News Posting by Strategic Users in a Social Network

Mangesh Gupte[1], MohammadTaghi Hajiaghayi[2], Lu Han[1], Liviu Iftode[1], Pravin Shankar[1], and Raluca M. Ursu[3]

[1] Department of Computer Science, Rutgers University, Piscataway, NJ 08854. {mangesh,luhan,iftode,spravin}@cs.rutgers.edu
[2] AT&T Labs – Research, Florham Park, NJ 07932. hajiagha@research.att.com
[3] Mount Holyoke College, South Hadley, MA 01075. ursu22r@mtholyoke.edu

Abstract. We argue that users in social networks are *strategic* in how they post and propagate information. We propose two models — greedy and courteous — and study information propagation both analytically and through simulations. For a suitable random graph model of a social network, we prove that news propagation follows a threshold phenomenon, hence, "high-quality" information provably spreads throughout the network assuming users are "greedy". Starting from a sample of the Twitter graph, we show through simulations that the threshold phenomenon is exhibited by both the greedy and courteous user models.

1 Introduction

Online social networks have become an increasingly popular medium for sharing information such as links, news and multimedia among users. The average Facebook user has 120 friends, and more than 30 million users update their status at least once each day [2]. More than 5 billion minutes are spent on Facebook each day (worldwide). As a direct consequence of these trends, social networks are fast overtaking traditional web as the preferred source of information [1].

In the early days of social networks, users tended to post predominantly personal information. Such information typically did not spread more than one hop, since only immediate friends were interested in it. Over time, online social networks have metamorphosed into a forum where people post information such as news that they deem to be of common interest. For example, during the recent Iran elections, traditional news media acknowledged the power and influence of social networks such as Twitter [3, 4].

Prior work has studied various aspects of information sharing on social networks. Domingos and Richardson studied the question of how to determine the set of nodes in a network that will most efficiently spread a piece of information for marketing purposes [5, 6]. Kempe, Kleinberg and Tardos proposed a discrete optimization formulation for this. Several recent studies focused on gathering intuition about influence spread from real-world data [9]. In [13], Leskovec, Singh and Kleinberg studied the patterns of cascading recommendations in social networks by looking at how individuals recommend the products they buy in an online-retailer recommendation system. In [10], Leskovec, Backstorm and Kleinberg developed a framework for the dynamics of news propagation across the web. Morris studied games where each player interacts with a small set of neighbors [16]. He proved conditions under which the behavior adopted by a small set of users will spread to a large fraction of the network.

An aspect that has been overlooked so far is to understand *why* users post information such as news or links on social networks. In this paper, we posit that users in a social network have transitioned from being passive entities to *strategic users* who weigh in various factors (such as how interested their friends will be in the news) to decide whether to post. This trend leads to several interesting questions, such as: What factors do users consider when deciding whether to post an item? How does information diffuse over the social network based on user strategies.

Our main result states that, assuming strategic users, the spread of news over an online social network exhibits a *threshold behavior*: The news spreads to a significant fraction of the network if its "quality" is larger than a certain threshold that depends on how aggressive users are about posting news. If the quality is smaller than this threshold, only a sub-linear number of nodes in the the network post the news.

The key contributions of this paper are:

1. We initiate the study of information propagation in social networks assuming strategic users.
2. We propose two models for strategic user behavior, greedy and courteous.
3. Assuming social networks can be modeled as certain random graphs, we prove that there is a threshold behavior when greedy users fully disseminate information.
4. We present a simulation study based on a real graph crawled from the Twitter social network, and show that the threshold phenomenon holds in both strategic models of user behavior.

In what follows, we provide a detailed description of our results. We start by defining the user model.

2 Strategic User Model

We propose a simple game to model the behavior of users posting news on social networks like Twitter and Facebook. For a particular user u in the network, whenever u first sees a previously unseen news item, she has the option of either posting it or not posting it. Her utility is 0 if she does not post it. If she does, then her utility depends on (i) The set $I_u = \{$Neighbors who are interested in the news$\}$ and (ii) The set $S_u = \{$Neighbors who, u knows, have already seen the news before$\}$. [4] Let N_u denote the set of u's neighbors. We propose two particular forms for her utility :

[4] u might not know the true set of neighbors who have seen the news. She knows that a friend has seen the news only if a mutual friend posted it. This also means that we assume that every user knows which of her friends are themselves friends.

