Walk Globally, Act Locally: Efficient Influential User Identification in Mobile Social Networks

Maotian Zhang[1], Panlong Yang[1], Chang Tian[1], Chaocan Xiang[1], Yan Xiong[2]
[1] PLA University of Science and Technology, China
[2]University of Science and Technology of China, China
maotianzhang@gmail.com, panlongyang@gmail.com, tianchang163@163.com, xiang.chaocan@gmail.com, yxiong@ustc.edu.cn

ABSTRACT

Being a fundamental and challenging research topic, influential user identification has attracted much attention with the rapid growth of mobile social networks. Previous studies either focus on relatively stable network structure, or need fairly large information overhead in achieving global maxima. In tackling the dynamic topologies, we propose an influential user identification scheme fully exploiting the active mobile users, where the stable state property is leveraged under information potential construction scheme. We present an efficient routing scheme in reaching the global maxima without relying on specific routing protocols. We validate our scheme with both synthetic and real-world mobility traces. The experimental results show that, the proposed scheme achieves considerable performance on influential user identification and route construction, while bringing forth less overhead.

1. INTRODUCTION

The proliferation of mobile devices (*i.e.*, smart phones, tablets, *etc.*) and WiFi access points gives rise to a new frontier for mobile social networks, where mobile users contact, interact and make friends with each other using their wireless devices. In mobile social networks, influential user identification could be more appealing and challenging, since it is a fundamental research issue in *e.g.*, the recommendation system [1], and information propagation analysis system [2].

Previous studies mainly focus on influential user identification in Internet [3] [4] , where network topologies are relatively stable, and the main metrics include network degree, closeness, betweenness, *etc.* [5]. With respect to mobile social networks, most of the studies [6] [7] can identify influential users, leveraging mobility patterns and social relationships, with the cost of excessive information overhead, which is not neglectable, especially for mobile users.

In this paper, we investigate the influential user identification problem in mobile social networks. Different from previous studies, we focus on utilizing the user mobility while requiring little network overhead. Our work is inspired by *information potential* scheme proposed by Loukas *et al.* [8] [9]. Information potential maps local information (*e.g.*, node degree) to a more meaningful value within the global context.

To achieve this, we have to address two major challenges when integrating this distributed technique into influence measurement schemes of mobile social networks. First, the connectivity of mobile social networks is more dynamic than that of traditional social networks due to users' mobility. The proposed solution should be tailored towards the specific connectivity properties of mobile social networks. Second, the communication and computation overhead should be minimized for energy-constrained mobile devices. Specifically, the information overhead incurred by the information potential computation should be lightweight and tolerable to user mobility.

In this work, we exploit the user mobility as well as its social relationship. The insight is that, under the information potential framework, the active mobile users will help propagate information to other communities effectively, which also impose great influence to mobile networks. In accounting this character, we've proved the convergence condition in information potential construction and propagation. Then, we incorporate the user mobility into the influential user identification procedure, leveraging the stable state property. We further propose a method to shape the influence landscape effectively, and an efficient route construction scheme leading to local maxima efficiently.

The contributions of this work are summarized as follows:

- We propose a distributed algorithm to compute mobile users' influence in mobile social networks. The presented algorithm has a number of desirable properties such as the stable state and the unimodality.
- We give a method to select the influential users based on the stable state of influence computed through our algorithm. Meanwhile, we construct a route that a mobile user can reach the influential users locally or the most influential user globally.
- We evaluate our algorithms and methods through simulations based on both a synthetic random-walk based trace and a real-world mobility trace [10]. We effectively construct the influence landscape and the influence-ascent route based on real-world mobility trace.

The rest of the paper is organized as follows. Section 2 describes the network model and the problem statement. We give the design of proposed algorithm and present the solution of the influential user identification problem in Section 3 and 4, respectively. Section 5 shows evaluation results. Related work is discussed in Section 6. Finally, we conclude our work in Section 7.

MSCC'14, August 11, 2014, Philadelphia, PA, USA.

http://dx.doi.org/10.1145/2633675.2633680.

