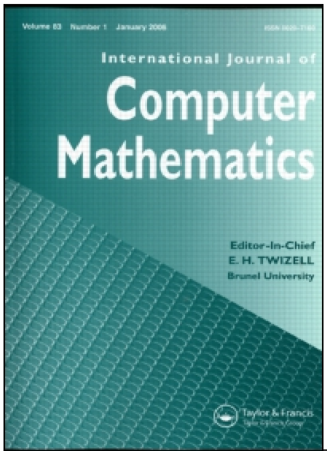

International Journal of Computer Mathematics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713455451

Dynamics of gossip-like information dissemination in complex computer networks

Maziar Nekovee [ab]; Yamir Moreno [c]
[a] Complexity Research Group, BT Research, Suffolk, UK
[b] Centre for Computational Science, University College London, London, UK
[c] Instituto de BiocomputaciÓn y FÍsica de Sistemas Complejos, Universidad de Zaragoza, Zaragoza, Spain

Online Publication Date: 01 August 2008

To cite this Article: Nekovee, Maziar and Moreno, Yamir (2008) 'Dynamics of gossip-like information dissemination in complex computer networks', International Journal of Computer Mathematics, 85:8, 1165 — 1173

To link to this article: DOI: 10.1080/00207160701763065
URL: http://dx.doi.org/10.1080/00207160701763065

International Journal of Computer Mathematics
Vol. 85, No. 8, August 2008, 1165–1173

Dynamics of gossip-like information dissemination in complex computer networks

Maziar Nekovee[a,b]* and Yamir Moreno[c]

[a] *Complexity Research Group, BT Research, Polaris 134, Adastral Park, Suffolk, UK;* [b] *Centre for Computational Science, University College London, London UK;* [c] *Instituto de Biocomputación y Física de Sistemas Complejos, Universidad de Zaragoza, Zaragoza, Spain*

(*Received 18 June 2007; revised version received 11 October 2007; second revision received 17 October 2007; accepted 18 October 2007*)

We investigate the dynamics of a gossip-like process for information dissemination in complex computer networks. We perform large-scale Monte Carlo simulations of this process on top of a scale-free network topology, as a prototype model of networks with strongly heterogeneous degree distributions, and compare the results with simulations performed for random graphs, which have a homogeneous degree distribution. In addition to the above static networks, we also investigate the spreading process on time-dependent networks created by mobile wireless nodes (mobile adhoc networks). Our study provides new insights on how the dissemination dynamics is affected by the complex interplay between network structure, mobility and the spreading process. Our results are also relevant to other complex networks where gossip-like information dissemination takes place.

Keywords: complex networks; stochastic prococess, gossip protocls; distributed computing

1. Introduction

Complex network-like structures appear in a wide variety of technological, social and biological systems. Some important examples in computer science include the Internet, peer-to-peer systems and email networks [2,5,14]. An interesting dynamic process taking place in complex networks is the spontaneous spreading of information via rumour-like mechanisms. In addition to its relevance to social sciences, such a mechanism also forms the basis of an important class of data dissemination algorithms in distributed computer systems. These algorithms are generally known as *gossip*, or *epidemic*, protocols (for recent reviews of epidemic algorithms see [8,20]).

The principle underlying these algorithms mimics the spread of rumour among humans or epidemics in populations. A process that wishes to disseminate a new piece of information to the system does not send it to a server, or a cluster of servers, in charge of forwarding, but rather to a (randomly chosen) set of other peer processes that it *knows*. In turn each of these processes does the same and also forwards the information to some of its peers. In addition to

*Corresponding author. Email: maziar.nekovee@bt.com

ISSN 0020-7160 print/ISSN 1029-0265 online

DOI: 10.1080/00207160701763065
http://www.informaworld.com

their attractive scalability promises, these algorithms exhibit high resilience in the presence of frequent link and/or process failures due to their distributed nature. Not surprisingly, the use of gossip algorithms has been explored in distributed applications involving a large number of processes with a dynamic connection topology. These include replication updates in distributed database systems, resource discovery in peer-to-peer networks and reliable multicasting in peer-to-peer and mobile adhoc networks. These experimentations have revealed several nontrivial performance issues that need to be addressed before gossip algorithms can be applied in practical settings.

