MINT: Maximizing Information Propagation in Predictable Delay-Tolerant Network

Shaojie Tang
Department of Computer and Information Sciences
Temple University
Philadelphia PA 19122
shaojie.tang@temple.edu

Jing Yuan
Department of Computer Science
University of Illinois at Chicago
Chicago IL
csyuanjing@gmail.com

Xiang-Yang Li
Department of Computer Science
Illinois Institute of Technology
Chicago IL
xli@cs.iit.edu

Yu Wang
Department of Computer Science
UNC at Charlotte
Charlotte NC 28223
yu.wang@uncc.edu

Cheng Wang
Department of Computer Science
Tongji University
Shanghai China
3chengwang@gmail.com

Xuefeng Liu
Department of Computing
Hong Kong Polytechnic University
Hong Kong China
csxfliu@gmail.com

ABSTRACT

Information propagation in delay tolerant networks (DTN) is difficult due to the lack of continues connectivity. Most of previous work put their focus on the information propagation in static network. In this work, we examine two closely related problems on information propagation in predicable DTN. In particular, we assume that during a certain time period, the interacting process among nodes is known a priori or can be predicted. The first problem is to select a set of initial source nodes, subject to budget constraint, in order to maximize the total weight of nodes that receive the information at the final stage. This problem is well-known influence maximization problem which has been extensively studied for static networks. The second problem we want to study is minimum cost initial set problem, in this problem, we aim to select a set of source nodes with minimum cost such that all the other nodes can receive the information *with high probability*. We conduct extensive experiments using $10,000$ users from real contact trace.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication, Network topology; G.2.2 [**Graph Theory**]: Network problems, Graph algorithms

Keywords

DTN, propagation, temporal, submodular.

1. INTRODUCTION

Information propagation in delay tolerant networks (DTN) is difficult due to the lack of continues connectivity and unstable network structure. Such dynamics are often ignored in most of previous works *e.g.*, most researches put their focus on the information propagation in static network [1] [2]. One simple way to model the time dependent DTN is to represent it using a static graph, nodes in the graph represent individuals, where we add an edge between any two nodes if and only if they have been interacted or met during one period. We may further assign certain weight to each edge to reflect their interacting frequency, *e.g.*, if two nodes meet each other m times among n time slots, we will add an edge between them with a probability (weight) m/n. Information propagation processes in such static networks have already been well studied [1]. However, this kind of static view fails to capture the network dynamics, that is shown to be very critical to the information propagation problem. For example, assume that there are three nodes v_a, v_b and v_c in DTN. Node v_a interacts with node v_b at the first timestep, and v_b meets v_c at the next timestep. When taking a static view of this dynamic network, we may conclude that v_a and v_c has the same importance since they meet v_b with the same frequency. It turns out that the final results will be totally different by choosing different one as the initial source node. In particular, if we select v_a as the source node, both v_b and v_c have chance to receive the information. In contrast, if we pick v_c as the source node, neither v_a nor v_b has chance to get the information.

In this work, we study two closely related problems on information propagation in predictable delay tolerant network (DTN). The first problem is weighted influence maximization problem, we assume that we are given a fixed set of nodes, and the interactions that happen among nodes over a period of time T can be known a priori or can be predicted. The objective is to select a subset of nodes as initial source nodes such that the total weight of nodes that can successfully receive the information within a period is maximized. As shown in [1] [3], find an optimal solution is NP-Hard in most existing models. We also study the minimum cost initial set problem. Given a predictable DTN, we aim to select a set of initial source nodes with minimum total cost such that *all* the nodes can receive the information with high probability.

The rest of the paper is organized as follows. Section 2 briefly review the related results. System model and problem formulation are presented in Section 3. We study the weighted influence maximization problem in Section 4 and the minimum cost initial set

MobiHoc'13, July 29–August 1, 2013, Bangalore, India.

problem in Section 5. We report extensive experimental results in Section 6 and conclude the paper in Section 7.

2. RELATED WORK

In many applications, data propagation in DTN is needed. Propagation protocols (including multicast and broadcast protocols) allow sending data packets from a source to multiple receivers. They are more effective, for data dissemination and multi-party communications, than unicast routing protocols. Chaintreau *et al.* [4] and Chierichetti *et al.* [5] both studied simple gossip-based protocols (including geographic and social gossip) to propagate information in mobile social networks. Zhou *et al.* [6] proposed new semantic multicast models for DTN multicast and developed several multicast routing algorithms with different routing strategies. Recently, Gao *et al.* [7] studied multicast in DTNs from the social network perspective. Unweighted influence maximization in dynamic networks is studied in [8]. Similar problem is also studied in [9].

