Collaboration in Social Network-based Information Dissemination

Aziz Mohaisen* Tamer AbuHmed† Ting Zhu‡ Manar Mohaisen§
University of Minnesota - Twin Cities, MN, USA* Inha University, Republic of Korea†
Binghamton University, NY, USA‡ Korea University of Technology & Education, South Korea§

***Abstract*—Connectivity and trust within social networks have been exploited to build applications on top of these networks, including information dissemination, Sybil defenses, and anonymous communication systems. In these networks, and for such applications, connectivity ensures good performance of applications while trust is assumed to always hold, so as collaboration and good behavior are always guaranteed. In this paper, we study the impact of differential behavior of users on performance in typical social network-based information dissemination applications. We classify users into either collaborative or rational (probabilistically collaborative) and study the impact of this classification and the associated behavior of users on the performance on such applications.By experimenting with real-world social network traces, we make several interesting observations. First, we show that some of the existing social graphs have high routing costs, demonstrating poor structure that prevents their use in such applications. Second, we study the factors that make probabilistically collaborative nodes important for the performance of the routing protocol within the entire network and demonstrate that the importance of these nodes stems from their topological features rather than their percentage of all the nodes within the network.**

keywords: Social networks, collaboration, routing, adversarial behavior, performance.

I. Introduction

The popularity of social networks have motivated a wide spectrum of new technologies: designs, protocols, and applications built based-on and atop these networks. These include random-walk based routing [1], [4], [6], [15], [16], shortest-path based routing [23], [10], Sybil defenses [28], [29], among other technologies.

While these systems serve different purposes and follow different operational models, all of these schemes strike a balance among their algorithmic properties, connectivity, trust, and collaboration within the underlying social networks, all of which are utilized for bootstrapping such systems. Collaboration is an essential feature of social networks; however, the assumptions underlying collaboration are usually made to support end-results: all nodes are assumed to be collaborative. To address this issue, we study the impact of classifying users into collaborative and rational on such algorithm of information dissemination on top of social networks.

The primary contribution of this work is as follows. First, we suggest a classification of users in social network-based systems on collaborative and non-collaborative (rational) and suggest the rationale of such classification. Second, we study the impact of such classification on information dissemination on top of social networks. We consider several routing algorithms, based on random walks, shortest path, and breadth-first search (BFS). By experimenting with real-world social network traces, our study unveils interesting results. First, regardless of the level of collaboration of the nodes, some social networks initially exhibit poor performance—in the sense that the number of steps required to reach a particular source in the social graph, at average, is large. In our analysis, we study the factors that impact the performance of the routing application on top of these networks and demonstrate that such performance does not merely depend on the percentage of rational nodes, but rather on the topological properties of these rational nodes (high degree, betweenness, etc). Unlike previous studies concluded, such topological factors of these nodes and their implications are essential and critical to the design and the performance.

The rest of this paper is organized as follows. Section II introduces terminologies and preliminaries contained and referenced throughout the paper. Section III introduces the model for classifying users in the network based on their collaboration. Section IV introduces our results for random walk-based routing on real-world social network traces. Related work is discussed in section V and section VI concludes the paper.

II. Preliminaries

In this section we outline preliminaries, used as ingredients in building the rest of this work. We outline the model of the network, and routing algorithms experimented with in the paper.

A. Network model

We represent the social network as an undirected and unweighted graph $G = (V, E)$, where $V = \{v_1, \ldots, v_n\}$ is the set of vertices, representing the nodes within the social graph, and $E = \{e_{ij}\}$ (where $1 \leq i \leq n$ and $1 \leq j \leq n$) is the set of edges connecting those vertices. $|V| = n$ denotes the size of G and $|E| = m$ denotes the number of edges in G. $A = [a_{ij}]_{n \times n}$ represents the adjacency matrix of G, where $a_{ij} = 1$ iff $v_i \sim v_j$ (adjacent) and 0 otherwise. In the rest of this paper, social network and graph are used interchangeably to refer to both the physical network and the underlying social graph.

B. Random walk-based routing

Random walk theory provides a straightforward framework for routing implementation in many networks, including wireless networks. In its simplest form, the random walk based routing uses transition matrix P associated with the social graph to randomly select forwarders at each node until the destination is reached. Recall A defined above, then $P = [p_{ij}]^{n \times n}$, where $p_{ij} = 1/\deg(v_i)$ iff $v_i \sim v_j$ and 0 otherwise is defined for the "simple random walk" which we use in this work.