The two basic ingredients that control the performance of gossip protocols are the forwarding policy that each node follows upon receiving a gossip message, and the logical connection network along which messages are forwarded (note that this connection network is simply an abstraction of a *who knows whom* relationship between the nodes, and should not be confused with the physical network). Most previous theoretical studies of gossip protocols have analysed the performance of these protocols using highly simplified models of connection network. Such simplified models become realistic if we assume that the gossip scheme is deployed within small-sized systems [4], and have the advantage that they can be treated analytically. However, in applications to large and dynamic systems, such as peer-to-peer networks, it is more realistic to assume that each process can only 'know' a small subset of other processes, and therefore highly complex topologies for the connection networks can emerge. It is imperative then to understand how such complex topologies impact the performance of gossip protocols.

In a recent paper [12], we derived a set of mean-field equations for the spreading of rumours in complex social networks and used these equations to numerically investigate rumour dynamics in such networks. These mean-field equations, however, were derived for *continuous time asynchronous* spreading of rumours, which is appropriate to human communication activities. Furthermore, they only consider static network topologies. In the current paper, we consider a *discrete time synchronous* gossip-based process, which is more relevant to information dissemination in computer networks. We use large-scale Monte Carlo simulations to investigate in detail the dynamics of this process on several models of complex computer networks. The simulations are performed for two models of static complex networks: the so-called scale-free networks (SFs) and random graphs (RGs), and on time-dependent peer-to-peer networks created by mobile wireless nodes (mobile adhoc networks).

Scale-free networks are models of networks in which the distribution of links (degrees) among nodes is strongly heterogeneous, 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 [19]. Scale-free architecture has also been observed in the popular peer-to-peer file sharing systems, such as Gnutella, which are formed as virtual networks on top of the Internet [15]. A RG topology, on the other hand, is a prototypical example of networks with a homogeneous degree distribution. Furthermore, the RG topology is being deployed in a number of recent implementations of gossip protocols [10]. Finally, mobile adhoc networks can be created on the fly by a collection of handheld devices (such as smartphones and PDAs), which communicate using short-range radio transmissions [9]. Adhoc technology has important applications in the provisioning of ubiquitous wireless Internet access, disaster relief operations and wireless sensor networks. From the perspective of complex network theory, the study of these networks is important as they provide a clear-cut example of the much less explored *time-dependent* networks.

Our studies provide new insights on how the dynamics of the spreading process unfold in these systems, and the complex way that control parameters of gossip algorithms interact with the underlying network structure and its dynamics. Furthermore, our study reveals that in SFs, both the delivery speed and the reliability of the spreading process are greatly affected by the degree distribution of the initial seed from which the process starts.

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 gossip protocols. In Section 3, we report on our Monte Carlo simulation studies of gossip spreading in these networks and examine the impact of network architecture and network dynamics on properties such as reliability, efficiency and delivery speed of gossip-like information dissemination. We close this paper in Section 4 with conclusions.

2. Models

In this section, we describe the network models used in our simulations and present our gossip algorithm.

2.1. *Network models*

As a model of strongly heterogeneous networks we use the SFs 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 time step, one new node is added to the network and is attached with probability

