Information Sciences 450 (2018) 182–199

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Exploring influence maximization in online and offline double-layer propagation scheme

Yongjian Yang, Yuanbo Xu, En Wang*, Kaihao Lou, Dongming Luan

Department of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China

ARTICLE INFO

Article history:
Received 22 November 2017
Revised 24 February 2018
Accepted 18 March 2018
Available online 19 March 2018

Keywords:
Online social networks
Offline mobile networks
Information propagation
Influence maximization

ABSTRACT

Information propagation in network environment is a widely studied research topic, especially in Online Social Networks (OSNs), where the problem has gained significant popularity. Recent studies attempt to pick up the key nodes, who could maximize the network influence in OSNs. However, in addition to propagation in OSNs, another information propagation way is through words of mouth among people in the offline mobile network, which is an indispensable factor and is not considered in most cases. Hence, the information propagation in both online social network and offline mobile network is a new valuable scheme. In this paper, we propose an Information Maximization strategy in Online and Offline double-layer Propagation scheme (IMOOP), where we first form the topological graph for online social network and offline connection graph of probability, respectively. Then, the two layers are compressed into a single-layer communication graph. We further prove that the influence maximization in double-layer propagation scheme is NP-hard, then we describe practical greedy heuristics for the resulting NP-hard problems and compute their approximation ratios. Our experiments with real mobility datasets (Brightkite, Gowalla and Foursquare) show that, the proposed propagation scheme achieves a higher information cover ratio, compared with the other propagation methods.

1. Introduction

Online Social Networks (OSNs) [15] have drawn a great amount of attention in recent years because they provide social applications, such as Facebook [33], Google Plus [5], Twitter [16], Wechat [44], Microblog [43], *etc*. Users could share humor, issue advertisements, make friends, and appoint activities through a variety of social application platforms [27]. There are plenty of existing works for online social networks' topological features, social group management, *etc*. Among them, one of the most important research issues is the influence maximization problem, which is to select a subset of nodes named seed nodes, so that the spread of information could be maximized if we treat the seed nodes as the initial spreading nodes. The influence maximization problem has been extensively studied in online social networks [1]. However, the existing works have a common limitation that only the influence propagation in the online social network is considered, while the influence propagation for offline mobile network is overlooked in most cases. In fact, an event can also be propagated through words of mouth among the people in offline mobile network. Sometimes, the spreading by words of mouth propagation even has greater and faster impact compared with the spreading of information in online social networks.

* Corresponding author.
E-mail addresses: yyj@jlu.edu.cn (Y. Yang), yuanbox15@mails.jlu.edu.cn (Y. Xu), wangen@jlu.edu.cn, wangen0310@126.com (E. Wang), loukh5513@mails.jlu.edu.cn (K. Lou), luandm5513@mails.jlu.edu.cn (D. Luan).

https://doi.org/10.1016/j.ins.2018.03.048
0020-0255/

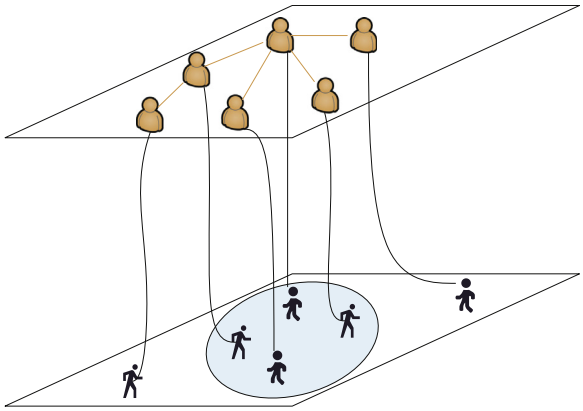

Fig. 1. An example of online and offline double-layer propagation scheme.

