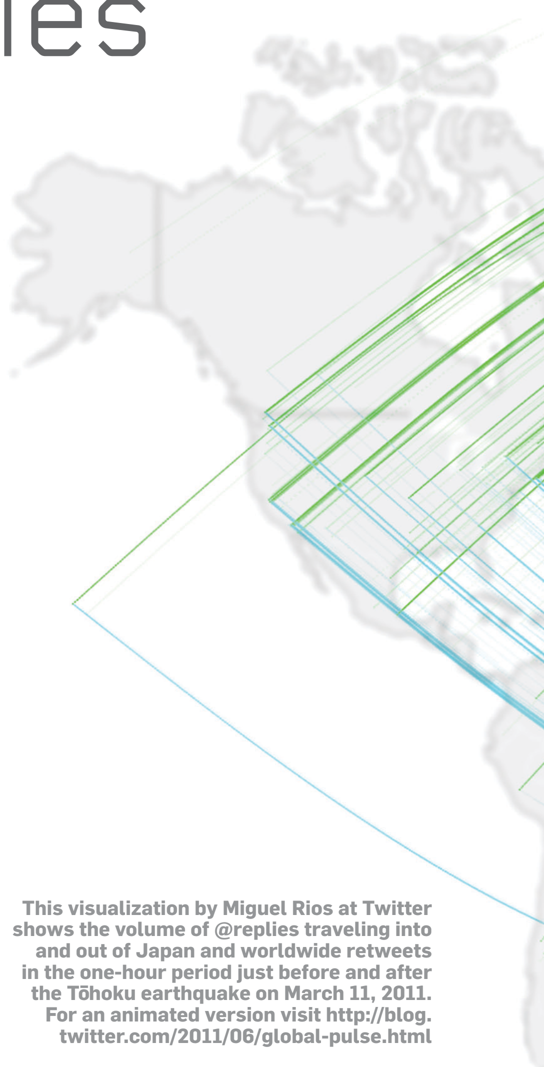
DOI:10.1145/2184319.2184338

A few hubs with many connections share with many individuals with few connections.

BY BENJAMIN DOERR, MAHMOUD FOUZ, AND TOBIAS FRIEDRICH

Why Rumors Spread So Quickly in Social Networks



This visualization by Miguel Rios at Twitter shows the volume of @replies traveling into and out of Japan and worldwide retweets in the one-hour period just before and after the Tōhoku earthquake on March 11, 2011. For an animated version visit http://blog.twitter.com/2011/06/global-pulse.html

UNDERSTANDING STRUCTURAL AND algorithmic properties of complex networks is important, due in part to the Internet's global social and commercial importance. Our focus here is to analyze how news spreads in social networks, simulating a simple information-spreading process in various network topologies and demonstrating that news spreads much more quickly in existing social-network topologies than in other network topologies. We support this finding by analyzing information spreading in the mathematically defined preferential attachment (PA) network topology, a common model for real-world networks, proving that sublogarithmic time suffices to spread news to all nodes of a network. All previously studied network topologies need at least logarithmic time. Surprisingly, nodes with few neighbors are crucial for quick dissemination.

Social networks like Facebook and Twitter are reshaping the way people communicate and take collective action, playing a crucial role in, for example, the 2011 Arab Spring uprisings and London riots. It has been argued that the "instantaneous nature" of these networks influenced the speed

» key insights

- **Social networks grow naturally; their graph structure is not designed for any particular use but still allows for the quick spread of news.**
- **Our mathematical proof shows rumors in social networks spread much more quickly than in most other network topologies, even complete graphs.**
- **The source of the speedy spread of information is fruitful interaction between the few nodes with many neighbors and the large number of nodes with few neighbors.**

VISUALIZATION BY MIGUEL RIOS AT TWITTER

actual real-world events unfolded.[4] Social networks spreading news so quickly is remarkable; the structure of social networks and the process that distributes news were not designed for this purpose. On the contrary, they were not designed at all but evolved in a random and decentralized manner.