Greedy Strategy: The utility is additive and for every neighbor who likes the news (irrespective of whether the neighbor has seen it before or not), she gets utility $+a$ and for every neighbor who does not, she gets utility $-b$. In this case, her decision to post only depends on a,b and $f_u = \frac{|I_u|}{|N_u|}$. User u posts only if her utility is positive, that is, the *fraction* f of users who like the news satisfies $a \cdot f_u - b(1 - f_u) > 0 \iff f_u > \frac{b}{a+b}$. Let us define $t = \frac{b}{a+b}$. In Section 4, we analyze this behavior and show that it depends critically on t.

Courteous Strategy: The main difference from the greedy strategy is that the user does not want to spam her friends. We model this by saying that if more than a c fraction of her friends have already seen the news before, she gets a large negative utility, when she posts the item. If a user does not post an item, then her utility is 0. In case the fraction $\frac{|S_u|}{|N_u|} \leq c$, then her utility is the same as in the greedy case. In particular, she gets utility $+a$ for every neighbor who likes the news and has not seen it before (the set $I_u \setminus S_u$), and she gets utility $-b$ for every neighbor who does not like it (the set $I_u^c \setminus S_u$). Hence, her strategy in this case is to post if the fraction of neighbors who have seen the news $\frac{|S_u|}{|N_u|} \leq c$ and if the fraction $\left(f_u = \frac{|I_u \setminus S_u|}{|N_u \setminus S_u|}\right)$ of neighbors in S_u^c who are interested in the news is $\geq t$. Note that, in this utility function, if a larger number of the user's neighbors have posted the news, she is *less* likely to post it. In section 5, we show simulation results for this behavior on a small sample of the Twitter Graph.

3 Notation and Preliminaries

Given a real symmetric matrix P with $0 \leq p_{i,j} \leq 1$, denote by $G(n, P)$ a random graph where edge (i, j) exists with probability $p_{i,j}$. For notational convenience we shall denote by $W_P = (V, E_P)$ the deterministic weighted graph with V as the vertex set and P as its adjacency matrix. Note that $p_{i,j}$ gives the weight of edge (i, j) in W_P.

Leskovec et al show that social graphs, such as Autonomous Systems on the Internet, the citation graph for high energy physics from arXiv, and U.S. Patent citation database, can be modeled well using Stochastic Kronecker Graphs [11]. $G(n, P)$ is a generalization of Stochastic Kronecker Graphs [11, 14] as well as the Erdõs and Rènyi model of Random Graphs $G(n, p)$ [7], and the model of random graphs with a given degree sequence [15]. In the following subsection, we prove properties of the graph $G(n, P)$ that will be used in the next section to analyze when a certain news spreads across the network.

3.1 Properties of G(n,P)

Definition 1 (Density of a cut). *Given a graph $G = (V, E)$, define the density of the cut $(S, V - S)$ as $\frac{|E(S,V-S)|}{|S||V-S|}$ where $E(S, V - S)$ denotes the set of edges between S and $V - S$. The partition $(S, V - S)$ that minimizes this density is called the* Sparsest Cut *in the graph.*

Definition 2 (α-balanced cut). *Given a graph $G = (V, E)$ and a cut $(S, V - S)$, the cut is α-balanced if and only if* $\min\{|S|, |V - S|\} \geq \alpha|V|$.

Hence, by a sparsest α-balanced cut we mean the cut with the minimum density over all cuts that are α-balanced. We start with a lemma from Mahdian and Xu [14] that proves when G is connected.

Lemma 3. *If the size of the min-cut in W_P is $> c \log n$, then with high probability, the sampled graph $G \sim G(n, P)$ is connected.*[5]

Given $U \subseteq V$, we shall denote by $G[U]$ the subgraph induced by U. The induced subgraph of W_P is denoted by $W_P[U]$. The following lemma gives a sufficient condition on the existence of a giant component in the graph $G(n, P)$.