This paper is inspired that in real-world, information can be transferred through both virtual world (Tweets or Weibo) and the physical world (mouth to ear). And in some special scenarios, we need to control the spread of information, to make most or least people to be influenced. Hence, we need to find the key nodes in the online and offline double-layer network to achieve the goal, such as Public Opinion Control (POC), advertising promotion, *etc.*

As shown in Fig. 1, actually, each node in Online Social Networks has the dual identities: online social node and offline mobile user. As a result, a node has two ways to spread the information: one is sharing the information with online social friends, the other one is through words of mouth among encountered people. In this propagation scheme, a node could disseminate the information to an unacquainted node quickly, which means that this scheme assists in information spreading. For example, node A is a net friend of B in Wechat, users B and C are colleagues. However, A is not familiar with C. Someday, A issues an interesting event through Wechat, and B quickly gets the event. Meanwhile, B is currently having a meeting with C, and he tells C the interesting event. In this way, A propagates the event to a stranger C, , which speeds up the spread of information. In online social network, a node's influence degree may be decided by the number of friends and its social activeness. In offline mobile network, a node's influence degree is always considered as the frequency that it encounters another node. Hence, a node has two influencing ways in both online and offline networks, the problem is to decide which node has the maximum influence in a given online and offline network topology. Furthermore, if we want to select top k nodes as seed nodes to spread the information, in order to maximize the influence, we should decide the optimal solution including k users. An easy solution is deciding an influence utility according to the double-layer topology, and selecting the top k nodes as the seed nodes. However, this method will lead a huddle phenomenon, which means that a group of nodes are selected, and they could not disseminate the information to other groups.

In this paper, we first propose an online and offline double-layer information propagation scheme, which assists in spreading information as widely as possible. This scheme allows the nodes with the information influence the other nodes through both online social network and offline mobile network. Then, in this scheme, we also propose the top k seeds selection algorithm taking the nodes' influences in the double-layer into consideration, in order to maximize the total influence of the selected k nodes. Furthermore, the two layers are combined into a communication graph. We prove that the influence maximization in double-layer propagation scheme is a NP-hard problem, practical greedy heuristics are discussed and their approximation ratios are also computed for the resulting NP-hard problem. Our experiments with real mobility datasets (Brightkite, Gowalla and Foursquare) show that, the proposed propagation scheme achieves a superior information cover ratio, compared with state-of-the-art baselines.

The main contributions of this paper are briefly summarized as follows:

- We propose a double-layer information propagation scheme, which consists of online social network and offline mobile network. The proposed propagation scheme could assist in spreading the information as widely as possible.
- In the proposed double-layer propagation scheme, to maximize the information influence, we propose a seeds selection algorithm taking the nodes' influences in the double layer into consideration.
- Extensive experimental results are presented, which demonstrate the meaningfulness of the proposed information propagation scheme and also verify that the seeds selection algorithm to maximize influence has high performance.

The remainder of this paper is organized as follows: We review the related work in Section 2. The network model and problem formulation are presented in Section 3. The online and offline double-layer propagation scheme is proposed in Section 4. In Section 5, we propose the seeds selection algorithm for influence maximization, in order to select the most valuable k nodes to spread the information as widely as possible. In Section 6, we evaluate the performance of the proposed propagation scheme and selection algorithm through extensive simulations. We conclude the paper in Section 7.