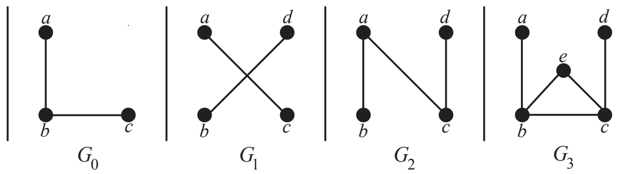


Figure 1: A sequence of dynamic network snapshots which are changing over time.

2. NETWORK MODEL AND PROBLEM STUDIED

In this section, we describe the network model and the problem of influential user identification.

2.1 Network Model and Notations

Basic notations: Assuming that there are N mobile nodes in set $V = \{n_1, n_2, \cdots, n_N\}$. Each node $n_i, 1 \leq i \leq N$, maintains a neighbor-node set $\mathcal{C}(n_i)$ including itself and its current neighbors, *i.e.*, $\mathcal{C}(n_i) = \{n_j : n_i \sim n_j \; or \; n_i = n_j\}$, where $n_i \sim n_j$ denotes that two nodes n_i and n_j are neighbors. Here, two nodes n_i and n_j are considered to be neighbors if they fall into the transmission range of each other. Notice that $\mathcal{C}(n_i)$ may change dynamically due to the mobility of nodes. We further define the degree of a node n_i (denoted by $d(n_i)$) is the cardinality of its neighbor-node set, *i.e.*, $d(n_i) = |\mathcal{C}(n_i)|$. For simplicity, we use i to denote node n_i when no confusion rises.

Let $G = (V, E)$ be an undirected unweighted graph, where V contains all N mobile nodes and E is the set of connects. We then define several matrix notations of the graph G. Let A be the $n \times n$ symmetric adjacency matrix of graph G, and it is given by

$$A_{ij} = \begin{cases} 1 & \text{if } i \sim j \text{ or } i = j \\ 0 & \text{otherwise} \end{cases}$$

Let D be the $n \times n$ diagonal degree matrix with $D_{ii} = d(i)$, and correspondingly the inverse degree matrix $D_{ii}^{-1} = 1/d(i), \forall i \in V$ and zero otherwise. With adjacency matrix A and diagonal degree matrix D, the transition matrix of random walk P can be given by $P = D^{-1}A$ according to the study in [11].

Dynamic network model: A mobile social network is intrinsically a highly dynamic network, which can be modeled as the specific time-varying graph [12]. Let $G_t = (V_t, E_t)$ be a time-varying network snapshot at time t and $G_0 = (V_0, E_0)$ be the original input network. Correspondingly, a dynamic network can be defined as a sequence of network snapshots changing over time: $\mathcal{G} = (G_0, G_1, G_2, \ldots)$. As demonstrated in Figure 1, there is a sequence of network snapshots at different times.

Based on above model, we analyze the proposed algorithm in Section 3, especially its desiable properties which be exploited to identify the influential users in Section 4.

2.2 Problem Studied

We try to address the following two questions: (*i*) How do we compute mobile users' influence distributedly? And (*ii*) how do we effectively select the influential users and efficiently construct a route to reach them?

Intuitively, a mobile user's influence is simply similar to his/her degree-based centrality in the mobile social topology. It is simple to compute the user centrality if we acquire the global topology through centralized solutions. However, centralized solutions suffer from the high computational complexity, and hardly capture the real-time topology of mobile social networks. We adopt a distributed algorithm to dynamically compute user's influence. We use the spirit of "information potential" rather than the degree itself to represent the user's influence. With this setup, we succeed to compute the users' influence and further choose the set containing several influential users. Additionally, constructing a route to reach these influential users is necessary. We introduce the *influence-ascent* to represent the largest influence increment among the user's local neighbors, and the user further reaches the local influential user according to these influence-ascents.