Lemma 4. *If there exists $U \subseteq V$, of size $\Theta(n)$, such that the sparsest α-balanced cut of $W_P[U]$ has density $> \frac{c}{|U|}$, then there is a giant connected component in $G[U]$ of size $\Theta(n)$ with high probability.*

4 Analysis of a Model for Strategic User Behavior

We analyze the *greedy* strategy defined in Section 2 when it is played over a random graph $G(n, P) = (V, E)$. Our results also apply to Stochastic Kronecker Graphs.[6] According to this model, a user posts only if the fraction of interested neighbors is $> t$. We assume that for a given news, each user in the network likes it with probability q, which is independent of everything else. Probability q could model the quality of the news item or the inherent interest the subject generates.[7] Throughout, we assume that q and t are constants that do not depend on the number of nodes n.

We color nodes in G yellow if they are interested in the news and blue if they are not. A yellow node is *responsive* if more than a *t fraction* of its neighbors are interested in the news. Color responsive nodes red. We denote these sets by Y, B and R, respectively. Note that $R \subseteq Y$. $G[R]$ is the graph induced by the red vertices. We are interested in the structure of the graph $G[R]$. We prove the following results in the next two subsections :

Proposition 5. *Suppose $G(n, P)$ is such that the min-cut in W_P is of weight $\geq c \log n$ and that $\log n$ random nodes in the network initially see the news. If $q > t$, with high probability, almost all nodes interested in the news will post it. On the other hand, if $q < t$ then only a sub-linear number of the vertices will post the news.*

Next, we give a condition on the sparsity of G, which gives a weaker result on the spread of the news.

---

[5] Throughout this paper, with high probability means with probability $1 - o(1)$.

[6] Due to space constraints, we leave the formal statement of the result and the proof of this claim to the Technical Report [8].

[7] It is an interesting question to relax this assumption, since in a typical social network we might expect nodes with similar interests to be clustered together.

Proposition 6. *Suppose $G(n, P)$ is such that the sparsest α-balanced cut in W_P has density $\geq c/n$ and that $\log n$ random nodes in the network initially see the news. If $q > t$ then, with high probability, a constant fraction of the nodes interested in it will post it.*

4.1 The Connectedness of $G[R]$

We prove Proposition 5 in this subsection. *Throughout this section, we shall assume that $G \sim G(n, P)$ where P is such that the min-cut of W_P has weight $\geq c \log n$, for a large enough constant c.* We start by looking at what happens to the min-cut in the sampled graph.

Lemma 7. *With high probability, the min-cut of the subgraph $G[Y]$ of G induced by the yellow vertices has size $> c' \log n$ for some constant c'.*

Now, we shall prove the main theorem of this section. We prove that $G[R]$ is connected by using the fact that its min-cut is large.

Theorem 8. *If $q > t$ and $G \sim G(n, P)$ where P is such that the min-cut of W_P has weight $\geq c \log n$, then, with high probability, every vertex in Y also belongs to R and so $G[R]$ is connected. When $q < t$, then, with high probability $G[R]$ only contains $o(n)$ vertices.*

Proof. Let Y_q be a random variable that takes value 1 with probability q and 0 otherwise. For a node v, let $d(v)$ denote its degree in G.

Case 1 : $q > t$: $Pr\left[v \notin R\right] = Pr\left[\sum_{i=1}^{d(v)} Y_q < qd(v) - (q-t)d(v)\right]$ $\leq \exp\left(-\frac{((q-t)d(v))^2}{2 \cdot q \cdot d(v)}\right) \leq n^{-\frac{(q-t)^2 c}{2 \cdot q}}$. This follows from Chernoff Bounds and the facts that Y_q are independent, $\mathbb{E}[Y_q] = q$ and $d(v) \geq c \log n$. We apply the union bound to get that with probability $\geq 1 - n^{1-\frac{(q-t)^2 c}{2 \cdot q}} \geq 1 - 1/n$ (when $c > \frac{2q}{(q-t)^2}$) all nodes in Y also belong to R and, from Lemmas 3 and 7, we get that $G[R]$ is connected.

Case 2 : $q < t$: $Pr\left[v \in R\right] = Pr\left[\sum_{i=1}^{d(v)} Y_q \geq qd(v) + (t-q)d(v)\right]$ $\leq \exp\left(-\frac{((t-q)d(v))^2}{2(qd(v)+(t-q)/3)}\right) \leq n^{-\frac{(t-q)^2 c}{t}}$. Hence, $\mathbb{E}[|R|] \leq |Y| \cdot n^{-\frac{(t-q)^2 c}{t}} \leq o(n)$. So, R only contains a sub-linear number of nodes of G.