Let v_s be the source and v_d be the destination to which a packet is intended. v_s uniformly at random selects one of her neighbors—say v_x in Figure 1(a), and forwards the packet towards her. At each time slot, the intermediate node on the random path between the source and the destination checks if the destination is among its neighbors. If so, the intermediate node directly forwards the packet to the destination. Otherwise, the intermediate node performs the same procedure by uniformly selecting one of her neighbors and forwarding the walk towards that neighbor.

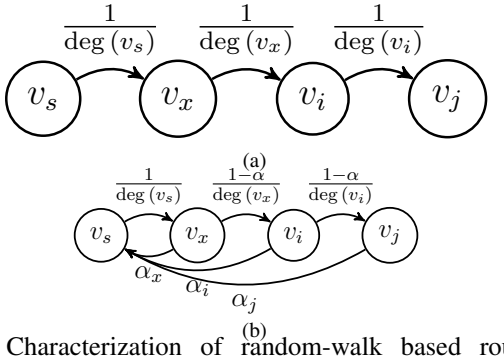

Fig. 1: Characterization of random-walk based routing. (a) Random-walk based routing on graph in normal fully-collaborative settings. (b) Random-walk based routing on graph in mixed settings, where users are classified into collaborative and non-collaborative users.

C. Shortest path based routing

The shortest path based routing uses the shortest directed distance between two nodes, the source and destination. Let $w : E \rightarrow \mathbb{R}$ be a weight function that assigns real-valued weights to edges in G— in case of undirected unweighted graphs, the weight will be equal to 1 across all edges. The weight of path $p = \langle v_1, v_2, \ldots, v_\ell \rangle$ is $w(p) = \sum_{r=1}^{\ell} w(v_{r-1}, v_r)$. The shortest path between nodes v_i and v_j is then defined as $\delta(v_i, v_j) = \min\{w(p) : v_i \overset{p}{\rightsquigarrow} v_j\}$ if there is a path from v_i to v_j, or $\delta(v_i, v_j) = \infty$ otherwise. Since it may not be unique, a shortest path between v_i and v_j is any path with weight $w(p) = \delta(v_i, v_j)$. Dijkstra's algorithm is an example of the shortest path based routing algorithm which is used in the Open Shortest Path First (OSPF) routing protocol. Dijkstra's algorithm mainly finds the shortest-path between nodes v_i and v_j depending on non-negative weights assigned to the edges of G, which is the case in our study.

In this context, and only for experiments, the weights of the edges are calculated using the Jaccard similarity coefficient, where the weight of the edge will be a reflection of the similarity between its vertices. The main motivation of using such weights is our interest in measuring how the topological structure of the graph influences its behavior. Any other meaningful weights can be used to replace the weights we used in our experiments, and to bring similar insights. Examples could be interactions, reputation-based, or cost-based weights.

D. Breadth-first search routing

Given a source and a destination pair of nodes v_i and $v_j \in G$, the BFS algorithm starts discovering all the reachable vertices from the root node v_i till it reaches the destination node v_j. The algorithm discovers all the vertices at distance k from v_i before discovering any nodes at distance $k + 1$. Recalling the definition of the aforementioned shortest path, BFS shortest path $\delta(v_i, v_j)$ in the case of unweighted graph is equal to the path with the smallest number of edges between v_i and v_j [3]. The BFS procedure uses first-in first-out (FIFO) queue Q data structure to manage the nodes to be traversed during the search process. This data structure can be replaced by a stack data structure to turn this algorithm into the depth-first search routing algorithm. Therefore, given a source node v_s and a destination node v_d, our scenario is to find the shortest needed path in order to deliver the transmitted packet to the destination.

III. Collaboration in Social Networks

In this section, we classify users in the social graphs into collaborative users and probabilistically collaborative users and study the impact of this classification on the performance of routing algorithms on top of social networks. Collaborative nodes are denoted by V_c and the level of altruism is denoted by γ.

