Why Rumors Spread Fast in Social Networks

Benjamin Doerr[1], Mahmoud Fouz[2], and Tobias Friedrich[1,2]

[1] Max-Planck-Institut für Informatik, Saarbrücken, Germany
[2] Universität des Saarlandes, Saarbrücken, Germany

Understanding structural and algorithmic properties of complex networks is an important task, not least because of the huge impact of the internet. Our focus is to analyze how news spreads in social networks. We simulate a simple information spreading process in different network topologies and demonstrate that news spreads much faster in existing social network topologies. We support this finding by analyzing information spreading in the mathematically defined preferential attachment network topology, which is a common model for real-world networks. We prove that here a sublogarithmic time suffices to spread a news to all nodes of the network. All previously studied network topologies need at least a logarithmic time. Surprisingly, we observe that nodes with few neighbors are crucial for the fast dissemination.

Social networks like Facebook and Twitter are reshaping the way people take collective actions. They have played a crucial role in the recent uprisings of the 'Arab Spring' and the 'London riots'. It has been argued that the 'instantaneous nature' of these networks influenced the speed at which the events were unfolding [4].

It is quite remarkable that social networks spread news so fast. Both the structure of social networks and the process that distributes the news are not designed with this purpose in mind. On the contrary, they are not designed at all, but have evolved in a random and decentralized manner.

So is our view correct that social networks ease the spread of information ("rumors"), and if so, what particular properties of social networks are the reason for this? To answer these questions, we simulate a simple rumor spreading process on several graphs having the structure of existing large social networks. We see, for example, that a rumor started at a random node of the Twitter network in average reaches 45.6 million of the total of 51.2 million members within only eight rounds of communication.

We also analyze this process on an abstract model of social networks, the so-called preferential attachment graphs introduced by Barabási and Albert [3]. In [17], we obtain a mathematical proof that rumors in such networks spread much faster than in many other network topologies—even faster than in networks having a communication link between any two nodes (*complete graphs*). As an explanation, we observe that nodes of small degree build a short-cut between those having large degree (*hubs*), which due to their large number of possible communication partners less often talk to each other directly.

Social Networks and Rumor Spreading

Social networks arise in a variety of contexts. They are formed by people, who are connected by knowing each other, *Facebook* members by agreeing on being friends (in Facebook), scientific authors by having a joint publication, or actors appearing in the same movie.

Despite this diversity, many networks share characteristic properties. Well known is the observation that any two individuals are connected through just "six degrees of separation", which was first formulated by Karinthy (see Barabási [2]) and became known to a broad audience through Milgram's "small world study" [25]. Similarly for the world wide web, Albert, Jeong, and Barabasi [1] predicted a *diameter* (maximum distance between two nodes in the graph) of only 19 in the network formed by web pages and links between them.

Several experimental studies [1, 13, 22] revealed another intrinsic property of social networks: the histogram of the 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 preferential attachment model for real-world networks that show a power-law. The model is widely used, also because of its simplicity. The paper [3] is currently the fifth most cited article in "Science" according to ISI Web of Knowledge. In the preferential attachment model, the graphs are constructed in a random, 'rich-get-richer' fashion: a newly entering node connects to m existing ones chosen randomly, but gives preference to nodes that are already popular, that is, have many neighbors. Note that the parameter m controls the density of the graph, i.e., the ratio of the number of present edges to the number of all possible edges. For these graphs, the authors of [3] empirically discovered a power-law of k^{-3}, which was later proven mathematically by Bollobás, Riordan, Spencer, and Tusnády [11]. A number of similar models emerged at the same time, all leading to a power-law distribution. It is known, though, that the PA model does not share all properties observed in the real-world networks, e.g., it is less clustered.