3. SYSTEM MODEL

3.1 Network Model

In this work, we represent the time-evolving DTN by a sequence of static networks. Each static network represents the contact information among users at that timeslot. See Fig. 1 (a) (b) for illustration. Here we highlight two important observations which can be found in many kinds of social based DTNs (*e.g.*, cell phone network, bus-based disruption-tolerant network) [10] [11]: 1) nodes usually interact with each other in a regular way instead of moving randomly; 2) the interaction occurs among nodes can be predicted with sufficient history information. We call these networks predictable DTN, and we mainly study the information propagation problem on predictable DTN in this paper. Throughout this paper, we assume that the contact information of the dynamic network within a certain period can be pre-obtained through certain existed prediction method or statistical analysis.

3.2 Information Propagation Model

We adopt *Independent Cascade Model* [1] [8] [12] [13] in this work. *Independent Cascade Model* describes a spreading process comprising of two sets of individuals, source node and non-source node. In a predictable DTN, a source node v tries to send the information to each of its current neighbors with success probability $p_{(vw)}$. If v succeeds in propagating information to w at timestep t, then w will act as new source node in the following timesteps. If v fails to transmit information to w at timestep t, v will try again in the subsequent timesteps as long as they meet each other. This process continues until timestep T, let $\mathrm{ProgSet}(S)$ denote the set of nodes which received the information until T, corresponding to the set of initial source nodes S. Let v^t denote node v at timestep t, and let V^t denote all nodes at timestep t.

3.3 Discussion on Network Modeling

Remember that in our problem formulation, we assume that the temporal contact pattern is known deterministically in advance. One may challenge the practicability of this assumption *e.g.* the current and future network topologies may not be predicted *accurately*. Actually, it will not affect the generality of our results even we take the "uncertainty" associated with the contact pattern into consideration. One way to resolve this issue is to regard the "uncertainty" as part of the success probability. For example, assume that the probability that nodes v_1 and v_2 contact each other during t and $t+1$ is $\mathbf{p}$, and the success probability in the original problem formulation is $p_{(v_1^t, v_2^{t+1})}$. Then we can define a new success probability as $\mathbf{p} \times p_{(v_1^t, v_2^{t+1})}$. By introducing the new success probability, we can ensure the feasibility of our results even under uncertain contact pattern. We may also formulate it as a robust optimization or stochastic optimization formulation which is left as possible future works.

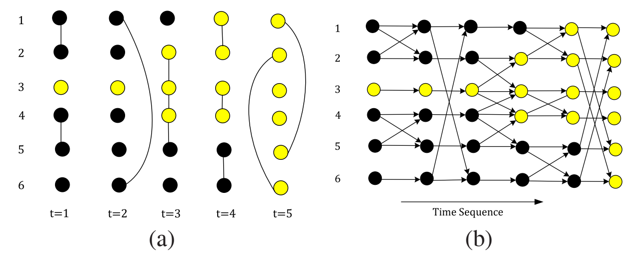

Figure 1: (a) illustrates the information propagation happens at each timestep by selecting node 3 as the initial source node; (b) describes the equivalent processing under the new constructed graph G. As shown in (a), if we choose node 3 as the initial source node, all of the nodes will receive the information (become light) at last. Equivalently, all nodes in V^T can be reached by node 3 in G.

4. WEIGHTED INFLUENCE MAXIMIZATION PROBLEM

We first study the weighted influence maximization problem. We assume that each node is associated with certain cost at which it can be selected as initial source node. We are interested at finding a set of initial source nodes, A, under the budget constraint such that the total weight of the propagated set is maximized.

Let $w(S)$ denote the the expected weight of the final propagated set under the initial source nodes S, we can prove that $w(S)$ is submodular monotone and non-negative. Our problem can be tackled by a simple greedy algorithm. We first compute two candidate sets for A: The first candidate set $A_{[1]}$ contains a single node which can maximize the expected total weight; the second candidate set $A_{[2]}$ is computed by Hill Climbing Algorithm in which we always add the node v that can maximize the expected incremental marginal gain: $w(A_{[2]} \cup \{v\}) - w(A_{[2]})/c(v)$ until the budget constraint is violated. Then we choose the better one as A. Based on the submodularity of $w(S)$ and similar proof in [14], we can prove that this algorithm achieves $\frac{1}{2} \cdot (1 - \frac{1}{e} - \epsilon)$ approximation.