A. Probabilistically collaborative users

These users, who are otherwise referred to as rational users, act less altruistically than collaborative users. In particular, while the altruism (i.e. γ) of collaborative users is close to one, and hence the selfishness characterized by $\alpha = 0$, the altruism of rational users is characterized by α where $0 < \alpha < 1$. Typically, a rational node may participate in the routing protocol with probability $1 - \alpha$ and deviate from it by dropping incoming packets or not collaborating with probability α. Such nodes in G are denoted by V_p whose size (as a fraction of the size of the network) is β. Note that $\beta + \gamma = 1$. We note that each node in the graph can only belong to one of the categories above—hence V_p and V_c are exclusive subsets of V where $V = V_p \cup V_c$. Second, we assume that a node may not change its behavior over the run time of the protocol, to simplify the analysis. Last, α may differ from a node to another, as we will see later.

B. Collaboration Impacts Information Dissemination Behavior

The behavior—and its description—of the different nodes in the graph is shown as a state diagram in Figure 1(b). The routing protocol is then described as a *biased* random walk where the event of not collaborating in the routing protocol is denoted by a loop from each node to the originator of the algorithm. For simplicity, in the figure, we remove collaborative nodes, which can be seen as nodes across the route with $\alpha = 0$.

As for the shortest-path and BFS-based routing, lack of collaboration results in shortest-path search failure. Accordingly, similar to above, we consider the percent of shortest-path (and BFS-based search) routing trials that fail among all possible trials in the graph among possible source-destination pairs.

IV. Results and Discussion

In this section, we introduce the results of this study and elaborate on the findings. The social graphs used in this study are shown in Table I. Some of these graphs are sampled from larger graphs using the breadth-first search—details on these graphs are in [22]. As an indicator of the topological structure of the different graphs, we compute both the diameter and radius of each graph. By defining the eccentricity as the set of maximal shortest paths from each and every source to other destinations in the graph, the diameter is defined as the maximal eccentricity

TABLE I: Social graphs with their size, diameter, and radius. Physics 1, 2, 3 are relativity, high-energy, and high-energy theory co-authorship respectively [12]. **D** stands for the diameter and **R** stands for the radius of the graph.

Social network	Nodes	Edges	D	R
Physics 1 [12]	4,158	13,428	17	9
Physics 2 [12]	11,204	117,649	13	7
Physics 3 [12]	8,638	24,827	18	10
Wiki-vote [11]	7,066	100,736	7	4
Enron [12]	10,000	108,373	4	2
DBLP [13]	10,000	20,684	8	4
Facebook [26]	10,000	81,460	4	2
Youtube [17]	10,000	58,362	4	2

and the radius is defined as the minimal eccentricity. We observe that these parameters differ greatly from a graph to another which implies different graph structures.

Evaluation metric. The evaluation metric of the different schemes used in this study is the normalized expected number of transmissions per single message delivery operation between a source and a destination. The definition slightly differs based on the algorithm in use, and is formally defined as $E[cost] = \frac{1}{S}\sum_{i=1}^{S} cost_i$, where S is the size of the sample for which the cost is computed, and $cost_i$ is the cost for a given pair of source and destination indexed by i.

A. Performance in ideal settings

Here we study the performance of routing on social graphs in ideal settings, without considering collaboration as a constraint.

1) Random walk-based routing on social graphs: Considering the different graphs in Table I, we first measure the performance of the simple random walk-based routing explained in section II. We define the cost of routing over graphs as the normalized expected number of transmissions, which is the average number of times that a node in the graph transmits a packet for a single routing session from a given source to a given destination. To avoid the random behavior and bias in the measurements, we consider the case of routing a single packet from a given source (selected uniformly at random from the graph) to $1,000$ arbitrary destinations, which are also selected uniformly at random from the graph. To reduce the bias in the measurements, we perform the same experiment, for the same source-destination pair, for $1,000$ times; totaling $1,000,000$ routing trials per data set. In addition, we take the average number of hops, which is then normalized by the network size, to give the expected number of transmissions. The results of these measurements are shown in Figure 2. While they are close in size, we observe that the cost of routing over the different graphs is basically different, and this is related to the structure of the underlying graph. In principle, the performance of the graphs can be classified into two categories: well-performing graphs (Epinion, Youtube, Wiki-vote, and Facebook) and poorly performing graphs (Physics-1 to 3 and DBLP). We observe that the poor-performing graphs exhibit a strong community structure, as evidenced by their high modularity—a measure of the community structure in social networks. On the other hand, well-performing graphs have less clear community structure evidenced by their small modularity. Furthermore, the poor performance is associated with larger radius and diameter of the graph contrary to well-performing ones