Still, the preferential attachment model has been successfully used to deduce many interesting properties of social networks. Famous structural results prove a small diameter of such graphs [10], determine their degree distribution [11], show high expansion properties [24], and a high robustness against random damage, but a vulnerability against malicious attacks [8, 9, 15, 18]. Algorithmic works show that in such networks, viruses spread more easily than in many other network topologies [5], or that gossip-based decentralized algorithms can approximate averages better [12].

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

We assume that nodes choose their communication partners uniformly at random from their neighbors, however, excluding the person they contacted just 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 works, the rumor spreading process is synchronized, that is, the process takes place at discrete points of time and in each time step, each node contacts a neighbor to exchange information. This is clearly a simplification over the real-world scenario where users act at different speeds, but we feel that in sufficiently large networks these differences balance out and thus assuming an average speed used by all nodes does not make a substantial difference.

We also note that the communication process is different in each social network. The push-pull model we regard naturally captures best a personal communication between two individuals as by phone or exchanging text messages, emails or other directed communications. Many online social networks allow also other ways of communication like posts on user's personal pages, possibly resulting and his friends to be notified of the post when they next log in, and them forwarding the news given that it is sufficiently interesting. Such forms of communication can be modelled only by more complicated mechanisms than the push-pull protocol.

Experimental Results

To support the common observation that news spreads very fast in social networks, we have simulated the rumor spreading process on samples of the *Twitter* and *Orkut* social networks (taken from [6, 23]) as well as on preferential attachment graphs. As most social networks have a roughly similar structure, we have chosen these large networks, for which data was readily available, as instances of social networks. For comparison, we have also included into our investigation complete graphs and *random-attachment graphs* (also called m-out model, see, e.g., Bohman and Frieze [7]), in which each node chooses m neighbors uniformly at random from all nodes.

Note that in both the random-attachment and the preferential attachment graph model, we can control the density of the graph via the parameter m. This allows us to run experiments on such graphs having equal number of nodes and density as the real-world graphs regarded.

Figures 2 and 3 show how a rumor started in a random node spreads in the graphs corresponding to the Orkut and Twitter social networks. Both figures show that news spreads much faster in the real-world networks and the preferential attachment graphs than in the complete and random-attachment graphs. For the Twitter experiment, a considerable difference between the preferential attachment model and the real-world graph is visible, indicating that the Twitter graph is captured less well by the theoretical model.

To determine how the graph size influences the rumor spreading speed, we ran the process on preferential attachment graphs, random-attachment graphs and complete graphs of varying size. The results depicted in Figure 4 indicate that a logarithmic time is needed both for random-attachment and complete graphs, whereas for preferential attachment graphs we observe times of a slightly smaller order of magnitude.
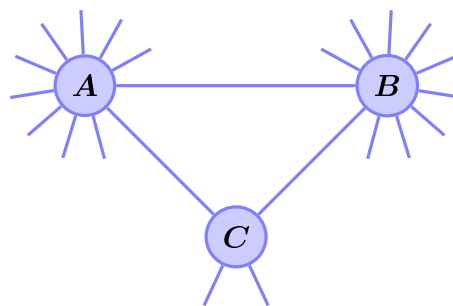
Mathematical Analysis

We support this empirical finding via a mathematical analysis, which proves that the rumor spreading process disseminates a news in sublogarithmic time. This refers to the time needed to inform all nodes as well as any constant fraction like 1%, 10% or 50%.

We denote by G_m^n the preferential attachment graph on n nodes with density parameter m. Since the graph G_m^n is a random graph, none of the statements mentioned previously holds with certainty. However, the probability that the random graph G_m^n does not satisfy the statements above, tends to zero for n growing to infinity. In the following, whenever a statement on a random object is made, it is meant to hold with high probability. For the preferential attachment graph G_m^n with m being any constant larger than one, we are able to prove that after a surprisingly short time a news spreads to all nodes.