Formally, given a dynamic network $\mathcal{G} = (G_0, G_1, G_2, \ldots)$, where G_0 is the input network and $G_1, G_2, \ldots$ are network snapshots obtained through a collection of network topology changing over time. The basic problem is to cumulatively compute the influence of each user in different network snapshots. Let $d, y : V \to \mathbb{R}$ be functions over V that assign a real value to each user. $d(i)$ is the degree obtained by each user $i \in V$ in a dynamic network $\mathcal{G}$. Function y is the user's instant influence derived from d, and we will present this function in Section 3. The first objective is to efficiently and iteratively compute y through a distributed method until it converges to a stable state π_y. Based on the stable state of influence of each user, the second objective is to find the influential user set U, which contains several influential users. Finally, we aim to construct a route to reach the influential users.

3. INFLUENCE COMPUTATION

In this section, as the first step to identify the influential users, we compute the mobile users' influence through a distributed algorithm. We then describe some desirable properties of the proposed algorithm. In particular, we analyze the convergence property in a dynamic and time-varying network.

3.1 The Distributed Algorithm

The proposed distributed algorithm is based on the constrained-averaging algorithm proposed by Andreas Loukas *et al.* [8] and is to compute a user's influence in mobile social networks.

Specifically, our algorithm differs from this previous study in the following aspects. First, we directly make the user degree be information computed by each user. A user's degree at different network snapshot can be obtained through mobile device discovery, *i.e.*, the smartphone periodically discovers its neighbors within its communication range during a specific time interval. Second, we consider that the user's influence of snapshot G_t is determined by that of the previous snapshot G_{t-1}. That means the influence computation is a cumulative process in which the converged influence increases with the evolution of network. The rationale behind this lies in the connectivity properties of mobile social networks, where users are highly mobile and the connectivity are changing over time [13], such that the user could independently aggregate its connections locally for computing its influence globally.

A mobile user i's influence $y(i)$ is given by:

$$y(i) = (1 - \varphi) \sum_{j \in \mathcal{C}(i)} \frac{y(j)}{d(i)} + \varphi d(i) \qquad (1)$$

where the parameter φ represents the inhibiting factor that inhibits or stimulates diffusion and controls the transition from global to local. When $\varphi = 0$, the equation is a pure diffusion process leading to uniform y and realizes a global view. On the contrary, when $\varphi = 1$, the influence only depends on a local one-hop view.

The iterative computation of user influence works as follows. During each time slot, user i exchanges the influence $y(i)$ with their neighbors through mobile device discovery and information exchange. During each round of iterative computation, the user receives neighbors' information. Then, the user updates its neighboruser set and degree, respectively. Further, the user updates its influence to the weighted sum of their locally degree $d(i)$ and of the average over the most recent influence values, including its own (as Equation 1 shows). The algorithm runs continuously adapting to any network dynamics, and if no dynamics are expected, termination is locally decided by comparing the difference of the influence at consecutive rounds against some error threshold.

3.2 Algorithmic Properties

We are able to analyze the algorithm properties by leveraging the spectral graph theory [11] and random walks on graphs [14]. Based on [8], our algorithm satisfies the following two desirable properties:

- **The Stable State**: The proposed algorithm converges to the stable state of the influence, denoted as π_y. According to the model described in Section II and the theory of random walks on graphs, the graph based on P is time-homogeneous, irreducible, and aperiodic. For irreducible, aperiodic, time-homogeneous graphs, it is well known that a unique stationary distribution state exists and that the graph converges to this distribution from any initial state [11]. As a result, the final influence π_y of each user is the converged influence through the iteratively distributed computation. (It is worth noting that the literature [8] has presented the convergence analysis of computing information potentials in a static network, as well as a dynamic but time-invariant network.)
- **The Unimodality Property**: A function y is unimodal if and only if it has a single extremum at user i, and for each user, at least one path to user i exists on which the values of y are monotonic. Our method can provide the influence's unimodality property, which guarantees a single extremum and allows us to combine the influence computed in different areas with chemotactic search, so that the user can independently navigate towards the most influential user.

4. IDENTIFYING INFLUENTIAL USERS

So far we have computed each user's influence and analyzed the algorithm convergence for dynamic and time-varying networks. In this section, we present a method to effectively identify and select the influential users, and efficiently construct a route in which each user can travel to its local influential user.