5. MINIMUM COST INITIAL SET PROBLEM

In the previous sections, we mainly study how to select a set of initial source nodes under limited budget such that the final propagation range is maximized. We next study a symmetric problem: Minimum Cost Initial Set Problem. In particular, given a predictable DTN, each node is associated with certain cost, we aim to select a set of initial source nodes with minimum total cost such that *all* the nodes can receive the information at last. We study this problem under both deterministic model and probabilistic model. Note that when we study the minimum cost initial set problem under the probabilistic model, we are interested at finding the set of initial source nodes in order to let all the nodes receive the information *with high probability*.

Deterministic Model: Under the deterministic model, we assume that the success probability on each link is 1. Therefore, each node is associated with a determined propagated set. Naturally, our problem can be converted to a standard weighted set cover problem where the ground set is composed of the nodes in V_T, and each node in V_0 acts as a subset which can cover the nodes in its corresponding propagated set, and the classic greedy algorithm is a $\ln n$-approximate algorithm.

Probabilistic Model: We also studied this problem under the probabilistic model. In the probabilistic model, the success probability on each link is some value in [0,1], we aim to select the minimum cost initial source nodes such that the information can be propagated to the whole network with high probability (*e.g.*, the probability is lower bounded by certain value). We propose several heuristics to tackle this problem. In some scenarios, the hill climbing greedy algorithm can achieve constant approximation ratio.

6. EXPERIMENT RESULTS

In this section, we conduct extensive experiments to evaluate the performance of our algorithm.

The Network Data: A typical predictable DTN can be found in university campuses, *e.g.*, the National University of Singapore (NUS) student contact trace model [15]. We download the schedules of the 4,885 classes and enrollment of $22,341$ students for each week of 77 class hours.

Information Propagation Model: We conduct extensive experiments under three basic information propagation models. In the first model, which is called *uniformly setting*, we assigned a uniform success probability p to each edge of the graph, and for simplicity we choose $p = 10\%$ in results reported here. In the second model, which is called *randomly setting* we assign p to each edge randomly at uniform, choosing from 0% to 100%. In the third model, which is called *deterministic setting*, we assigned the propagation probability as 1 to each edge of the graph.

The Algorithms and Implementation: For the weighted influence maximization (WIM) problem, we implement a greedy algorithm (HC Selection) that selects the initial source nodes one by one such that the expected incremental marginal gain is maximized, until the budget is violated. For the Minimum Cost Initial Set (MCIS) problem, we modify the greedy algorithm to select the nodes one by one until the influential probability for each node is met. We compare our greedy algorithm with the following heuristics: 1). *Randomized Selection:* for the WIM problem, we randomly select the initial source nodes one by one until the budget constraint is violated; for the MCIS problem, we randomly select the nodes one by one until the influential probability for each node is met. We run the experiment 1000 times and choose the average weight as final results; 2). *Top-k Selection*: We first order all the nodes in the decreasing order of $|\operatorname{ProgSet}(v)|$. For the WIM problem, we pick the nodes one by one in order until the budget is violated; and for the MCIS case, we select the nodes in order until the influential probability constraint is met. 3). *Minimum Cost Selection:* This method is evaluated only for the WIM problem. We select the nodes one by one in the increasing order of their cost until the budget is violated.

The Results: First we present the results of our algorithms under the WIM problem settings. In this set of experiments, we compare all above algorithms under weighted probabilistic model with budget constraint. The network size is fixed to $10,000$ if not other specified.