Theorem 1. *There is a constant K such that the rumor spreading process described above 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 preferential attachment graphs [14], which is of order $\log(n)^2$. Theorem 1 for the first time shows that rumor spreading in preferential attachment graphs is much faster than for the other network topologies regarded so far. For random-attachment graphs, it is easy to see that the diameter is of order $\log(n)$, which is clearly also a lower bound for the rumor spreading time. A similar reasoning also shows that, independent of the starting node, a constant fraction of all nodes has distance $\Theta(\log(n))$ from the starting node. Hence a logarithmic number of rounds is already needed to inform any constant fraction of the nodes. Similar bounds follow for hypercube networks, which are common communication networks in computer science applications.

However, it is not always the diameter that tells us the time needed for rumor spreading. The complete graph obviously has a diameter of exactly one, but the time needed to spread a rumor from one node to all others is of logarithmic order. This result was proven by Karp et al. [20] for a rumor spreading process where nodes are allowed to choose their random communication partners among all neighbors, including the one they have just talked to. It is not difficult to see that their proof remains valid in our setting. Furthermore, it can be seen from their proof that a logarithmic number of rounds is still necessary to inform any constant fraction of the nodes. Similar results hold for the classical Erdős-Rényi random graphs as can be deduced from [19, 20]. All these results were proven by mathematical means, that is, they do not rely on experiments conducted for certain graph sizes n,

Figure 1: How a rumor progresses from a large-degree node A to another one B. Due to their large degrees, A and B need long to push or pull the rumor. C having a small degree however quickly pulls the rumor from A and again quickly pushes it to B.

but are valid for all graph sizes. Unfortunately, the proof of Theorem 1 is too long to be

presented here (it can be found in [17]). However, we will sketch one main argument, that also explains why rumor spreading in social networks is that fast.

To this aim, 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 around d_A rounds until A contacts B and thus A pushes the news to B. Similarly, it will take an expected number of around d_B time steps until B calls A and thus pulls the rumor from there. If d_A and d_B are large, e.g., $n^{1/3}$, then it would take an expected number of almost $\frac{1}{2}n^{1/3}$ rounds to propagate the rumor from A to B along the direct link.

Note that in Theorem 1 we stated a much smaller bound. Hence there must be a better way to get the news from A to B. Interestingly, it is the small-degree nodes that make the difference. Assume for the moment that there is a third node C that is a neighbor of both A and B, and that has a small degree d_C of say 4 only. Now after an expected number of roughly $d_C = 4$ rounds, C will have contacted A and thus learned the news from A. Similarly, after another expected number of about $d_C = 4$ rounds, C will have contacted B and told the news to B. So in expectation, in around $2 \cdot d_C = 8$ time steps, the rumor went from A to B via C. Fortunately, such a node C exists with high probability—the preferential attachment rule ensures that newly entering nodes put enough preference on connecting with A and B.

So one mechanism which allows fast rumor spreading in social networks is that small-degree nodes quickly learn the rumor from an informed neighbor and 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 in a short time (without doing anything). Such mechanisms are missing, e.g., in complete graphs, because all nodes have a high degree of $n - 1$. Consequently, all neighbors of the starting node A have a very small probability of calling A and asking for the news (A is just one of their $n - 1$ neighbors).

Let us add that such fast links are abundant in preferential attachment graphs. To ease the language, let us call a node *popular*, if it has a degree of $\Theta(\log(n)^2)$ or higher. We can 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 seen above, these nodes act as fast links, propagating the rumor in an expected number of roughly $2 \cdot m$ rounds. In consequence, the expected time the rumor needs to traverse the whole path is about m times its length. With some extra care, namely by showing that there is a huge number of such paths between any two popular nodes, we can even show that once any popular node is informed, then 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 faster, namely in $O(\log(n)^{3/4}\log(\log(n)))$ rounds. Once all popular nodes are informed, we can use a symmetric argument showing that after another $O(\log(n)^{3/4}\log(\log(n)))$ rounds, the remaining small-degree nodes, mostly by calling more popular nodes, all become informed. This finishes the proof sketch. For more details, we refer to [17].