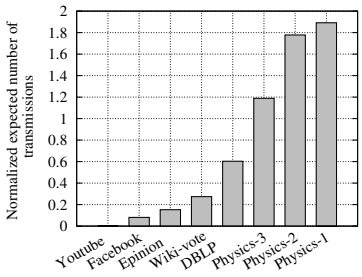

Fig. 2: Random-walk based routing on graph in normal fully-collaborative settings—all nodes act according to the protocol and behave honestly. Different social graphs have different structures and different qualities of the performance of the routing protocol.

(shown in Table I) which are associated with smaller radius and diameter values.

2) Shortest path based routing on social graphs: Following the evaluation scenario of the Random walk-based routing, we evaluated the expected number of transmissions between two randomly selected source and destination nodes in the case of BFS and Dijkstra routing algorithms. In Figures 4 and 5, we show the CDF of the number of transmissions per node in several social graphs using BFS and Dijkstra routing. These graphs illustrate the relation between the number of transmissions, the structure and connectivity of the graph. These results coincide with our finding regarding to the random walk-based routing in social graphs, where the cost of routing over the different graphs is strongly related to the structure and connectivity of the underlying graph and loosely affected by the applied routing algorithm.

B. Collaboration in random walk-based routing

We measure the performance of the routing protocol, with the same settings as above, when considering probabilistically collaborative nodes in the graph. We uniformly at random sample subsets of the nodes $V_p \subset V$ as the set of probabilistically collaborative users, with the remaining nodes in the graph as totally collaborative (altruistic). As explained earlier, each probabilistic node v_i follows the protocol with probability α_i, which is uniformly selected in the range of 0.1 to 1, or drop the routing request with probability $(1-\alpha_i)$ (we trim the distribution from 0 to 0.1 for that the existence of very adversarial nodes may block traffic entirely due to the lack of multi-path). The results of the performance of the protocol on the different social graphs are shown in Figure 3. By considering β as the percent of rational users, we consider different values of β (i.e., from 0 to 0.8 with 0.2 steps) where $\beta = 0$ in each graph represents the performance of the corresponding social graph in Figure 2. In brief, we make the following observations on the different experiments.

While the performance of the different social graphs initially differs greatly, as evidenced by the first experiment, the impact of the increasing percent of rational nodes is not linear but rather depends on the underlying social graph. For example, social graphs with strong community structure have rather fairly regular behavior (Figure 3(a) to Figure 3(d)). However, we observe that even with these social graphs, relatively large β dramatically increases the cost of routing. To understand this behavior, we list the rational nodes, and map them to their degrees. We observe that, while in the first case—where less impact is made on the performance in random walk-based routing due to lack of collaborative of some nodes—the degree is fairly distributed, in