Is it correct to assume social networks ease the spread of information ("rumors")? And, if they do, which of their properties make it possible? To answer, we simulated a simple rumor-spreading process on several graphs reflecting the structure of existing large social networks (see Figure 1). For example, a rumor begun at a random node of the Twitter network reaches on average 45.6 million of the total of 51.2 million members within only eight rounds of communication.

We also analyzed this process on an abstract model of social networks, the so-called PA graphs introduced in 1999 by Barabási and Albert.[3] In Doerr et al.,[17] we devised a mathematical proof that rumors in these networks spread much more quickly than in many other network topologies—even quicker than in networks with a communication link between any two nodes (complete graphs). As an explanation, we observe that nodes with relatively few neighbors build a shortcut between nodes with large degree (hubs) that, due to their large number of possible communication partners, talk less often to each other directly.

Rumor Spreading

Social networks arise in a variety of contexts, formed by people connected by knowing each other, Facebook members agreeing to be friends (in Facebook), scientific authors having a joint publication, and actors appearing in the same production.

Despite this diversity, many different types of networks share characteristic properties. Well known is that any two individuals are connected through just "six degrees of separation," as first formulated in 2003 by Frigyes Karinthy (see Barabási[2]) and became known to a broad audience through Stanley Milgram's "small world study."[25] Likewise, for the Web, Albert et al.[1] predicted a diameter (maximum distance between two nodes in the graph) of only 19 connections in the network formed by Web pages and the links between them.

Several experimental studies[1,13,22] have revealed another intrinsic property of social networks: The histogram of node connectivity follows a power law; the number of nodes with k neighbors is inversely proportional to a polynomial in k. To explain this phenomenon, Barabási and Albert[3] suggested the PA model for real-world networks that show a power law. The model is widely used, due to its simplicity, and the article was the fifth most cited in *Science*, as of April 2012, according to ISI Web of Knowledge (http://www.webofknowledge.com/). Graphs in the PA model are constructed in a random, "rich-get-richer" fashion; a newly entering node connects to m existing nodes chosen randomly but gives preference to nodes that are already popular; that is, they have many neighbors. Note that the parameter m controls the density of the graph, or the ratio of the number of present edges to the number of all possible edges. For these graphs, Barabási and Albert[3] empirically discovered a power law of k^{-3} later proved mathematically by Bollobás et al.[11] Similar models emerged at the time, all leading to a power-law distribution. Also known is the PA model does not share all properties observed in real-world networks; for example, it is less clustered.

Still, the PA model has helped deduce many interesting properties of social networks. Famous structural results prove small diameters for such graphs,[10] determine their degree (in terms of number of neighbors) distribution,[11] and show high expansion properties[24] and robustness against random damage, along with vulnerability against malicious attacks.[8,9,15,18] Algorithmic work shows that in such networks, viruses spread more easily than in many other network topologies[5] and that gossip-based decentralized algorithms are better at approximating averages.[12]

Here, we assume a rumor is sufficiently interesting so people learn it when talking to someone who knows it. This is substantially different from the probabilistic virus-spreading model[5] where the probability of being infected is proportional to the number of neighbors infected. Two types of rumor-spreading mechanisms have been identified in the literature: In the push model, only nodes that know the rumor contact neighbors to inform them, and is used to transmit information in computer networks.[16,21] In contrast, to capture the effect of gossiping in social networks, it is more appropriate for uninformed nodes to actively ask for new information. We use the push-pull model of Demers et al.[16] (see also Karp et al.[20]) in which all nodes regularly contact a neighbor and exchange all the information they have.