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{1}$$

to $m(\leq m_0)$ randomly chosen nodes of the network. After sufficient iterations of this algorithm, a SF builds up with the following power-law degree distribution: $P(k) \sim k^{-\gamma}$ (for $k \geq m$), with $\gamma = 3$ and average degree distribution $\langle k \rangle = 2m$. 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 RGs we consider the Watts–Strogatz (WS) small world networks [21] in the limit of complete random rewiring. In this case, one starts from a ring with N nodes, each of them connected symmetrically to $2m$ neighbours. With probability p, each link connected to a clockwise neighbour is rewired to a randomly chosen node; otherwise it is preserved. After enough iterations, this algorithm produces a network which, in the RG limit of the model ($p = 1$), has the following degree distribution [2]:

$$P(k) = \frac{m^{k-m}}{(k-m)!} \mathrm{e}^{-m}, \tag{2}$$

which gives an average degree $\langle k \rangle = 2m$. Henceforth, we will consider the RG limit of the WS network and use $m_0 = m = 3$ for the SF network and $m = 3$ for the RG network. These choices give $\langle k \rangle = 6$ for the average connectivity of both networks. We note that the BA model generates only a specific class of networks with a power-law degree distribution. Other more elaborate algorithms do exist, which can generate a whole ensemble of SF networks [16]. However, the interesting feature of our choice of the BA model is that while the resulting network has degree distributions, which are radically different from the RG network, it shares two other important properties with the RG network, namely, the small-world property, that is, the mean shortest distance between vertex pairs in the network scales logarithmically or slower with N, and a very small clustering coefficient. Consequently, using these models allows us to isolate the impact of degree distribution on the resulting dynamics of the rumour spreading process.

Finally, we model the topology of wireless adhoc networks in the following way [13]. We distribute N wireless nodes randomly and uniformly in a two dimensional plane of area A, and create a link between any two nodes, which are within a communication distance r_t of each other. In the case that nodes do not move the topology of the resulting network can be described as a two dimensional random geometric graph (RGG) [6,18]. Like RGs, these graphs have a heterogeneous degree distribution, which peaks at an average value $\langle k \rangle = \pi r_t^2 N/A$ and shows small fluctuations around this value. However, other properties of a RGG are radically different from RG. Most notably, these networks are characterized by a large cluster coefficient and a mean shortest distance between vertex pairs, which scales $\sim\sqrt{N}$ with the network size. A time-dependent network topology is obtained from RGG by allowing nodes to move in and out of each other's communication range according to a random walk mobility model.

2.2. *Model of gossip propagation*

Many variants of gossip protocols exist and are typically distinguished by the forwarding policy that is used during the dissemination process. The gossip algorithm we consider here was introduced in the pioneering work of Demers and co-workers [7] in the context of file replication in large distributed databases, and shows close analogies with the way rumours spread in human populations.

We consider a system consisting of N nodes and logical connections between them, which form a gossip network. Each of the N nodes of the network can be in one of three possible states. We call a node holding an update and willing to forward 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 any further are called *stiflers*. At each communication round, each spreader contacts a randomly selected subset of its neighbours. 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 during the forwarding process. This latter mechanism ensures that the algorithm terminates in a finite time; the algorithm terminates when there are no spreaders left. The parameter α can be tuned to control the behaviour of the algorithm while the parameter λ mimics the impact of non-permanent link and node failures on performance.

We note that the above algorithm is very generic and its parameters could be tuned and further extended to mimic a number of more elaborate protocols. These include gossip protocol in which the parameters λ and α are functions of various network parameters or can change as a function of time. However, since our focus here is to investigate the impact of network topology on performance, we choose to perform our analysis for the simple version of the protocol in which the protocol parameters are fixed during the spreading process.

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 communication round, 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 IIN and a given realization of the networks, we performed 100 Monte Carlo runs of the gossip protocol and averaged the results. Furthermore, unless stated otherwise, the final results were obtained by performing these Monte Carlo runs for 100 different randomly chosen IINs, 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.

3.1. *Reliability, delivery speed and load*

Qualitatively the dynamics of the gossip 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 starts to decline and vanishes at which the point dissemination stops.

In Figure 1, we show time evolution of the number of stiflers in our networks, as obtained from Monte Carlo simulations. The results are shown for RGs and SFs, 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 SFs 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 SFs. In RGs, 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 Figure 2 the variations of reliability with α^{-1}. It can be seen that the reliability of the protocols on both networks decreases monotonically with decreasing α^{-1}, but the drop in reliability is much more graceful in random networks.