Conclusion

We simulated a natural rumor spreading process on different graphs representing real-world social networks and several classical network topologies. We also performed a mathematical analysis of this process in preferential attachment graphs. Both works demonstrate that rumor spreading is extremely fast in social networks.

A key observation in the mathematical proof and a good explanation for this phenomenon is that small-degree nodes quickly learn a rumor once one of their neighbors knows it, and then again quickly forward the news to all their neighbors. This in particular facilitates sending a rumor from one large-degree node to another.

What does this mean for our everyday life? It partially explains why social networks are observed to spread information extremely rapidly even though this process is not organized centrally and the network is not designed in some intelligent way. Crucial is the fruitful interaction between hubs, which have many connections, and average users with few friends. The hubs make the news available to a broad audience, whereas average users quickly convey the information from one neighbor to another.

References

[1] R. Albert, H. Jeong, and A.-L. Barabasi. Diameter of the world-wide web. *Nature*, 401:130–131, 1999.

[2] A.-L. Barabási. *Linked: How Everything Is Connected to Everything Else and What It Means.* Plume, 2003.

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

[4] P. Beaumont. The truth about twitter, facebook and the uprisings in the arab world, Feb 2011. `www.guardian.co.uk/world/2011/feb/25/twitter-facebook-uprisings-arab-libya`.

[5] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. On the spread of viruses on the internet. In 16*th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 301–310, 2005.

[6] B. Bhattacharjee, P. Druschel, K. Gummadi, et al. Online social networks research at the Max Planck Institute for Software Systems. `socialnetworks.mpi-sws.org`.

[7] T. Bohman and A. M. Frieze. Hamilton cycles in 3-out. *Random Structures & Algorithms*, 35:393–417, 2009.

[8] B. Bollobás and O. Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1:1–35, 2003.

[9] B. Bollobás and O. Riordan. Coupling scale-free and classical random graphs. *Internet Mathematics*, 1:215–225, 2003.

[10] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24:5–34, 2004.

[11] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18:279–290, 2001.

[12] S. P. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52:2508–2530, 2006.

[13] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33:309–320, 2000.

[14] F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumor spreading in social networks. *Theoretical Computer Science*, 412:2602–2610, 2011.

[15] C. Cooper, A. M. Frieze, and J. Vera. Random deletion in a scale-free random graph process. *Internet Mathematics*, 1:463–483, 2004.

[16] A. J. Demers, D. H. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. E. Sturgis, D. C. Swinehart, and D. B. Terry. Epidemic algorithms for replicated database maintenance. *Operating Systems Review*, 22:8–32, 1988.

[17] B. Doerr, M. Fouz, and T. Friedrich. Social networks spread rumors in sublogarithmic time. In *43rd ACM Symposium on Theory of Computing (STOC)*, pages 21–30, 2011.

[18] A. D. Flaxman, A. M. Frieze, and J. Vera. Adversarial deletion in a scale-free random graph process. *Comb., Probab. & Comput.*, 16:261–270, 2007.

[19] N. Fountoulakis, A. Huber, and K. Panagiotou. Reliable broadcasting in random networks and the effect of density. In *IEEE INFOCOM*, pages 2552–2560, 2010.

[20] R. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking. Randomized rumor spreading. In *41st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 565–574, 2000.

[21] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *44th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 482–491, 2003.

[22] S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Extracting large-scale knowledge bases from the web. In *25th International Conference on Very Large Data Bases (VLDB)*, 1999.

[23] J. Leskovec. Stanford large network dataset collection. http://snap.stanford.edu/data/.

[24] M. Mihail, C. Papadimitriou, and A. Saberi. On certain connectivity properties of the internet topology. *J. Comput. Syst. Sci.*, 72:239–251, 2006.