We assume nodes choose their communication partners uniformly at random from their neighbors, excluding the person they contacted immediately before. In this model, we regard the spread of a single piece of information initially present at a single node. For simplicity, as in most previous work, the rumor-spreading process is synchronized; that is, it takes place at discrete points over time and each time step, with each node contacting a neighbor to exchange information. This simplifies a real-world scenario where users act at different speeds, but in sufficiently large networks these differences balance out and thus assume an average speed used by all nodes.

Note that the communications process is different in each social network. The push-pull model we analyze is naturally best at capturing personal communication between two individuals by, say, phone, text message, email message, or other directed communication. Many online social networks also allow other ways to communicate (such as posts on personal Web pages), possibly resulting in friends being notified of a post when they next log in and then forwarding the news, given that it is of sufficient interest. Such forms of communication can be modeled only by more complicated means than the push-pull protocol.

Experimental Results

Supporting the notion that news spreads quickly in social networks, we have simulated the rumor-spreading process on samples from the Twitter and Orkut social networks (from Bhattacharjee et al.[6,23]), as well as on PA graphs. As most social networks have roughly similar structure, we chose these large networks (data readily available) as instances of social networks. For comparison, we also included complete graphs and random-attachment graphs (also called the m-out model, as in, for example,

Figure 1. How a rumor progresses from a large-degree node *A* to another node *B*; due to their having large number of neighbors, *A* and *B* need more time to push or pull the rumor.

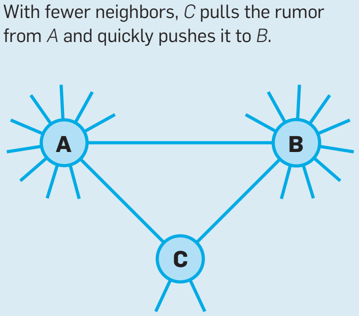
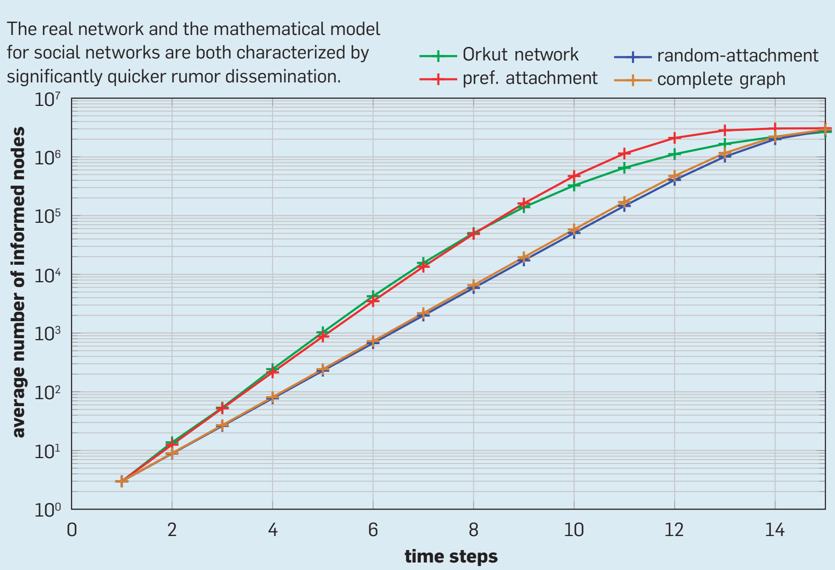


Figure 2. Average number of informed nodes as a function of time for the Orkut network and preferential attachment, random attachment, and complete graphs of the same size n = 3,072,441 nodes and density parameter m = 38.



Bohman and Frieze[7]) in which each node chooses m neighbors uniformly at random from all nodes.

Note that in both the random attachment (RA) and the PA graph model, we are able to control the density of the graph through the parameter m, allowing us to run experiments on graphs with the same number of nodes and density as equivalent real-world graphs.

Figures 2 and 3 reflect how a rumor begun in a random node spreads in graphs corresponding to the Orkut and Twitter social networks. News spreads much more quickly in real-world networks and in PA graphs than in complete and RA graphs. In the Twitter experiment, a considerable difference is apparent between the PA model and the real-world graph, indicating the Twitter graph is captured less well by the theoretical model.