Initially, one might expect that the existence of nodes with a large number of connections (the so-called *hubs*) in SFs should help disseminate the gossip to a larger fraction of nodes than in RGs, as is the case in the spreading of epidemics [17]. 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 neighbours before turning into a stifler. Once a few hubs are turned into stiflers, many of the neighbouring nodes could be isolated and never receive the

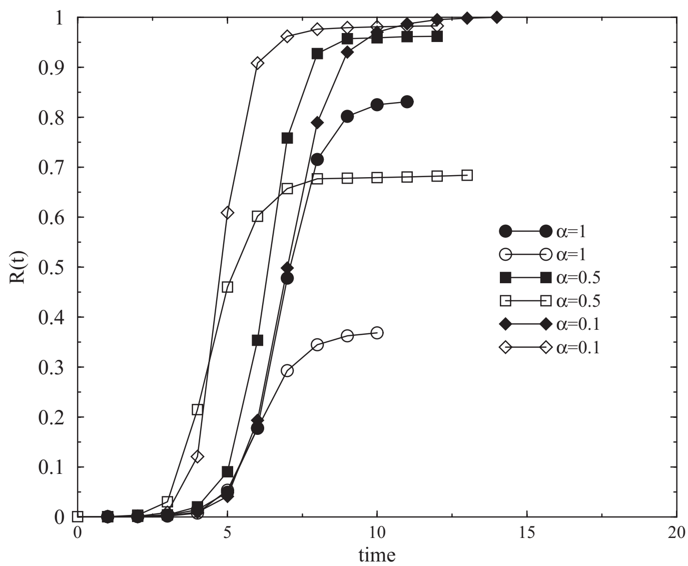

Figure 1. Time evolution of the density of stifler nodes for different values of α in scale-free (open symbols) and random graphs (filled symbols) is shown. The system size is $N = 10{,}000$.

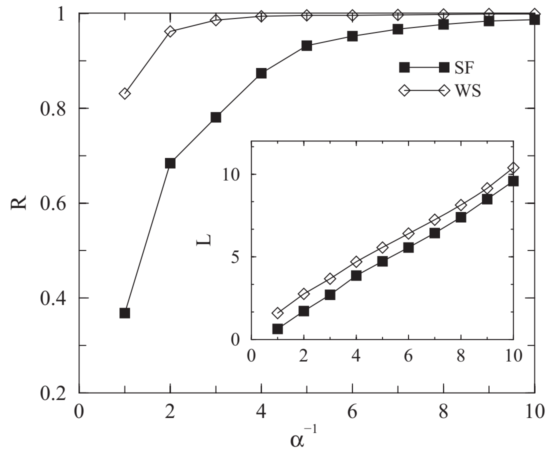

Figure 2. Variations of reliability (main figure) and load (inset) with α are shown for the same system as in Figure 1.

update. Random graphs, on the other hand, allow for a more reliable dissemination of information, since all nodes contribute equally to the message passing.

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 Figure 2 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 ∼10% [11].

3.2. *The impact of the initial infected node*

In RG, 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 IIN. In SFs, 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 SFs using three different initial infective nodes, which had $k = 3$, $k = k_{\rm av} = 6$, and $k = k_{\rm max} = 280$ links, respectively. The simulation runs were averaged over 10 different realization of the network and 200 Monte Carlo runs, per IIN and per network realization.