Fig. 2(a) and Fig. 2(b) demonstrate the comparison results under the weighted probabilistic model, with $p = 10\%$ and random $p \in [0, 1]$ respectively. As shown in Fig. 2(a), when the budget is relatively small, *e.g.* less than 100, there is no huge difference on the performance among these four algorithms. However, as budget goes to larger, our greedy algorithm performs much better than the other three methods, *e.g.* with a budget over 700, the expected weight of propagation set exceeds $10,000$ using greedy algorithm, the greedy algorithm outperforms the other heuristics by 30%. Fig. 2(b) tell us similar results, there is no huge difference among different algorithms when the budget is small, when the budget becomes larger and larger, the greedy algorithm outperforms the other heuristics by up to 30%. Remember that in the Top-k Selection, we choose the nodes which can perform best individually. The experimental results show that the performance of this algorithm is almost the worst. This basically tells us that significantly better marketing results can be obtained by explicitly considering the dynamics of information in a network, rather than relying solely on some properties of single node.

Furthermore, we present the results of our algorithms under the MCIS problem settings. We compare three algorithms: greedy algorithm, Top-k Selection and Random Selection. For random algorithm, the result is averaged over $1,000$ experiments. Fig. 2(c) shows the results under the weighted deterministic model where propagation probability of each link is set to be 1. We plot the cost of the initial node set such that all the nodes receive the information with at least a certain probability (influential probability). As shown in the figure, when the influential probability is low, there is no significant difference observed among the three algorithms. However, as the influential probability increases, the greedy algorithm outperforms the other two algorithms. Moreover, we evaluate the performance of the algorithms under the weighted probabilistic model. The results in Fig. 2(d) and Fig. 2(e) compare the algorithms with $p = 10\%$. Fig. 2(d) shows that the cost of initial node set increases as the influential probability increases, and that the greedy algorithm achieves the lowest initial cost. Then we fix the influential probability to 90% and vary the number of nodes and the results are shown in Fig. 2(e). Again, the greedy algorithm performs the best. At last, we fix the influential probability to 90% and evaluate the performance of the algorithms under varied propagation probability p. As shown in Fig. 2(f) the greedy algorithm achieves the lowest initial cost and outperforms the other two algorithms in all cases.

7. CONCLUSION

In this work, we studied several information propagation problems in DTN by exploring the temporal information of the intermittent connections. For both problems, we develop several efficient algorithms. There are a number of interesting problems left for future research. One problem is to design practically efficient methods when we know that the DTN network is some random network, instead of being an arbitrary network. For example, the network structure at every time slot could be a graph with power-law degree distribution, while the graphs at different time slots are independent. The other direction is to study both problems under new information propagation models such as linear threshold model.

8. ACKNOWLEDGEMENT

The research of authors is partially supported by NSF CNS-0832120, NSF CNS-1035894, NSF ECCS-1247944, NSF CNS-0915331, NSF CNS-1050398, National Natural Science Foundation of China under Grant No. 61170216, No. 61228202, China 973 Program under Grant No.2011CB302705, No. 61202383, the Program for New Century Excellent Talents in University under grant No. NCET-12-0414, the Natural Science Foundation of