To determine how graph size might influence the speed a rumor spreads, we ran the process on PA graphs, RA graphs, and complete graphs of varying size. The results (see Figure 4) indicate logarithmic time is needed for RA and complete graphs, whereas for PA graphs time is needed of a slightly smaller order of magnitude.

Surprisingly, nodes with few neighbors are crucial for quick dissemination.

Mathematical Analysis

We supported this empirical finding through mathematical analysis, proving the rumor-spreading process disseminates news in sublogarithmic time, or the time needed to inform all nodes, as well as any constant fraction (such as 1%, 10%, or 50%).

We denote by G_m^n the PA graph on n nodes with density parameter m. Since the graph G_m^n is a random graph, none of the statements mentioned earlier holds with certainty. However, the probability that the random graph G_m^n does not satisfy our assumptions and observations tends to zero for n growing to infinity. In the following paragraphs, when we make a statement concerning a random object, that statement is meant to hold with high probability. For the PA graph G_m^n, with m being any constant larger than one, we were able to prove that after a surprisingly short time any given news item spreads to all nodes.

Theorem 1. There is a constant K such that the rumor-spreading process we described spreads a rumor from any starting node to all other nodes in at most $K \cdot \log(n)/\log(\log(n))$ time steps. This result improves the previous best bound for rumor spreading in PA graphs[14] of order $\log(n)^2$. Theorem 1 showed for the first time (in 2011) that rumor spreading in PA graphs is much quicker than for the other network topologies covered here so far. For RA graphs, it is easy to see that the diameter is of order $\log(n)$, which is also clearly a lower bound for rumor-spreading time. Similar reasoning also shows that, independent of starting node, a constant fraction of all nodes has distance $\Theta(\log(n))$ from the starting node. Hence a logarithmic number of rounds is needed to inform any constant fraction of the nodes. Similar bounds follow for hypercube networks common in computer science applications.

However, the diameter of the network does not always reveal the time needed to spread a rumor. The complete graph has a diameter of exactly one, but time to all other networks is of logarithmic order. This result was proved in 2000 by Karp et al.[20] for a rumor-spreading process in which nodes were allowed to choose their random communication partners from among all neighbors, including the one they might have just talked to. It is not difficult to see the Karp et al. proof is valid in our setting. That proof also shows that a logarithmic number of rounds is still necessary to inform any constant fraction of the nodes. Similar results hold for the classical Erdös-Rényi random graphs as might be deduced from Fountoulakis et al.[19] and Karp et al.[20] All these results were proved through mathematical means; that is, they did not rely on experiments conducted for certain graph sizes n but are valid for all graph sizes. For the proof of Theorem 1 see Doerr et al.,[17] though here we outline a main argument that also explains why rumor spreading in social networks is so speedy.

Toward this goal, let A and B be neighboring nodes in G_m^n. We denote their degrees by d_A and d_B. Assume that A is informed and B is not. How does the rumor progress from A to B? Since A contacts its neighbors randomly, it will take approximately d_A rounds until A contacts B; thus A pushes the news to B. Likewise, it takes an expected number of approximately d_B time steps until B calls A, pulling the rumor from there. If d_A and d_B are large (such as $n^{1/3}$), then it would take an expected number of almost $1/2n^{1/3}$ rounds to propagate the rumor from A to B along the direct link.

Theorem 1 established a much smaller bound. Hence there must be a better way to get the news from A to B. It is the small-degree nodes that make the difference. Now assume there is a third node C that is a neighbor of both A and B and has a small degree d_C of, say, only four neighbors. After an expected number of roughly $d_C = 4$ rounds, C will have contacted A and thus learned the news from A. Likewise, after another expected number of $d_C = 4$ rounds, C will have contacted B and told it the news. That is, in $2 \cdot d_C = 8$ time steps, the rumor went from A to B through C. Fortunately, such a node C exists with high probability; the PA rule ensures newly entering nodes put enough preference on connecting with A and B.