Figure 3 displays time evolution of the density of nodes in the stifler class when the gossip 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 IIN, the faster the system reaches its stationary state (that is, the delivery latency is the lowest for $k = 280$). Furthermore, the level of reliability increases with increase in the connectivity of IIN. Clearly, when the probability of spreader to 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

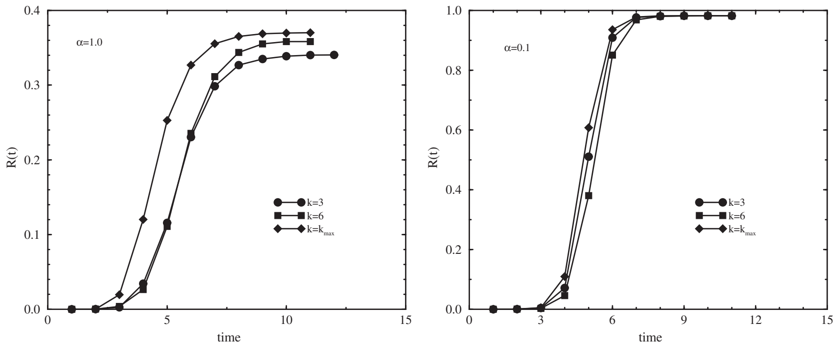

Figure 3. Time evolution of the number of stiflers is shown when the spreading starts at a node with $k = 3$, $k = 6$ and $k = 280$ connections, respectively, and for $\alpha = 1$ (left panel) and $\alpha = 0.1$ (right panel).

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 into an arbitrary node of a SF network is quickly routed to network hubs [1]. Thus the connectivity of the IIN does not 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.

3.3. *The impact of node mobility: time-dependent topologies*

Finally, we briefly discuss the impact of node mobility on the spreading process in the context of mobile adhoc networks. Unlike the networks considered in previous section, which are connected, mobile adhoc networks are usually fragmented into isolated clusters which merge and disintegrate dynamically as nodes move in and out of each other's communication range. It is important, therefore, to investigate how our rumour protocol performs in such highly dynamic settings.

In Figure 4, we display representative examples of our simulations of gossip spreading in a wireless adhoc network consisting of 200 nodes, each having a 100-m transmission range. The spreading process was first simulated on top of a static network. Subsequently, nodes were allowed to perform random walks during the spreading process. It can be seen that in the static network a substantial fraction of nodes do not receive the rumour message. This is due to the above-mentioned network fragmentation, which prevents the spreading of the gossip beyond the network cluster in which it was generated. In contrast, in the time-dependent network the rumour can spread throughout the network thanks to mobile spreaders, which carry the message with

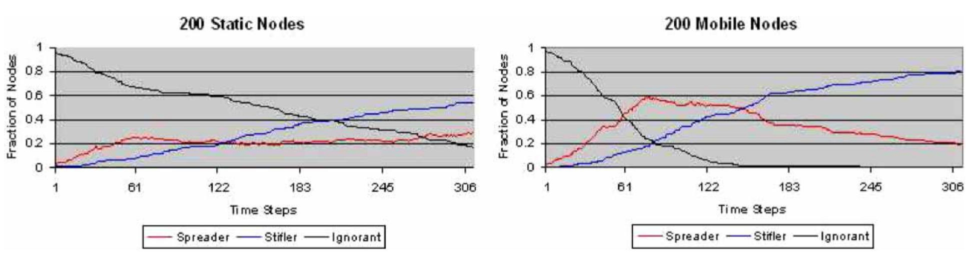

Figure 4. Time evolution of rumour spreading in a fragmented wireless adhoc network. Results are shown for a static network (left panel) and a time-dependent network where nodes follow random walks. The protocol parameters used are $\lambda = 1$ and $\alpha = 0.1$.

themselves as they move from one network cluster to another. Furthermore, it can be seen that mobility also greatly increases the speed of spreading.

4. Conclusions

In this paper, we have considered the dynamics of a gossip-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 RG topology. However, the process evolves faster, and imposes a somewhat lower load on the network, in SFs, which have a strongly heterogeneous degree distribution. We found, however, that on the whole the impact of network topology on the gossip spreading process is not as dramatic as in the case of epidemic spreading. In particular, recent studies of epidemic models on complex networks have shown that the threshold for an epidemic outbreak is vanishing in infinitely large SFs, while it has a finite value in network with homogeneous degree distribution [17]. Our simulations, on the other hand, show that due to the conflicting roles played by hubs in the spreading process, the reliability of gossip protocols do not show such a threshold behaviour at all, regardless network topology.