2. Related work

2.1. Mobile social networks

There are plenty of works that focus on propagation in social networks [23,25,39]. He et al. [12] proposes a kind of epidemic model, which includes two methods to describe rumor spreading in MSNs. Moreover, in order to block the rumor spreading, two cost-efficient strategies are proposed, one is Real-Time Optimization (RTO), another is Pulse Spreading truth and Continuous Blocking rumor (PSCB). Thilakarathna et al. [31] proposes a low-cost content storage and distribution system. In this system, the content replication problem is proved NP hard. Then this paper proposes a method replicating the content on a minimum friends' devices to maximize availability based on a greedy heuristic algorithm. Li et al. [20] proposes the model that extends classical dynamic routing, in which the collection of information is aggregated over all users' trips. Then the paper combines the advantages of social information sharing and the routing problem, in the problem non-atomic selfish users choose between a high-cost and a low-cost path in a non-cooperative game. In order to protect location privacy in the location-based services, Gong et al. [6] motivates mobile users to participate in pseudonym change. For users, whether to change pseudonym is cast as a socially aware pseudonym change game (SA-PCG). For SA-PCG, it is proved that there exists a socially aware Nash equilibrium (SNE). Xiao et al. [37] proposes an Average makespan sensitive Online Task Assignment (AOTA) algorithm and a Largest makespan sensitive Online Task Assignment (LOTA) algorithm in order to solve the task assignment problems. Besides, two greedy strategies which are respectively adopted for the task assignments in AOTA and LOTA are proposed. In [35] they propose a propagation- and mobility- aware content replication strategy for edge-network regions, social contents are assigned to users according to a comprehensive consideration. Then paper formulates the replication strategy as an optimization problem and designs a distributed algorithm to solve it. Finally, the trace-driven experiments demonstrate the superiority of this proposal. Gan et al. [4] proposes a novel game-based incentive mechanism for multi-resource sharing, which involves a combination of some processes and satisfies truthfulness, individual rationality and robustness. Moreover, in order to maintain the social fairness-efficiency tradeoff, the paper develops a resource sharing algorithm on the basis of Dominant Resource Fairness(DRF). In [13], they propose a general framework to model the video transmission among mobile users and user QoS. Then the paper uses the hierarchical structure to divide this problem into two problems, one is bitrate adjustment problem and the other is a spectrum allocation problem. For the first problem, the paper proposes an online bitrate adjustment strategy. For the second problem, this paper proves the problem is a potential game and designs a decentralized algorithm to find the Nash equilibrium. In order to address the diffusion minimization problem, Lu et al. [24] improves the approximation algorithm for the asymmetric k-center problem and proposes a community based algorithm and a distributed set-cover algorithm. Gong et al. [7] studies mobile users' data usage behavior by both considering the network effect and the congestion effect. Then the paper develops a Stackelberg game for data usage, paper analyzes the two-stage game. For Stage I, paper develops an optimal pricing algorithm to maximize the wireless provider's revenue. For Stage II, paper provides conditions of a user demand equilibrium (UDE) and proposes algorithm to find the UDE and the distributed manner for users. There are also some works using the trust measures in propagating the influence. In [26], they provide a positive answer to the question: centrality-based reputation scores allow for predicting helpfulness-based reputation ones and Eigenvector Centrality was found to be the most important predictor. In [3], they demonstrate, by an extended set of experiments on datasets extracted from real communities, that trust measures can effectively replace profile matching in order to optimize group's cohesion.

The above works focus on the area of MSNs. These works could be regarded as the preliminary work of this paper, and also an important part in online social network. However, the works do not consider the influence in offline mobile networks. Consequently, the proposed propagation scheme could not be directly used in the double-layer propagation scheme of this paper.

2.2. Location-based social networks

There are plenty of works that focus on influence maximization problem in social networks [10,11,18,28,29], especially for location-based social networks. In [22], they propose an overall geographical probabilistic factor model (Geo-PFM) framework which considers various factors. Based the Geo-PFM framework, they have developed a Poisson Geo-PFM with a more rigorous probabilistic generative process. In [21], they present a Content-aware Collaborative Filtering (ICCF) framework to combine semantic content and avoid negative sampling. Gu et al. [8] proposes a Home Location Global Positioning System named HLGPS using an overall iteration algorithm based on IME model that aims to address home location recognition issues in Location-Based Social Networks (LBSNs). In order to make full use of the geographical influence, in [42], they present a personalized and efficient geographical location recommendation framework named iGeoRec with a approximation algorithm with approximation and efficiency guarantees. Yang et al. [38] builds a Spatial Temporal Activity Preference (STAP) model which first takes the spatial and temporal activity preference into consideration respectively and adopts a context-aware fusion framework to combine them together. Yao et al. [40] proposes a new way which includes the temporal matching between users and POIs into POI recommendations and builds a Temporal Matching Poisson Factorization Model (TM-PFM) to profile the temporal popularity of POIs. In [9], they present a trust-based impact model called TSU to model edges in LBSN. Based on TSU model, they introduce a Home Location Identification method with two stages and they are the first using the idea of "social trust" for identification. Wen et al. [36] presents a new social impact based user recommenda-