Figure 3. Average number of informed nodes over time for the Twitter network and comparable preferential attachment, random attachment, and complete graphs (size n = 51,217,936 nodes, density parameter m = 32).

The two plots modeling social networks again outperform the random attachment and complete graph.

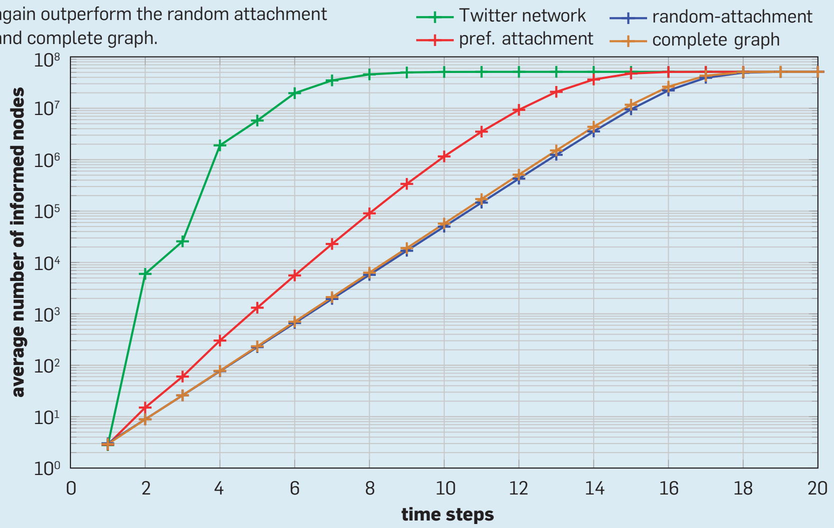
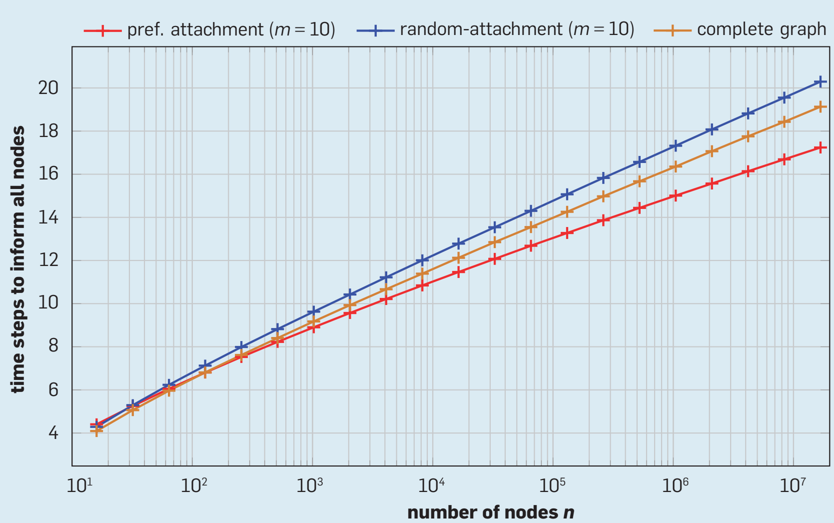


Figure 4. Average time needed to inform all nodes of different networks of varying size; the data shows logarithmic dependence for random-attachment and complete graphs.

For preferential attachment graph (a mathematical model for social networks), the times appear to be of lower order. Note that the generally speedy propagation partially obscures the advantage of the preferential attachment network structure. Alternatively, if one is willing to include 16 rounds of rumor-spreading communications between people (on average), then one can share a rumor with more than 300,000 others, as in the preferential attachment model, but with only around 30,000 people organized in random-attachment fashion.