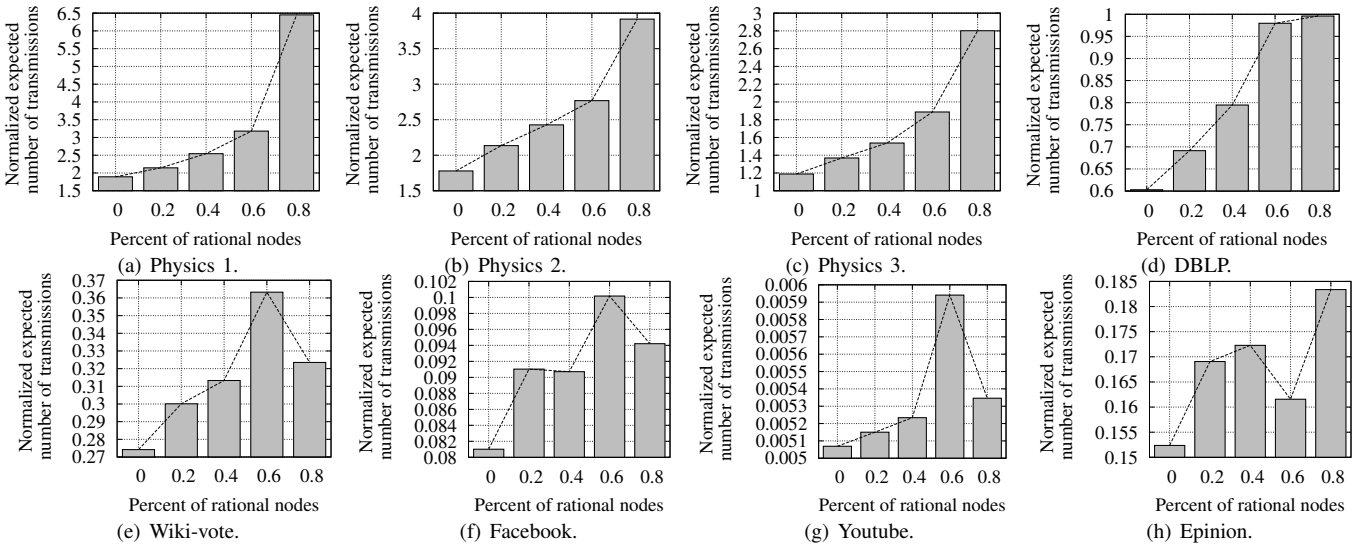

Fig. 3: The normalized expected number of transmissions per node in each social graph when using random walk-based routing.

the second case—where higher impact is observed—some high degree nodes, with small α blocks flows between communities, and thus dramatically increases the cost of routing. In such graphs, the cost is exponential in β.

On the other hand, the behavior of the initially well-performing graphs is in part harder to anticipate. In general, the routing protocol performs well, even when considering larger values of β in these graphs, though, like the previous case of poorly performing graphs it may have some odd behavior when high degree, intra-communities nodes behave rationally, with small α.

The difference between the two sets is that while the well-performing graphs are sensitive and any node can be of importance to the random routing on them, as evidenced by many high degree nodes, yet, the performance on such graphs is reasonable, and within the theoretically acceptable bounds.

C. Collaboration in shortest path-based routing

Considering the existence of probabilistically collaborative nodes in the graphs, we measured the performance of the shortest path based routing when a random sample subset of the nodes $V_p \subset V$ are probabilistically collaborative and the remaining are collaborative. As in the case of random walk-based routing, the results of the performance of the protocol with underlying shortest path based routing algorithm on the different social graphs are shown in Figures 6 and 7. However, due to the deterministic nature of the shortest path based routing, we consider the delivery failure rather than the cost of routing as the evaluation criterion, for different β values. The results show that, in terms of packet delivery rate, the well-performing graphs (Epinion, Youtube, Wiki-vote, and Facebook) are less sensitive to the existence of the rational users compared to the poorly performing graphs (Physics-1 to 3 and DBLP), which are affected by the rational users.

One implication of these findings is that, since the collaboration of high-degree nodes is more important to the overall performance of the network, investigating providing incentives for such nodes to be always collaborative to improve the performance is an interesting issue. For example, in DTNs' routing built on social networks [14], [2], [4] it is assumed that all nodes are collaborative. Since it is not always the case, it will be interesting to deploy these observations and build incentives for collaboration, especially for those critical nodes, in that context.

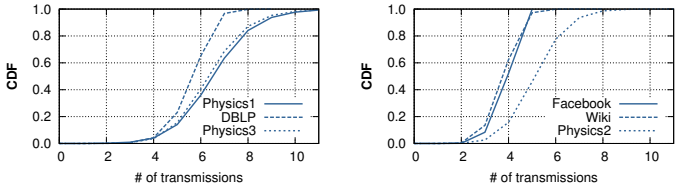

Fig. 4: The CDF of the number of transmissions per node in different social graph using BFS routing.

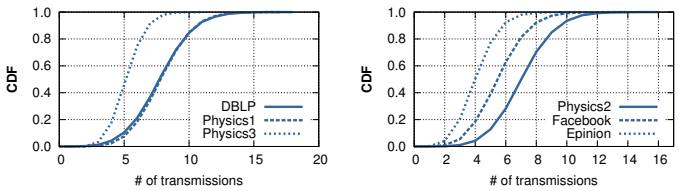

Fig. 5: The CDF of the number of transmissions per node in different social graph using Dijkstra routing.