4.1 Identifying Influential Users

Taking the mobility and activeness of mobile users into consideration, we define the influential user on a graph as follows:

DEFINITION 1. *The influential users on a graph are the most frequently visited nodes among their local neighbors by random walks on the graph.*

As previous section mentions, we obtain the distribution of the user's influence which converges to a stationary distribution state. Then, we capture these local maxima as influential users and further shape an influence landscape. Recalling the previous influence computation, we obtain each user's influence that has converged to a stable state π_y. The influential user $s \in \mathcal{C}(i)$ has the highest π_y among user i's local neighbors (including user i itself).

4.2 Constructing Influence-ascent Route

One of our contributions is to efficiently construct a route to reach the influential users. The weighting factor of degree is $\big((1-\varphi)P\big)^k$. P_{ij}^k is the probability matrix that a random walker starts from user i and reach user j in k steps. The probability is associated with the connectivity and the distance between user i and j. The proposed algorithm assigns higher significance to the information (*i.e.*, the user density) residing in nearby users and in users with higher centrality. Drawing on the experience of Authority-Ascent Shift presented in [15], we introduce the influence-ascent $\mathcal{A}(i)$, which means the neighboring node of i with the highest expectation of influence increment by a random walk. It is formulated as

$$\mathcal{A}(i) = \arg\max_{j \in \mathcal{C}(i)} P_{ij}^k \big(\pi_y(i) - \pi_y(j)\big) \tag{2}$$

THEOREM 1. *The influence-ascent procedure for any starting user is finite and converges to a local maximum (i.e., the influential user among local neighbors).*

PROOF. The influence-ascent of the successive shifts keep strictly increasing until it reaches a user whose influence-ascent is itself. So, the length of the sequence is at most the number of local users, and the final user has the highest π_y. $\square$

As a result, starting from a user, successive shifts to influence-ascents progress toward its local maximum. Thus, a mobile user is able to reach the influential user according to an influence-ascent route. Compared to the greedy routing, our method exhibits both the computation-efficiency and energy-efficiency properties.

Algorithm 1 describes the procedure of selecting influential user set and constructing influence-ascent route. At first, we obtain the stable state of the influence of each mobile user. We can find the local influential user whose influence is the local maximum among local neighbors (line 2). Meanwhile, the influence-ascent value $\mathcal{A}$ is computed according to Equation 2 (line 3), and it will be exploited to construct the influence-ascent route R (line 6). Besides, we choose the influential users who hold local maxima into the influential user set (line 4).

Algorithm 1 Influential User Selection and Influence-Ascent Route Construction

Input:
  The converged influence π_y.
Output:
  Influential user set U, and a route R towards the local influential user.

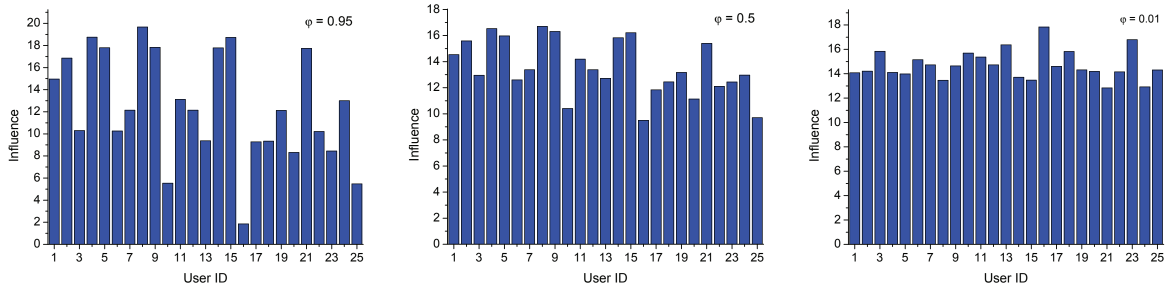
```
1: for each user i do
2:    Find the local maximum, i.e., the highest π_y.
3:    Compute the influence-ascent A(i) according to Equation 2
4:    Add user s ∈ C(i) of local maxima into U, U = U ∪ {s}
5: end for
6: Construct R: associate each user i with its local maximum by
   tree traversals along A(i).
```

5. EVALUATION

In this section, we present the evaluation of the proposed method. We first perform a synthetic random-walk based simulation to validate our algorithm. We then conduct a real-world trace-driven simulation to show the effects of selecting different parameters. Finally, based on the chosen real-world mobility trace, we shape the influence landscape and construct the influence-ascent route.