One mechanism enabling the spread of rumors in social networks is that small-degree nodes learn the rumor from an informed neighbor, then quickly forward it to all other neighbors. In a sense, they act as an automatic link between their neighbors; once one neighbor is informed, then all other neighbors are informed—without doing anything. Such a mechanism is missing (such as in complete graphs) because all nodes have a high degree of $n - 1$. Consequently, all neighbors of the starting-node A have a small probability of calling A and asking for the news; A is just one of their $n - 1$ neighbors.

Also note that such high-speed links are abundant in PA graphs. To clarify, call a node popular if it has a degree of $\Theta(\log(n)^2)$ or higher. We can now show that between any two popular nodes, there is a path of length $O(\log(n)/\log(\log(n)))$ such that every second node on the path has the minimal possible degree of m. As per our assumptions, equations, and observations these nodes function as quick links, propagating rumors in an expected number of roughly $2 \cdot m$ rounds. Consequently, the expected time a rumor must traverse the whole path is about m times its length. With extra care, namely by showing there is a huge number of such paths between any two popular nodes, we can even show that once a popular node is informed, after $O(\log(n)/\log(\log(n)))$ rounds with high probability, all popular nodes are informed.

Since nodes tend to attach to popular nodes, a rumor started in a small-degree node is propagated to some popular node even quicker, namely in $O(\log(n)^{3/4} \log(\log(n)))$ rounds. Once all popular nodes are informed, a symmetric argument can show that after another $O(\log(n)^{3/4} \log(\log(n)))$ rounds, the remaining small-degree nodes, mostly by calling more popular nodes, would all be informed; for more, see Doerr et al.[17]

Conclusion

We simulated a natural rumor-spreading process on various graphs representing real-world social networks and several classical network topologies. We also performed a mathematical analysis of the process in PA graphs. Simulation and analysis both demonstrate the speediness of rumor spreading in social networks.

A key observation in the mathematical proof, as well as being a good explanation for this phenomenon, is that small-degree nodes learn a rumor once one of their neighbors knows it, then quickly forward it to their neighbors. This propagation scheme facilitates sending rumors from one large-degree node to another.

How does this play out in everyday life? It partially explains why social networks are observed to spread information quickly, even though the process is not organized centrally, and the network is not designed in an intelligent way. Crucial is fruitful interaction between hubs with many connections and average users with few friends. Hubs make the news available to a big audience, whereas average users quickly convey the information from one neighbor to the next.

A rumor begun at a random node of the Twitter network reaches on average 45.6 million of the total of 51.2 million members within only eight rounds of communication.