tion framework (SIR) to discover users with influence in LBSN and design a dynamic weight tuning method to incorporate the features obtained into unified follow probability scores. In [19], they utilize mobile crowdsourced data collected from LBSNs services to research the maximization of users' influence in LBSNs. A new network model and a new influence propagation model is proposed, which takes the online social networks and the physical world into considerations. Kempe et al. [15] formulates the processes of choosing the most influential sets of nodes as a NP-hard optimization problem and obtains an algorithm with the provable approximation guarantees. Zhou et al. [45] researches the issues of influence maximization under O2O environment. A new model named TP for O2O considering both online and offline parts is proposed. Moreover, they also introduce a maximization algorithm called TPH.

The above works focus on the work of influence maximization in LBSNs. However, to the best of our knowledge, there is no work focusing on proposing the top k nodes selection algorithm in double-layer propagation scheme. Meanwhile, the double-layer propagation scheme proposed in this paper matches the information propagation in the real world, and could select the optimal top k nodes to propagate the information.

3. Network model and problem formulation

3.1. Network model

Online social networks can be grouped into undirected online social networks (Facebook) and directed online social networks (Twitter). Statistics show that nodes act differently in the above two kinds of online social networks. In this paper, we focus on the influence propagation in undirected online social networks. Besides, in offline mobile network, the edges are also undirected, which means that a pair of nodes could share their information with each other.

Suppose there are N nodes denoted by $V = \{v_1, v_2, \cdots, v_N\}$ deployed in the double-layer networks. Each user is equipped with a smart phone and has an account in the undirected online social network. Each smart phone is equipped with a positioning component which can obtain its geographical position in the offline mobile network and a mobile telecommunication component which is able to send information to the online social network. We use $v_i.x_t$ and $v_i.y_t$ to denote the x-coordinate and y-coordinate of node $v_i \in V$ at time t. For two nodes v_i, $v_j \in V$, the Euclidean distance between v_i and v_j at time t are denoted as follows:

$$d(v_i, v_j, t) = \sqrt{(v_i.x_t - v_j.x_t)^2 + (v_i.y_t - v_j.y_t)^2} \tag{1}$$

Two nodes v_i, $v_j \in V$ are in communication in the offline mobile network at time t if $d(v_i, v_j, t) \le r$, where r is the predefined propagation radius. In this paper, we assume that the information maybe probabilistically propagated from an influenced user v_i to v_j, if v_i and v_j are friends in the online social network, and the information will be propagated if they are in communication in the offline mobile network. Then the online social network and nodes' graphical locations in offline mobile network can be described as a double-layer graph. $G^t = (V, E_{on}, E^t_{off})$, where V is the set of nodes, E_{on} and E^t_{off} are two undirected edge sets. $\forall v_i, v_j \in V$, $(v_i, v_j) \in E_{on}$ if and only if v_i and v_j are friends in the online social network. Similarly, for time t and $\forall v_i, v_j \in V$, $(v_i, v_j) \in E^t_{off}$ if and only if $d(v_i, v_j, t) \le r$. We assume E_{on} remains the same while E^t_{off} changes over time. In other words, a dynamic geometric graph is used to denote users' relationships in the offline mobile networks.

In this paper, we assume that an event such as an advertisement needs to be assigned to k nodes in online social networks. Influenced nodes may propagate this event with their friends in online social networks or their neighbors in offline mobile networks. Generally, the influence of the event will propagate in both online social networks and offline mobile networks simultaneously. For time t, if two nodes are friends in online social network, and the node with the interesting event could share this with its friend in the form of probability. At the same time, the node with the interesting event could also tell its neighbor (the distance is smaller than r) in offline mobile network. Fig. 2 shows two instances, $G1$ and $G2$, of our network model. Each double-layer graph has two subgraphs to describe nodes' relationships in the online social network and in the offline mobile network, respectively. Two nodes with the same ID denote the same node. Suppose an event is activated in the offline mobile network and the nodes within the shadowed influencing region are influenced. Then the influence will propagate in both the online social network and the offline mobile network. For G1, we can see that although node 5 is not in the influencing region, node 5 may still be influenced since node 4 and node 6 are node5's friends in the online social network and node 4 and node 6 are in the influencing region. Similarly, although node 1 and node 7 are not friends in the online social network, the influence may propagate from node 1 to node 7 since they are neighbors in the offline mobile network. However, for G2, nodes 5 and 7 become neighbors in offline mobile network, in G1, they could not communicate with each other, but in G2, they could share their information and influence each other. It is worth noting that, the topology of online social network is fixed, while the offline mobile network is dynamic.