[25] S. Milgram. The small-world program. *Psychology Today*, 2:60–67, 1967.
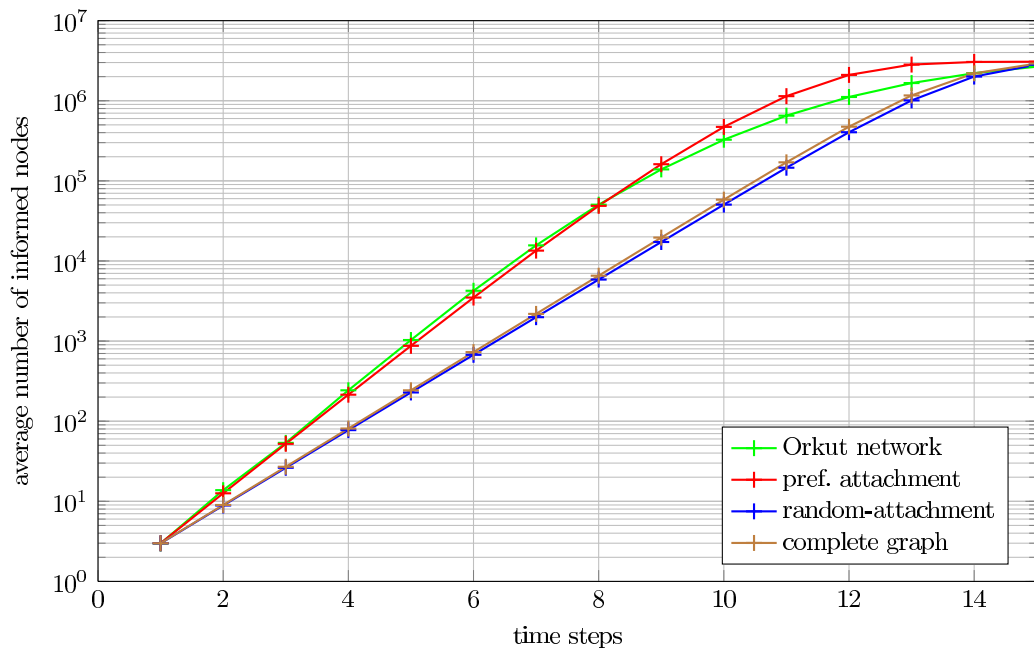
Figure 2: Average numbers of informed nodes over time for the Orkut network and preferential attachment, random-attachment and complete graphs of same size $n = 3,072,441$ and density parameter $m = 38$ (where applicable). Both the real network and the mathematical model for social networks show a significantly faster dissemination of the rumor.
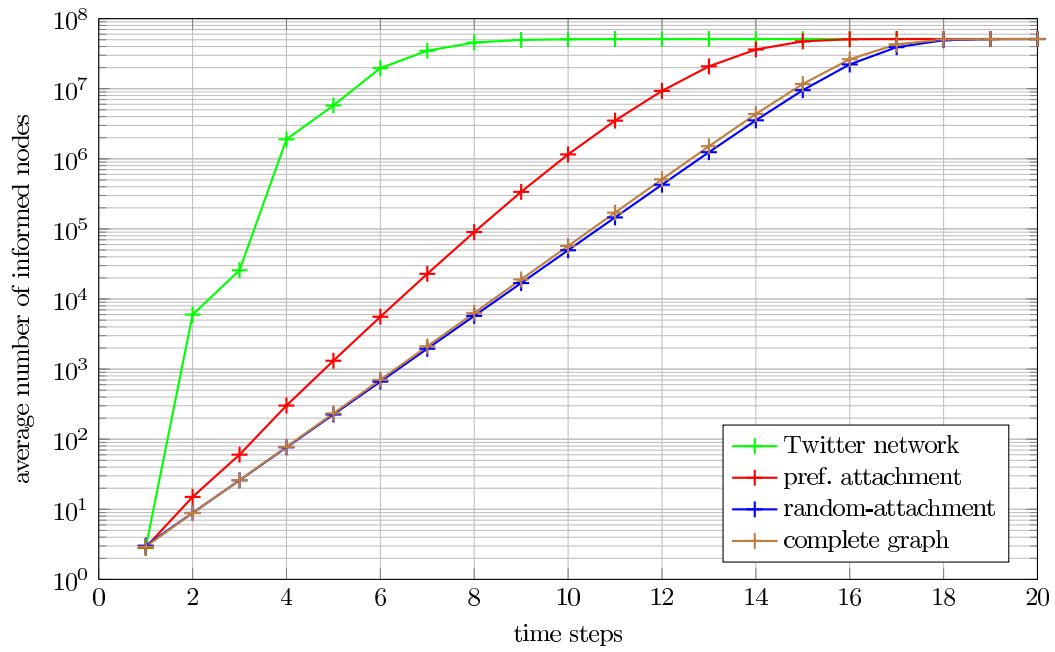
Figure 3: Average numbers of informed nodes over time for the Twitter network and comparable preferential attachment, random-attachment and complete graphs (size $n = 51,217,936$, density parameter $m = 32$). Again, the two graphs modelling social networks greatly outperform the random-attachment and complete graph.