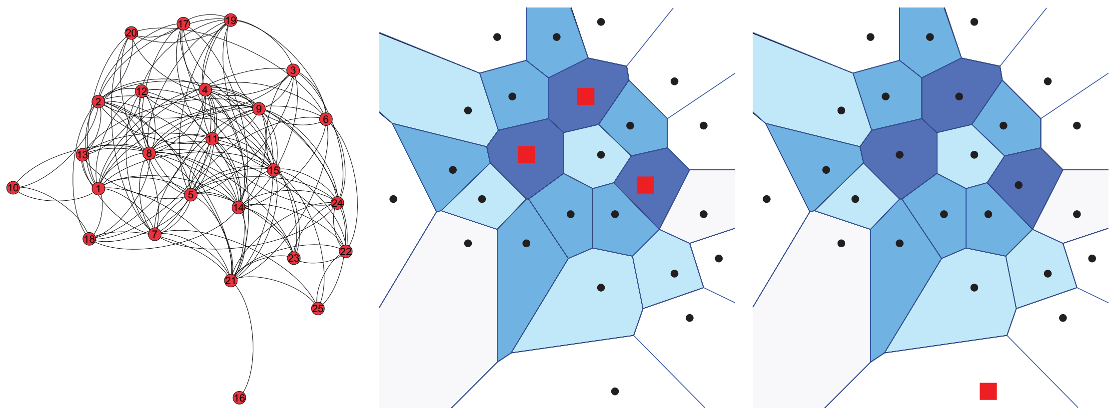
5.1 Synthetic Random-Walk based Simulation

We perform a synthetic random-walk based simulation to validate our model and the influence computation algorithm. In particular, we evaluate the convergence property of the proposed method.



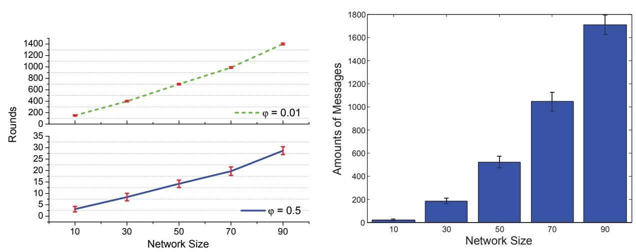
(a) The computed influence of $\varphi = 0.95$ (b) The computed influence of $\varphi = 0.5$ (c) The computed influence of $\varphi = 0.01$

Figure 3: The effects of selecting different inhibiting factor φ.



(a) The social graph of Sassy (b) $\varphi = 0.5$, 3 influential users (c) $\varphi = 0.01$, 1 most influential user

Figure 4: The influence landscapes (b-c) on the social graph of 25 mobile users (a). The red squares in (b) and (c) represent the influential users. There are three influential users with $\varphi = 0.5$ and one most influential user with $\varphi = 0.01$.



(a) # of rounds until the algorithm converges (b) # of messages

Figure 2: The overheads versus network size

We create a simulator in Matlab to generate the synthetic random-walk based mobility trace. Each mobile user performs the random walk during a fixed time. When two user discover each other within the communication range, they record this encounter and exchange the influence computed by the proposed algorithm. The trace records the user IDs and their corresponding encounters, and consequently we can capture the random-walk based social graph. We test five network sizes: 10, 30, 50, 70, and 90 users. We evaluate two representative values of the inhibiting factor, *i.e.*, $\varphi = 0.5$, and 0.01, respectively.

Fig. 2a shows the number of rounds until the algorithm converges. We record the rounds when the convergence error becomes smaller than 0.05. The number of rounds increases with increasing the network size. In addition, the proposed method incurs an overhead that is on the order of $1/\varphi$, which means that the number of rounds of 10 users translates from about 3 for $\varphi = 0.5$ to about 150 for $\varphi = 0.01$. For small networks, it is preferable to choose high values of φ.

