CS590 Assignment 2 – A Review of Contributed Paper in Communications of ACM

Why Rumors Spread So Quickly in Social Networks

by *Benjamin Doer, Mahmoud Fouz, and Tobias Friedrich*
CACM 55(6), 2012, 70-75

Shi Yin

This paper [5], as its title suggests, tried to find out what is the reason of fast spread of news and rumors in today's social networks. To achieve this, the authors firstly introduced an empirical finding that information spreads much faster in social networks than other topologies of a graph by building a model of information spreading process and then running it on different topologies. Then they briefly discussed about mathematical analysis of their experimental results, including proposing a potential reason for the fast flooding of information in social networks and in preferential attachment (PA) graphs, a model used to describe real social networks. They also claimed that it is the small-degree nodes that accelerate this spreading process.

The authors first validated the widely use of PA graph to describe a social network [2], by pointing out a commonly accepted conclusion that the number of nodes with *k* neighbors is inversely proportional to a polynomial in *k* [1, 3, 8]. They described the construction of a PA graph by adding new nodes based on certain rules, but did not bother to provide details of this process and of these rules, such as how the 'preferences' are given to popular nodes (a), why every node has the same number of *m* neighbors when first added into the graph (b), or how to build the initial state of the graph before new nodes can be added in using the rules the authors mentioned (c).

What interests me is that even in the original paper of Barabasi et al. [2], they did not give answers to (b) and (c). But they did describe how the preference was given, which makes flaw (a) of this paper more unacceptable.

The authors then made some assumptions in order to use push-pull principle to represent the spread process of information. One assumption was that nodes actively asked information from their neighbors, just like people in social networks frequently check recent updates of their friends on Facebook or people they are following on Twitter. This assumption fitted the real-world scenario pretty well, but the following two assumptions were not as good as this one. The authors assumed that a node chooses from its neighbors uniformly at random as its communication partner (d) and that all nodes communicate at the same speed (e).

The former assumption (flaw (d)) is only accurate in social networks like Facebook, where the relationship (friend-ship) between two 'friends' is mutual. In other scenarios such as Twitter, the relationship (follow-ship) is unidirectional. As for the later simplification (flaw (e)), the authors only claimed that it was acceptable in a sufficiently large network in which different acting speeds between nodes balance out each other and thus could be represented by an average speed used by all nodes, but they neither justified this statement nor provided reference where this claim was proved.

Thus, a better model that I can think of is to connect a new node to old nodes using unidirectional edges under ‘preferential attachment’ rule; and all nodes ask for information actively in each step to one of its neighbors in a random manner. A bold guess is this model would make the spread even faster, since this model decreased the number of nodes that a ‘popular’ node can receive information from.

Following, the authors showed their experimental results using three figures. The first two figures showed that rumors spread much faster in actual social networks or PA graphs than random attachment (RA) graphs and complete graphs. The big difference of rumor-spread speed in PA graph and Twitter social network also suggested that PA graph is not a very accurate representation of Twitter social network. The third figure showed how the time to inform all nodes in a graph increased as the size of the graph became larger. It clearly indicated not only that topology of PA graphs accelerate the spread of information compared to RA graphs and complete graphs but also that RA graphs and complete graphs need logarithmic steps to inform all their nodes about some piece of information.

The figures in this part of the article were very decent. The discussion was brief but compact, clear and complete. Although the authors did not give many details about their experiments, one can easily repeat it based on the information provided in previous discussion, except for how to set the preference of attaching to existing nodes for a new node (See flaw (a) in page 1).

At the last part of this article, the authors first briefly introduced many results of mathematical analysis, from previously proved arguments including the lower bound for information spread in RA graphs, that in complete graphs [7] and that in classical Erdos-Renyi random graphs [6, 7], to the authors’ own conclusion that only a sub-logarithmic time is needed for a piece of rumor to flood throughout the whole PA graph, rather than previous lower bound of order $\log(n)^2$, where n is the number of nodes in the PA graph [4].

All the processes of these mathematical proofs were not included in this paper, but the authors then talked pretty much about one of their arguments, which might also be the answer to the question in the title of this paper: it is the small-degree nodes that contribute a lot to the swiftness of the rumor-spread process in PA graphs, by bridging two large-degree nodes that are both its neighbors. This mechanism is missing in RA graphs and complete graphs.

This part could have been both interesting and convincing, however, the authors failed to include either brief explanation or to give reference in many mathematical derivation steps (f). One example is that in the next-to-last paragraph of page 74, they called ‘a node popular if it has a degree of $\Theta(\log(n)^2)$ or higher’, then immediately claim that they could ‘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’. They never gave more clues to help readers know how they could show this property. Similar examples in this part are numerous. If they could have mentioned more about their mathematical derivations in this part or at least had given us references that we can look into, as they did in previous sections, it would be a much better ending.

Last but not least, the paper is not well written with respect to language (g). Although this article is very easy to follow in general, it has a lot of grammatical errors, from “the singular or the plural” mistakes, “the trailing ‘s’ of third person singular” mistakes, to “the past tense or the present tense” mistakes. Most of the grammar issues are easy to be found by a simple run of any spelling/grammar checker software. For this rank of a magazine, it is an unacceptable defect to have so many problems in English grammar. Besides, maybe because the authors are from Germany, they used one particular structure of sentence a lot. One example of this structure is in the first paragraph of page 71: “Social networks spreading news so quickly is remarkable”. This is not a mistake in grammar, but in my experience, a native speaker would not write like this, instead, they would probably write this sentence like this: “It is remarkable that social networks spread news so quickly”.

In spite of the several flaws (from (a) to (g)) in this paper, the research itself was of high quality and of great importance in general. The full introduction of this research [4] (which included) was published in 2011, but already cited by 22 papers or books (including this paper). Currently, most of the papers that cited this paper are in the area of Theory of Computing. This paper will probably help other scientists to think of more advanced models to represent real social networks, or maybe directly used by sociologists, or maybe both, in the future.

[1] Albert, R., Jeong, H., and Barabási, A.-L. Diameter of the World-Wide Web. *Nature*, Vol. 401 (9), 1999, 130–131

[2] Barabasi, A.-L., and Albert, R. Emergence of scaling in random networks. *Science*, Vol. 286 (10), 1999, 509-512

[3] Broder, A., Kumar, R., Maghoul, F., raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. Graph structure in the Web. *Computer Networks*, Vol. 33 (1–6), 2000, 309–320

[4] 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.

[5] Doer, B., Fouz, M., and Friedrich, T. Why Rumors Spread So Quickly in Social Networks, *Communications of ACM*, Vol. 55 (6), 2012, 70-75

[6] 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

[7] Karp, R., Schindelhauer, C., Shenker, S., and Vocking, B. Randomized rumor spreading. In *Proceedings of the 41st IEEE Symposuum on Foundations of Computer Science* (Redondo Beach, CA, Nov. 12-14). IEEE Computer Society Press, 2000, 565-574

[8] Mumar, 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