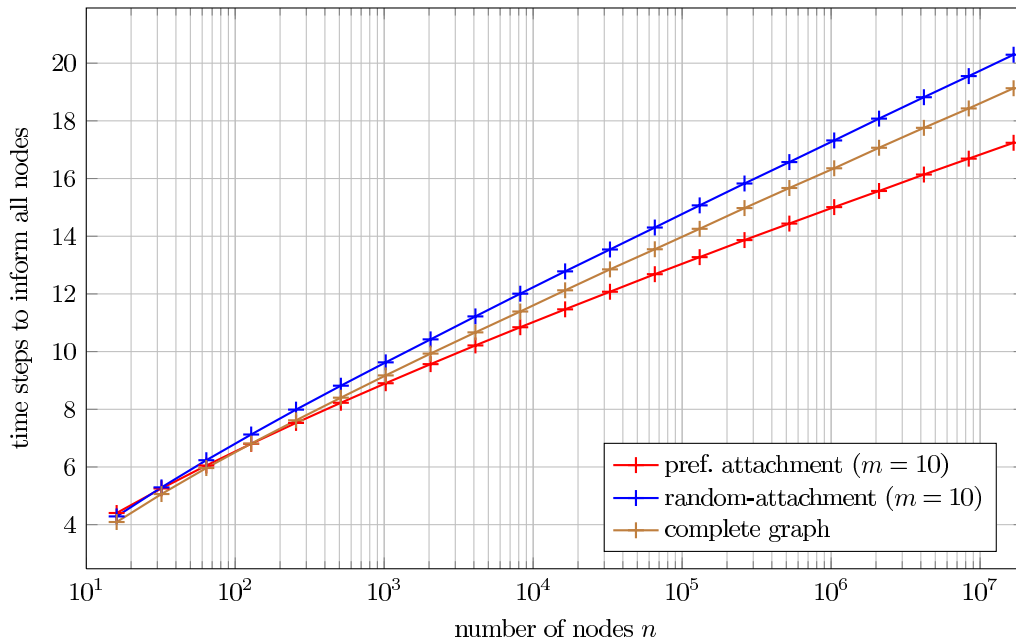
Figure 4: Average times needed to inform all nodes of different networks of varying size. The data show a logarithmic dependence for random-attachment and complete graphs. For preferential attachment graph (a mathematical model for social networks), the times appear to be of lower order. Note that the generally small times partially obscure the advantage of the preferential attachment network structure. The data above can also be read also as follows. If one is willing to spend an average number of 16 rounds, then one can inform more than 300,000 people connected as in the preferential attachment model, but only around 30,000 organized in a random-attachment fashion.