Proof (of Proposition 5). If $q < t$, then the proposition follows directly from Theorem 8. When $q > t$, Theorem 8 tells us that $G[R]$ is connected. If any node in R receives the news, it will be propagated to all the nodes. However, the probability that none of the nodes in R get the news is $\leq (1 - \frac{|R|}{n})^{\log n} = O(n^{-q}) = o(1)$. Hence, with high probability, almost all the nodes interested in the news post it.

In the next subsection, we shall identify conditions on the distribution from which G is sampled, which are enough to show when a constant fraction of the nodes who are interested in the news actually receive it.

4.2 Existence of a giant component in $G[R]$

We now prove Proposition 6. *In this section, we shall assume that* $G \sim G(n, P)$, *where* P *is such that the sparsest* α*-balanced cut of* W_P *has density* $\geq c/n$. We prove that the size of the sparsest-cut in $G[Y]$ is not small.

Lemma 9. *Consider the subgraph* $G[Y]$ *induced by the yellow vertices. With high probability,* $G[Y]$ *has a subgraph of size* $\Theta(n)$ *whose sparsest* α*-balanced cut has density* $\geq \frac{c'}{|Y|} \geq \frac{c'}{n}$, *for some constant* c'.

Theorem 10. *Let* G *be a random graph sampled from the distribution* $G(n, P)$ *where the density of the sparsest* α*-balanced cut in the graph* W_P *is greater than* c/n. *If* $q > t$, *then every yellow node is red with probability* $> 1 - \epsilon_c$. *Further, the induced graph* $G[R]$ *has a giant connected component of size* $\Theta(n)$ *with high probability.*

Proof. Let Y_q be a random variable that takes value 1 with probability q and 0 otherwise. Let us denote the degree of node v by $d(v)$. Since $(v, V \setminus v)$ is a cut, $\frac{d(v)}{1 \cdot (n-1)} \geq \frac{c}{n} \implies d(v) \geq c$.

For $v \in Y, Pr\left[v \notin R\right] = Pr\left[\sum_{i=1}^{d(v)} Y_q < qd(v) - (q-t)d(v)\right]$ $\leq \exp\left(-\frac{((q-t)d(v))^2}{2 \cdot q \cdot d(v)}\right) \leq e^{-\frac{(q-t)^2 c}{2 \cdot q}} = \epsilon_c$. Hence, a constant fraction $f \geq 1 - \epsilon_c$ of the vertices in Y belong to R. From this and from Lemmas 4 and 9, we can prove that $G[R]$ also has a giant connected component with high probability.

Proof (of Proposition 6). By Theorem 10, $G[R]$ contains a giant component C of size $(1 - \epsilon_c)n$. If any node in C receives the news, it will be propagated to all the nodes in C. However, the probability that none of the nodes in C get the news is $\leq (1 - \epsilon_c)^{\log n} = o(1)$. Hence, a constant fraction of the nodes interested in the news actually receive it with high probability.

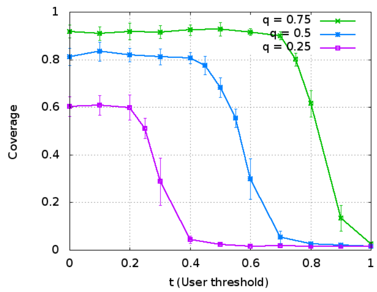
5 Simulation Results

In this section, we present results from the simulation of our two strategic user models over a partial crawl of the Twitter social network.[8]

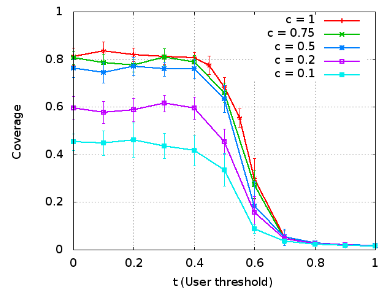
The dataset is obtained by means of a partial crawl of the Twitter social network. Since we are interested in link postings by typical users, we remove hubs (users with more than 500 friends) from our graph. The resulting graph we use in our experiments consists of 5978 users with 110152 directed friendship edges. Each simulation starts with a set of seed nodes (of size $\log n$), and in each round, nodes decide whether to post the link using one of the two models described earlier in the paper.

We define coverage as the fraction of interested users who get to see the news item. Figure 1 plots how the coverage varies for the *greedy* and *courteous*

[8] All the data was obtained in accordance with Twitter's terms of use.