We plot the message overhead of the proposed scheme with different network sizes in Fig. 2b. The messages consumption increases with increasing the network size. Recalling the proposed algorithm in Section 3, the mobile user i discovers neighbors and broadcasts a message which consists of ID i and the computed influence $y(i)$ to these neighbors. As a result, the inhibiting factor φ does not affect the messages consumption.

5.2 Real-World Trace-Driven Simulation

We introduce a real-world mobility trace, Sassy [10], with which we use for performance evaluation, and then present the results from a trace-driven simulator.

To simulate the mobile social network, we utilize a real-world mobility trace, the Sassy dataset, which records the encounters of each mobile device. The simulator first generates the encounter trace of mobile users. A mobile user continuously records its local encounters. Two mobile users will start exchanging the computed influence after they know each other through periodic device dis-

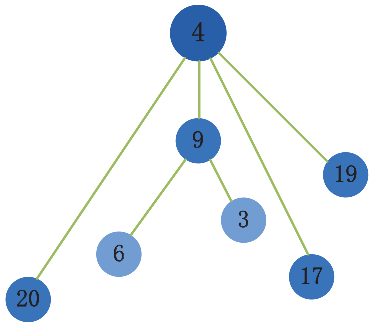


Figure 5: An influence-ascent route with a specific influential user, and $\varphi = 0.5$.

covery. Each mobile user iteratively computes its influence based on received information and its local encounters (*i.e.*, its degree) until the time of this mobility trace is up.

5.2.1 Effects of The Inhibiting Factor φ

The inhibiting factor φ determines the aggregation scope. To show the effects of selecting different inhibiting factor, we plot each user's converged influence for different inhibiting factor $\varphi = 0.95, 0.5$, and 0.01 in Figure 3. The influence differences among each user for $\varphi = 0.01$ are smaller than that for $\varphi = 0.05$ and $\varphi = 0.95$, which means a global view. These results validate our algorithm that the scope are controlled from local to global by decreasing the inhibiting factor. In addition, if we select the influential users into the influential user set U, it will be the user sets $\{4, 8, 9, 15\}$, $\{4, 8, 15\}$ and $\{16\}$ for $\varphi = 0.95, 0.5$, and 0.01, respectively. The number of influential users is varied by choosing different inhibiting factor, and the reason lies in its smoothness property.

We are able to adjust the inhibiting factor φ to control the shape of the *influence landscape*. An instance of the social graph of the Sassy trace is shown in Figure 4a. The maxima are shown as red squares in the Voronoi cells, as Figure 4b and 4c show. Figure 4c eliminates two of the three local maxima (influential users) of Figure 4b, and thus has a single extremum (the most influential user). Therefore, we can reshape the influence landscape through controlling the inhibiting factor.

5.2.2 The Influence-Ascent Route

According to Algorithm 1, we construct the influence-ascent route based on the Sassy trace. The converged influence of each user is first computed by the influence computation algorithm. Based on this, each user compute the influence-ascent according to Equation 2. We use the first random walk step of $k = 1$ [15]. Consider the network size of the Sassy dataset, we set the inhibiting factor φ be 0.5, and thus the influential user set U is $\{4, 8, 15\}$. Each user shifts to the local influential user according to his/her influence-ascent. The construction of the influence-ascent route is shown in Figure 5. For simplify, we show a trajectory tree in which the root node is mobile user 4. For a specific user 6, the influence-ascent route can be obtained as $6 \to 9 \to 4$.

6. RELATED WORK

Influential user identification in mobile social networks: With the increasing of social networking, some research issues are proposed, such as the influence maximization problem and influential user identification problem. In traditional social networks, Domingos and Richardson [3] are the first to study the influence maximization as an algorithmic problem and propose a probabilistic solution. Kemple *et al.* [4] formulate the problem of finding a influential user set as an optimization problem, and propose a Greedy Algorithm guaranteeing $(1-1/e)$ approximation. More recently, Han and Srinivasan [7] proposed a distributed and lightweight protocol for identifying mobile influential users through random walks. Our solution differentiates their work in the two following aspects: The proposed approach achieves that mobile users can realize that how influential they are; We propose an efficient way of navigating towards the influential users.