V. Related Work

There has been a number of papers on the use of social networks for building communication and security systems, studying the performance of such designs on top of social networks, and analyzing the assumptions used in these designs as well. The most close to this study is the work in [7], where nodes are basically assumed to have some selfish behavior in each and every one of them, which follows some distribution (e.g., uniform, normal, or geometric). The major difference between our work and the work in [7] is actually twofold. First, while [7] considers traces of encounter-like wireless networks, we consider traces of static social graphs. While in the general sense both types of traces, static and encounter-based, could be of potential use to routing, we believe that static traces are more favorable to the assumption of trust which most routing protocols weigh a big value on to demonstrate effectiveness [19]. Second, and more important, the conclusions in this paper are at contradiction with the findings in [7]—most likely due to the different types of graphs used. In particular, whereas it is shown in [7] that selfishness does not affect the behavior of the routing algorithm much due to the

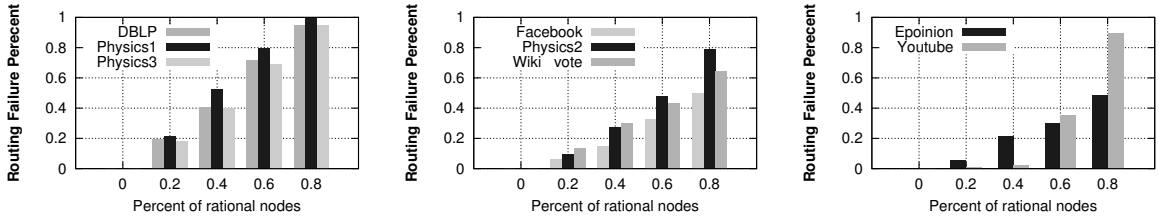

Fig. 6: BFS failure rate, expressed as the routing failure percent per β value (rational users) in each social graph.

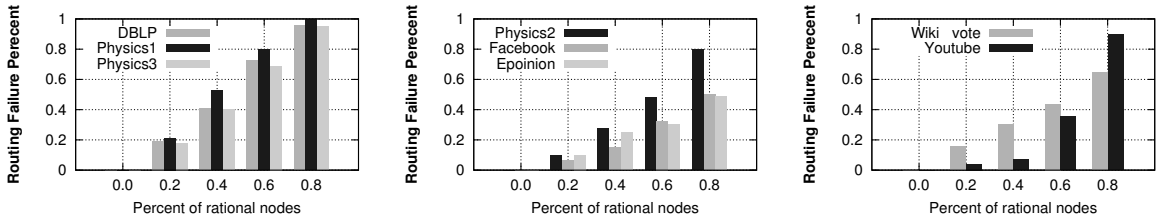

Fig. 7: Dijkstra's failure rate, expressed as the routing failure percent per β value (rational users) in each social graph.

multi-path characteristics of the underlying connections and links among nodes in the graph, we show strong evidence that the lack of collaboration by a few nodes with a particular characteristic (e.g., degree distribution) in static social graphs could greatly affect the effectiveness of the routing protocol built on top of the social network (see details in section IV). In total, our work, while brings conclusions contradicting with the prior work in [7], can be considered an effort in the same direction to understanding collaboration in settings where social networks are used for improving routing in networked systems.

Other systems built on top of social networks include works on Sybil defenses in [18], [5], [25] and understanding the assumptions used in these defenses in [27], [19], [22], [21]. Distributed computing services on top of social networks have been recently studied in [20], [24]. Routing on small world networks is initially investigated in [9], and delay of routing on such networks is most recently rigorously studied in [8].

VI. Conclusion and Future Work

In this paper, by classifying nodes in social graphs into collaborative and probabilistically collaborative, we studied the impact of collaboration in social networks on the performance of information dissemination techniques, including random walk based routing, shortest-path based routings, all of which are built on top of social networks. Even without the classification part, we experimentally demonstrated that the cost of such protocols on top of some of real-world graph is large while it is reasonable on others. We further show that some networks are very well performing and meet the potential of such applications, whereas other networks are quite sensitive to the users' behavior.

Exploring theoretical models to characterize the performance of the routing algorithms under behavior of users expressed as parameters would be the work to be considered in the near future. This will benefit from recent works, e.g., [8].