Our first challenge is to develop an appropriate influence propagation scheme to simulate the influence propagation in the above circumstance. The influence propagation in both online social network and offline mobile network should be considered in this scheme. Based on the proposed influence propagation scheme, we can then evaluate the influence of a node. Obviously, the influence results may be quite different for different choices of selected nodes. Then our challenge is to seek k optimal nodes to be seeds of information so that the influence of the information can be maximized. The main notations are illustrated in Table 1.

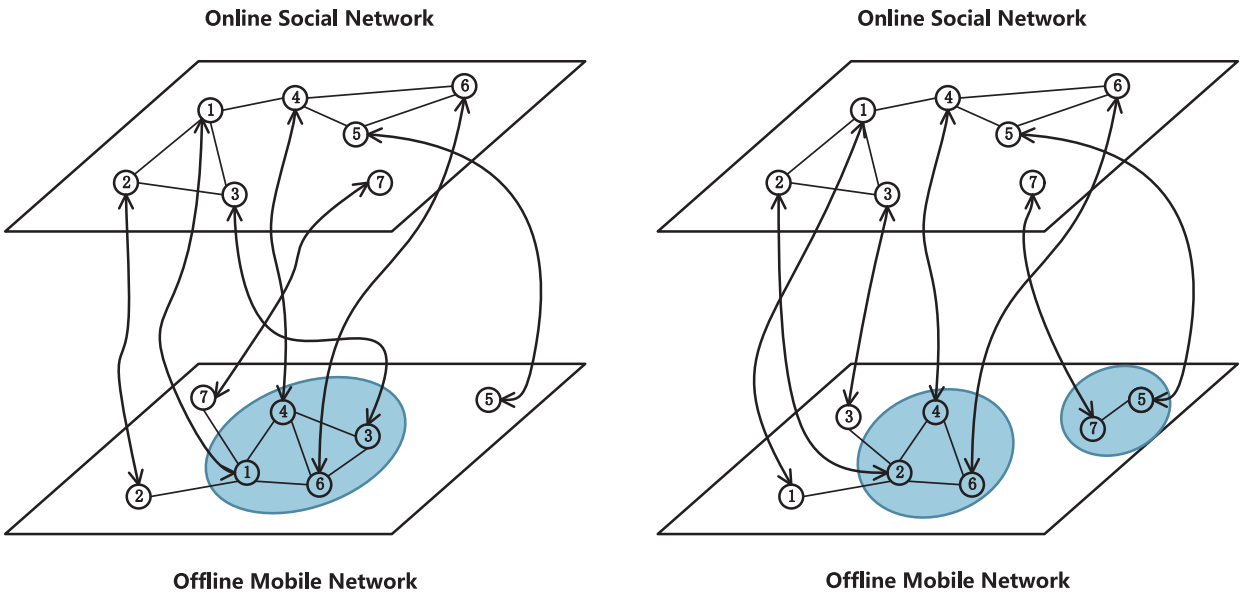

Fig. 2. Two instances of double-layer propagation scheme ($G1$ in left and $G2$ in right).

Table 1
Main notation used throughout the paper.