References

1. Albert, R., Jeong, H., and Barabási, A.-L. Diameter of the World-Wide Web. *Nature 401* (Sept. 1999), 130–131.
2. Barabási, A.-L. *Linked: How Everything Is Connected to Everything Else and What It Means.* Plume, 2003.
3. Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science 286* (Oct. 1999), 509–512.
4. Beaumont, P. The truth about Twitter, Facebook, and the uprisings in the Arab world. *The Guardian* (Feb. 25, 2011); http://www.guardian.co.uk/world/2011/feb/25/twitter-facebook-uprisings-arab-libya
5. Berger, N., Borgs, C., Chayes, J.T., and Saberi, A. On the spread of viruses on the Internet. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms* (Vancouver, B.C., 2005), 301–310.
6. Bhattacharjee, B., Druschel, P., Gummadi, K. et al. Online Social Networks Research @ The Max Planck Institute for Software Systems; http://socialnetworks.mpi-sws.org
7. Bohman, T. and Frieze, A.M. Hamilton cycles in 3-out. *Random Structures & Algorithms 35*, 4 (2009), 393–417.
8. Bollobás, B and Riordan, O. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics 1*, 1 (2003), 1–35.
9. Bollobás, B and Riordan, O. Coupling scale-free and classical random graphs. *Internet Mathematics 1*, 2 (2003), 215–225.
10. Bollobás, B and Riordan, O. The diameter of a scale-free random graph. *Combinatorica 24*, 1 (2004), 5–34.
11. Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. The degree sequence of a scale- free random graph process. *Random Structures & Algorithms 18*, 3 (2001), 279–290.
12. Boyd, S.P., Ghosh, A., Prabhakar, B., and Shah, D. Randomized gossip algorithms. *IEEE Transactions on Information Theory 52*, 6 (2006), 2508–2530.
13. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. Graph structure in the Web. *Computer Networks 33*, 1–6 (2000), 309–320.
14. Chierichetti, F., Lattanzi, S., and Panconesi, A. Rumor spreading in social networks. *Theoretical Computer Science 412*, 24 (2011), 2602–2610.
15. Cooper, C., Frieze, A.M., and Vera, J. Random deletion in a scale-free random graph process. *Internet Mathematics 1*, 4 (2004), 463–483.
16. Demers, A.J., Greene, D.H., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H.E., Swinehart, D.C., and Terry, D.B. Epidemic algorithms for replicated database maintenance. *Operating Systems Review 22*, 1 (1988), 8–32.
17. Doerr, B., Fouz, M., and Friedrich, T. Social networks spread rumors in sublogarithmic time. In *Proceedings of the 43rd ACM Symposium on Theory of Computing* (San Jose, CA). ACM Press, New York, 2011, 21–30.
18. Flaxman, A.D., Frieze, A.M., and Vera, J. Adversarial deletion in a scale-free random graph process. Combinatorics, *Probability & Computing 16*, 2 (2007), 261–270.
19. Fountoulakis, N., Huber, A., and Panagiotou, K. Reliable broadcasting in random networks and the effect of density. In *Proceedings of the IEEE Conference on Computer Communications* (San Diego, Mar. 15–19). IEEE Press, 2010, 2552–2560.
20. Karp, R., Schindelhauer, C., Shenker, S., and Vöcking, B. Randomized rumor spreading. In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science* (Redondo Beach, CA, Nov. 12–14). IEEE Computer Society Press, 2000, 565–574.
21. Kempe, D., Dobra, A., and Gehrke, J. Gossip-based computation of aggregate information. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science* (Cambridge, MA, Oct. 11–14). IEEE Computer Society Press, 2003, 482–491.
22. Kumar, S.R., Raghavan, P., Rajagopalan, S., and Tomkins, A. Extracting large-scale knowledge bases from the Web. In *Proceedings of the 25th International Conference on Very Large Data Bases* (Edinburgh, Scotland, Sept. 7–10). Morgan Kaufmann, 1999, 639–650.
23. Leskovec, J. Stanford Large Network Dataset Collection; http://snap.stanford.edu/data/
24. Mihail, M. Papadimitriou, C., and Saberi, A. On certain connectivity properties of the Internet topology. *Journal of Computer and System Sciences 72*, 2 (2006), 239–251.
25. Milgram, S. The small-world program. *Psychology Today 2* (1967), 60–67.

Benjamin Doerr (doerr@mpi-inf.mpg.de) is a senior researcher in the Department of Algorithms and Complexity at Max-Planck-Institut für Informatik, Saarbrücken, Germany, and an adjunct professor at Saarland University, Saarbrücken, Germany.

Mahmoud Fouz (mahmoudfouz@rocket-internet.com.tr) is a managing director of Rocket Internet, Dubai, U.A.E.; this work was done while he was a Ph.D. student in the Computational Complexity group of the Department of Computer Science at Saarland University, Saarbrücken, Germany.

Tobias Friedrich (t.friedrich@mpi-inf.mpg.de) is an independent research group leader in the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University, Saarbrücken, Germany, and a senior researcher in the Department of Algorithms and Complexity at Max-Planck-Institut für Informatik, Saarbrücken, Germany.

© 2012 ACM 0001-0782/12/06 $10.00