***Acknowledgement*—**This work was supported in part by Binghamton University academic program and faculty development fund. A. Mohaisen is supported in part by a doctoral dissertation fellowship from UMN and M. Mohaisen is supported by research subsidy for newly-appointed professor at Korea University of Technology and Education for the period 2010-2011

References

[1] G. Bigwood and T. Henderson. Social dtn routing. In *CoNEXT '08: Proceedings of the 2008 ACM CoNEXT Conference*, pages 1–2. ACM, 2008.

[2] R. J. Clark, E. Zasoski, J. Olson, M. H. Ammar, and E. W. Zegura. D-book: a mobile social networking application for delay tolerant networks. In *Challenged Networks*, pages 113–116, 2008.

[3] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001.

[4] E. M. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *MobiHoc*. ACM, 2007.

[5] G. Danezis and P. Mittal. Sybilinfer: Detecting sybil nodes using social networks. In *NDSS*. The Internet Society, 2009.

[6] J. Davitz, J. Yu, S. Basu, D. Gutelius, and A. Harris. ilink: search and routing in social networks. In *KDD*. ACM, 2007.

[7] P. Hui, K. Xu, V. Li, J. Crowcroft, V. Latora, and P. Lio. Selfishness, altruism and message spreading in mobile social networks. In *NetSciCom*, 2009.

[8] H. Inaltekin, M. Chiang, and H. Poor. Delay of social search on small world graphs. *Journal of Mathematical Sociology*, 2012.

[9] J. Kleinberg. Navigation in a small world. *Nature*, 406(6798), 2000.

[10] S. Kwon and N. B. Shroff. Analysis of shortest path routing for large multi-hop wireless networks. *IEEE/ACM Trans. Netw.*, 17(3):857–869, 2009.

[11] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Predicting positive and negative links in online social networks. In *WWW*, pages 641–650, 2010.

[12] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *KDD*. ACM, 2005.

[13] M. Ley. The DBLP computer science bibliography: Evolution, research issues, perspectives. In *String Processing and Information Retrieval*, 2009.

[14] Q. Li, S. Zhu, and G. Cao. Routing in socially selfish delay tolerant networks. In *INFOCOM*, 2010.

[15] I. Mabrouki, X. Lagrange, and G. Froc. Random walk based routing protocol for wireless sensor networks. In *ValueTools '07*, pages 1–10. ICST, 2007.

[16] S. Marti, P. Ganesan, and H. Garcia-Molina. Dht routing using social links. In *IPTPS*, pages 100–111, 2004.

[17] A. Mislove, M. Marcon, P. K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *IMC*, 2007.

[18] P. Mittal, M. Caesar, and N. Borisov. X-Vine: Secure and pseudonymous routing using social networks. In *Proc. of NDSS*, 2012.

[19] A. Mohaisen, N. Hopper, and Y. Kim. Keep your friends close: Incorporating trust into social network-based sybil defenses. In *INFOCOM*. IEEE, 2011.

[20] A. Mohaisen, H. Tran, A. Chandra, and Y. Kim. Socialcloud: Using social networks for building computing services. Technical report, UMN, 2011.

[21] A. Mohaisen, H. Tran, N. Hopper, and Y. Kim. Understanding social networks properties for trustworthy computing. In *ICDCS Workshops*, 2011.

[22] A. Mohaisen, A. Yun, and Y. Kim. Measuring the mixing time of social graphs. In *IMC*, pages 383–389. ACM, 2010.

[23] T. Zhu, S. Xiao, P. Yi, D. Towsley, and W. Gong. A Secure Energy Routing Protocol for Sharing Renewable Energy in Smart Microgrid. In *IEEE SmartGridComm*, 2011.

[24] D. Tran, F. Chiang, and J. Li. Friendstore: cooperative online backup using trusted nodes. In *Proc. of SNS*, pages 37–42, 2008.

[25] N. Tran, B. Min, J. Li, and L. Subramanian. Sybil-resilient online content voting. In *NSDI*, 2009.

[26] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *EuroSys*, 2009.

[27] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai. Uncovering social network sybils in the wild. In *IMC*, 2011.

[28] H. Yu, P. B. Gibbons, and M. Kaminsky. Toward an optimal social network defense against sybil attacks. In *PODC*, pages 376–377. ACM, 2007.

[29] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: defending against sybil attacks via social networks. In *SIGCOMM*, 2006.