(a) ***Greedy*** **strategy: The coverage exhibits a step behavior when** $q = t$**.**



(b) ***Courteous*** **strategy. (q = 0.5). The coverage decreases logarithmically as the user is more courteous.**

Fig. 1. Coverage of *greedy* and *courteous* strategies. q **is the probability with which each user likes the link.** t **and** c **are the thresholds for the greedy and the courteous strategies, respectively.**

strategies. For the courteous strategy, we fixed $q = 0.5$. Each simulation was repeated 10 times, and the standard deviations are indicated in the form of error bars.

Greedy Strategy: Figure 1(a) shows that, for all values of q, the coverage exhibits a step behavior and the step happens around $q = t$, which is consistent with Theorem 10. For different values of q, the percentage of coverage decreases with q. This is true even when $t = 0$, which means that the size of the connected components drops when we sample the graph. The Densification Power Law ($E \alpha N^k$) [12] that has been observed in other social networks would predict this behavior.

Courteous Strategy: Figure 1(b) shows the effect of the courteous strategy on the coverage. The parameter c indicates the threshold of neighbors that have already seen the link. A courteous user posts the link only if less than c fraction of neighbors have already seen the link. The figure shows that the coverage decreases logarithmically as the user is more courteous. This means that even when the users are courteous, if $q > t$ then, news can still reach a reasonable fraction of the graph.

6 Conclusion and Future Work

We proposed the model of strategic user behavior in online social networks for news posting, defined two user models (greedy and courteous), presented formal analysis of the greedy model and simulated both models on the data set we collected from Twitter. We propose the following directions:

Mine the Twitter data set: Search for patterns in the way users post news links in order to validate the model and provide further insights about user strategies over social networks.

Analyse the Courteous Strategy and Multiple Strategies: We leave open a formal proof that the courteous strategy also exhibits threshold behavior. Further, we want to test whether similar results hold in a social network in the presence of multiple user strategies.

Design a framework for advertisement on online social networks of strategic users: We believe that the strategic user model has applications in advertising and marketing. We plan to investigate incentive schemes for marketeers to encourage strategic users to advertise products over social networks.

Acknowledgments: We would like to thank S. Muthu Muthukrishnan for the inspiring discussions. We would also like to thank the anonymous reviewers. This work has been supported in part by the NSF grant CNS-0831268 and by the Summer 2009 DIMACS/DyDAn REU Program.

References

[1] Compete — Site Comparison. `http://siteanalytics.compete.com/facebook.com+google.com/?metric=uv`.

[2] Facebook Statistics. `http://www.facebook.com/press/info.php?statistics`.

[3] Journalism Rules Are Bent in News Coverage From Iran. `http://www.nytimes.com/2009/06/29/business/media/29coverage.html`.

[4] Twitter Revolution in Iran. `http://money.cnn.com/video/news/2009/06/22/n_twitter_revolution_iran_pentagon.cnnmoney/`.

[5] P. Domingos and M. Richardson. Mining the network value of customers. *Proc. 7th ACM SIGKDD*, 2001.

[6] P. Domingos and M. Richardson. Mining knowledge-sharing sites for viral marketing. *Proc. 8th ACM SIGKDD*, 2002.

[7] P. Erdos and A. Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5, 1960.

[8] M. Gupte, M. Hajiaghayi, L. Han, L. Iftode, P. Shankar, and R. M. Ursu. News posting by strategic users in a social network. Technical Report DCS-tr-657, Rutgers University, Department of Computer Science, October 2009.

[9] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. *Proc. 9th ACM SIGKDD*, 2003.

[10] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proc. 15th ACM SIGKDD*, 2009.

[11] J. Leskovec and C. Faloutsos. Scalable modeling of real graphs using kronecker multiplication. In *International Conference on Machine Learning (ICML)*, 2007.

[12] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. *Proc. 11th ACM SIGKDD*, 2005.

[13] J. Leskovec, A. Singh, and J. Kleinberg. Patterns of influence in a recommendation network. *Proc. PAKDD*, 2006.

[14] M. Mahdian and Y. Xu. Stochastic kronecker graphs. In *Proceedings of the 5th Workshop on Algorithms and Models for the Web-Graph*, 2007.

[15] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6, 1995.

[16] S. Morris. Contagion. *Review of Economic Studies*, 67, 2000.