Information potentials: Information potentials are usually employed as mediums of information aggregation or discovery pertaining to sensor networks [16]. Lin *et al.* [16] construct smooth harmonic gradients towards sources so that local forwarding guarantees their discovery. However, their approaches do not concern the landscape formation.

Loukas *et al.* [9] [8] propose a localized mechanism for determining the information potential on each node based on local process and the potential of neighboring nodes. They also analyze the convergence of information potentials for static networks, and provide a mechanism for reshaping the information landscape. While our work is partially motivated by the spirit of information potentials, ours focuses more on the dynamic and time-varying network, and addresses the influential user identification problem.

7. CONCLUSION

In this paper, we investigate the problem of identifying influential users in mobile social networks. To compute the mobile users' influence, we present a distributed algorithm which is inspired by information potentials. In particular, we analyze the algorithm properties with the stable state and the unimodality property. Furthermore, considering the mobility and activeness of mobile users, we present a method to effectively identify the influential users and efficiently construct an influence-ascent route so that a mobile user can reach the local influential user. We leverage both the synthetic random-walk based trace and the real-world mobility trace to evaluate our methods. The simulation results validate that the presented method brings forth comparably less overhead. Additionally, we show the shaping of influence landscape, as well as the construction of influence-ascent route.

In future work, we want to validate our methods through real-world implementation and integrate them into the mobile applications.

Acknowledgments

This research is partially supported by NSF China under Grants, 61232018, 61272487, 61170216, 61172062, 61371124.

8. REFERENCES

[1] M. Jamali and M. Ester, “Trustwalker: a random walk model for combining trust-based and item-based recommendation,” in *ACM KDD*, 2009.

[2] M. Cha, A. Mislove, and K. P. Gummadi, “A measurement-driven analysis of information propagation in the flickr social network,” in *ACM WWW*, 2009.

[3] P. Domingos and M. Richardson, “Mining the network value of customers,” in *ACM SIGKDD*, 2001.

[4] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *ACM SIGKDD*, 2003.

[5] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.

[6] N. P. Nguyen, T. N. Dinh, S. Tokala, and M. T. Thai, “Overlapping communities in dynamic networks: their detection and mobile applications,” in *ACM MobiCom*, 2011.

[7] B. Han and A. Srinivasan, “Your friends have more friends than you do: identifying influential mobile users through random walks,” in *ACM MobiHoc*, 2012.

[8] A. Loukas, M. Zuniga, M. Woehrle, and M. C. K. Langendoen, “Think globally, act locally: On the reshaping of information landscapes,” in *ACM/IEEE IPSN*, 2013.

[9] A. Loukas, M. Woehrle, P. Glatz, and K. Langendoen, “On distributed computation of information potentials,” in *ACM FOMC*, 2012.

[10] G. Bigwood, D. Rehunathan, M. Bateman, T. Henderson, and S. Bhatti, “CRAWDAD data set st_andrews sassy (v. 2011-06-03),” Downloaded from http://crawdad.cs.dartmouth.edu/st_andrews/sassy/.

[11] F. R. Chung, “Spectral graph theory,” 1997.

[12] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro, “Time-varying graphs and dynamic networks,” in *Ad-hoc, Mobile, and Wireless Networks*. Springer, 2011, pp. 346–359.

[13] F. De Pellegrini, D. Miorandi, I. Carreras, and I. Chlamtac, “A graph-based model for disconnected ad hoc networks,” in *IEEE INFOCOM*, 2007.

[14] L. Lovász, “Random walks on graphs: A survey,” *Combinatorics, Paul erdos is eighty*, vol. 2, no. 1, pp. 1–46, 1993.

[15] M. Cho and K. M. Lee, “Mode-seeking on graphs via random walks,” in *IEEE CVPR*, 2012.

[16] H. Lin, M. Lu, N. Milosavljevic, J. Gao, and L. J. Guibas, “Composable information gradients in wireless sensor networks,” in *ACM/IEEE IPSN*, 2008.