for $k = 280$). Furthermore, the level of reliability increases with

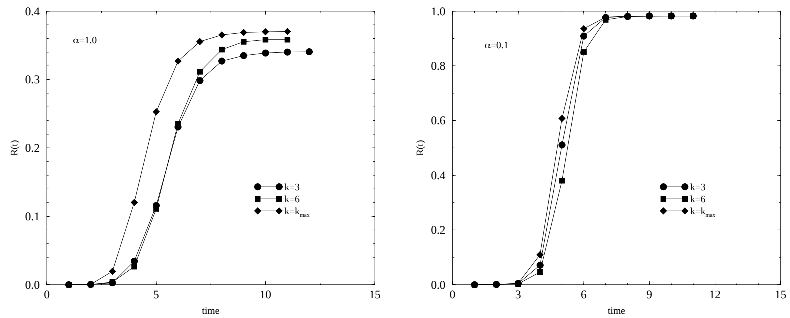

Figure 1.2: Time evolution of the number of stiflers is shown when the spreading starts at a node with $k = 3$, $k = 6$ and $k = 280$ connections, respectively, and for $\alpha = 1$ (left panel) and $\alpha = 0.1$ (right panel).

increase in the connectivity of IIN. Clearly, when the probability of spreader$\rightarrow$ stifler transition is high (α is close to 1) it is highly advantageous to start the spreading process from a hub. This is, however, not the case when α is close to 0. The reason is that for small values of α propagation paths do not get blocked by stifler nodes and so a message injected in an arbitrary node of a SF network is quickly routed to network hubs[2]. Thus the connectivity of the initial infected doesn't matter much. For $\alpha \approx 1$, however, this mechanism is much less effective and thus injecting the message directly into a hub is a more effective way of spreading the gossip.

1.4 Conclusions and outlook

In this paper we considered the dynamics of a rumor-like mechanism for information dissemination in complex computer networks, and investigated the impact of node degree distribution on this process through large scale Monte Carlo simulations. We found that the spreading process generally reaches a higher fraction of nodes when the underlying network has a homogeneous degree distribution, corresponding to a random graph topology. However, the process evolves faster, and imposes a somewhat lower load on the network, in scale-free networks, which have a strongly heterogenous degree distribution.

In the current study we mainly focused on the impact of degree distribution on the dynamics. Two other important characteristic of complex networks are the so-called clustering coefficient C, which is a measure of clique formation in a network [1], and the degree correlation function $P(k'|k)$, which is the conditional probability that a vertex of degree k is connected to a vertex of degree k'. We plan to extend our studies to more realistic models of scale-free computer networks which, unlike the BA model and random graphs, can have a high clustering coefficient and show a high level of degree correlations.

We thank Sverrir Olafsson and Keith Briggs for helpful comments.

Bibliography

[1] Albert R., Barabási, A-L. "Statistical mechanics of complex networks", *Rev. Mod. Phys.* **74** (2002), 45–512; Newman M. E. J. "The structure and function of complex networks" *SIAM Review* **64** (2003), 046135.

[2] Adamic, L. A., Lukose, R., Puniyani, A., Huberman, B. *Phys. Rev. E* **64** (2003), 046135.

[3] Barabási, A-L., Albert, R. "Emergence of scaling in random networks", *Science* **286** (1999), 509–512.

[4] Demers, A., Grene, D., Hauser, C., Irish, W. Larson, J. "Epidemic algorithms for replicated database maintenance", *Proc. of the 6th Annual ACM Symposium on Principles of Distributed Computing* (1987).

[5] Imanitchi, A., Foster, I., "On fully decentralized resource discovery in Grid environments", *Proc. of the Second International Workshop on Grid Computing* (2001), 51–62.

[6] Kermarrec, A-M., Massouli, L., Ganesh, A. J., "Probabilistic reliable dissemination in large-scale systems", *IEEE Trans. Parallel and Distributed Systems* **14** (2003), 1–11.

[7] Moreno, Y., Nekovee, M., Vespignani, A., "Reliability and efficiency of epidemic data dissemination in complex networks", *Phys. Rev. E. (R)* (in press).

[8] Oram, A. (ed.), *Peer-to-peer: Harnessing the power of of disruptive technologies*, O'Reily and Associates (2001).

[9] Pastor-Satorras, R., Vespignani A., *Phys. Rev. Lett,* **86** *(2001), 3200–3204*; May, R., Lloyd A., *Phys. Rev. E,* **64** *(2001), 066112–06117*;

[10] Siganos G., Faloutsos M., Faloutsos P., Faloutsos C., "Power-law and the AS-level Internet topology", *IEEE/ACM Transactions on Networking* **11** (2003), 514-524.

[11] Vogel, W., van Renesse, R., Birman, K., "The power of epidemics: Robust communication for large-scale distributed systems", *Proc. of HotNets-I* (2002).

[12] Watts D. Strogatz Watts, D. J., Strogatz, S. *Nature* **393** *(1998), 393-395.*

[13] "Dynamics of rumor propagation in small-world networks", *Phys. Rev. E. (R)* **64** *(2003), 050901-050905.*