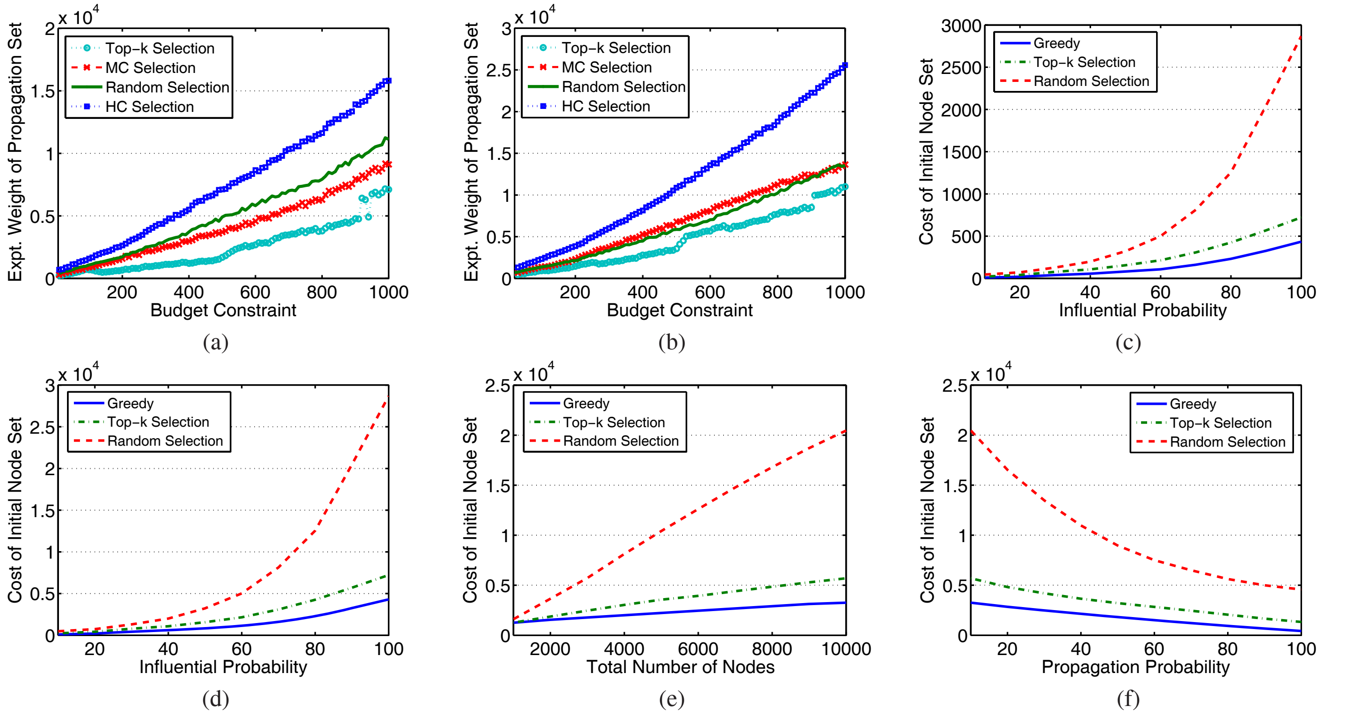

Figure 2: Performance evaluation under weighted probabilistic model and deterministic model.

Shanghai under grant No. 12ZR1451200, the National Science Foundation for Postdoctoral Scientists of China under grant No. 2012M510118, the Research Fund for the Doctoral Program of Higher Education of China under grant No. 20120072120075, and the Special Foundation of China Postdoctoral Science under grant No. 2013T60463.

9. REFERENCES

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *KDD*. ACM, 2003, pp. 137–146.

[2] S. Tang, J. Yuan, X. Mao, X.-Y. Li, W. Chen, and G. Dai, "Relationship classification in large scale online social networks and its impact on information propagation," in *INFOCOM 2011*. IEEE, 2011, pp. 2291–2299.

[3] P. Domingos and M. Richardson, "Mining the network value of customers," in *KDD*. ACM, 2001, pp. 57–66.

[4] A. Chaintreau, P. Fraigniaud, and E. Lebhar, "Opportunistic spatial gossip over mobile social networks," in *Proceedings of the first workshop on Online social networks*. ACM, 2008, pp. 73–78.

[5] F. Chierichetti, S. Lattanzi, and A. Panconesi, "Gossiping (via mobile?) in social networks," in *Proceedings of the fifth international workshop on Foundations of mobile computing*. ACM, 2008, pp. 27–28.

[6] W. Zhao, M. Ammar, and E. Zegura, "Multicasting in delay tolerant networks: semantic models and routing algorithms," in *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*. ACM, 2005, pp. 268–275.

[7] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: a social network perspective," in *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2009, pp. 299–308.

[8] T. Y. Berger-Wolf, "Maximizing the extent of spread in a dynamic network," 2007.

[9] B. Han, P. Hui, V. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *TMC*, vol. 11, no. 5, pp. 821–834, 2012.

[10] Q. Yuan, I. Cardei, and J. Wu, "Predict and relay: an efficient routing in disruption-tolerant networks," in *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2009, pp. 95–104.

[11] X. Zhang, J. Kurose, B. N. Levine, D. Towsley, and H. Zhang, "Study of a bus-based disruption-tolerant network: mobility modeling and impact on routing," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007, pp. 195–206.

[12] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing letters*, vol. 12, no. 3, pp. 211–223, 2001.

[13] G. etal., "Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata," *Academy of Marketing Science Review*, vol. 9, no. 3, pp. 1–18, 2001.

[14] S. Khullera, A. Mossb, and J. Naor, "The budgeted maximum coverage problem," *Information Processing Letters*, vol. 70, pp. 39–45, 1999.

[15] V. Srinivasan, M. Motani, and W. T. Ooi, "Analysis and implications of student contact patterns derived from campus schedules," in *MobiCom*. ACM, 2006, pp. 86–97.