Finally, we presented initial results of rumour spreading in time-dependent networks created by mobile wireless devices. Our results indicate that node mobility can significantly impact the dynamics of spreading in these networks. In particular, we found that mobility greatly improves the reliability of information dissemination in fragmented mobile adhoc networks.

In the current study, we mainly focused on the impact of degree distribution on the dynamics. Two other important characteristics of complex networks are the so-called clustering coefficient, C, which is a measure of clique formation in a network [2,14], 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 RGs, can have a high clustering coefficient and show a high level of degree correlations. Furthermore, we are developing new simulation tools and analytical models in order to further investigate spreading dynamics in large-scale mobile adhoc networks.

References

[1] L.A. Adamic et al., *Search in power-law networks*, Phys. Rev. E 64 (2003), pp. 046135–046138.
[2] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys. 74 (2002), pp. 450–512.
[3] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, Science 286 (1999), pp. 509–512.
[4] K.P. Birman et al., *Bimodal multicast*, ACM Trans. Compu. Syst. 17(2) (1999), pp. 41–88.
[5] S. Boccaletti et al., *Complex networks: Structure and dynamics*, Phys. Rep. 424 (2006), pp. 175–308.
[6] J. Dall and M. Christensen, *Random geometric graphs*, Phys. Rev. E. 66(2002), 0161121, pp. 1–9.
[7] A. Demers et al., *Epidemic algorithms for replicated database maintenence*, Proc. of the 6th Annual ACM Symposium on Principles of Distributed Computing, Vancouver, Canada, 1987.
[8] P.T. Eugster et al., *Epidemic information dissemination in distributed systems*, IEEE Computer Mag. 20 (2004), pp. 60–67.
[9] R. Hekmat, *Adhoc Networks: Fundamental Properties and Network Topologies*, Springer, Dordrecht, The Netherlands, 2006.
[10] A.-M. Kermarrec, L. Massouli, and A.J. Ganesh, *Probabilistic reliable dissemination in large-scale systems*, IEEE Trans. Parallel Distrib. Syst. 14 (2003), pp. 1–11.
[11] Y. Moreno, M. Nekovee, and A. Vespignani, *Efficiency and reliability of epidemic data dissemination in complex networks*, Phys. Rev. E. 69 (2004), 055101 (R), pp. 1–4.
[12] M. Nekovee et al., *Theory of rumour spreading in complex social networks*, Physica A. 374 (2007), pp. 457–470.
[13] M. Nekovee, *Worm epidemics in wireless adhoc networks*, New. J. Phys. 9 (2007), pp. 189–192.
[14] M.E.J. Newman, *The structure and function of complex networks*, SIAM Rev. 45 (2003), pp. 167–256.
[15] A. Oram (ed.), *Peer-to-peer: Harnessing the Power of Disruptive Technologies*, O'Reilly and Associates, New York, 2001.

[16] J. Park and M.E. Newman, *Statistical mechanics of networks*, Phys. Rev. E. 70 (2004), 066117, pp. 1–13.

[17] R. Pastor-Satorras and A. Vespignani, *Epidemic spreading in scale-free networks*, Phys. Rev. Lett. 86 (2001), pp. 3200–3204.

[18] M. Penrose, *Random Geometric Graphs*, Oxford University Press, Oxford, 2003.

[19] G. Siganos et al., *Power laws and the AS-level internet topology*, IEEE/ACM Trans. Netw. 11 (2003), pp. 514–524.

[20] W. Vogels, R. Van Renesse, and K.P. Birman, *The power of epidemics: robust communication for large-scale distributed systems*, ACM SIGCOMM Compu. Commun. Rev. 33 (2003), pp. 131–135.

[21] D.J. Watts and S. Strogatz, *Collective dynamics of small-world networks*, Nature 393 (2003), pp. 393–395.