Symbol	Meaning
N	total number of nodes in the network
V	set of nodes $V = \{v_1, v_2, \cdots, v_N\}$
r	communication radius in offline mobile network
E_{on}	undirected edge sets in online social network
E^t_{off}	undirected edge sets at time t in offline mobile network
$P^{on}_{n_i n_j}$	communication probability in online social network for nodes n_i and n_j
$t_{n_i n_j}$	encountering duration for n_i and n_j in a time cycle
T	time cycle
$P^{off}_{n_i n_j}$	communication probability in offline mobile network for nodes n_i and n_j, $P^{off}_{n_i n_j} = \frac{t_{n_i n_j}}{T}$
$P_{n_i n_j}$	communication probability in both online and offline double-layer scheme
U_{n_i}	$n_i{}'$ total communication probability with all the other nodes
U_S	nodes set S' total communication probability with all the other nodes

3.2. Problem formulation

The network is composed of N nodes, and they formulate a double-layer network environment. In online social network, two friends could use a probabilistic method to disseminate information to each other. In offline social network, two nodes in communication could disseminate the information to each other. Hence, for a node in the double-layer scheme, it has two ways to influence the other nodes. The online social network topology dominates the propagation way in online friends, while the offline mobile network topology depends on the contact probability among nodes, and also dominates the propagation way in physical world. In this situation, the question is that, how to select the optimal k seeds to disseminate the information, in order to maximize the information influence before the information's deadline.

To solve the above problem, we first define the weights of graphs in both online social network and offline mobile network as the communication probabilities in both online and offline networks. An easy solution is that we firstly decide the node's influence utility according to the weight in the double-layer network, which means that the utility is the total weight of the node's edges in both online social network and offline mobile network. Then we select the top k nodes with the highest utilities as the seeds.

However, the above method obviously has flaws, because the nodes with the highest utility may have common friends in online social network and common encounters in offline mobile network. For example, in [19], they prove that if users are friends with each other in the online social networks, their distances in the physical world tend to be closer. Hence, the nodes with the highest utilities may form a group, which could not disseminate the information to the other groups.

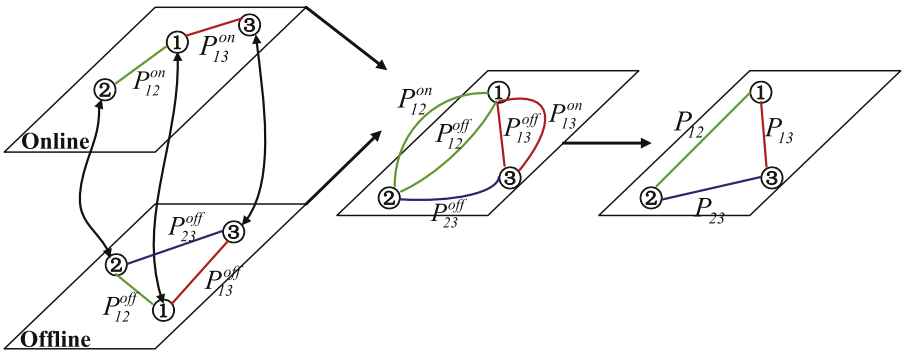

Fig. 3. The process to compress the double-layer propagation scheme into the single-layer propagation scheme.

So they may be not the optimal solution. To address the above problem, we propose a seeds selection algorithm taking the nodes' influences in the double-layer into consideration, in the proposed double-layer propagation scheme, to maximize the information influence. We prove that the influence maximization in double-layer propagation scheme is the NP-hard problem, then we describe practical greedy heuristics for the resulting NP-hard problems and compute their approximation ratios.

4. Double-layer propagation scheme

As shown in Fig. 3, the compressing process of the double-layer is illustrated. In the online social network, node 1 and node 2 are friends, node 1 and node 3 are also friends. In offline mobile network, the pairs of contact probabilities are listed as $P_{n_i n_j}^{off}$. Due to the reason that, the weights of edges in both online and offline networks are defined as the probabilities. Hence, we could compress the double-layer into a single-layer propagation scheme, which is easy for us to discuss the node selection problem.

4.1. Online social network layer

In online social network layer, there is an edge between a pair of nodes if and only if they are friends. The weight of the edge between n_i and n_j is defined as the communication probability of them. We use $P_{n_i n_j}^{on}$ to express the weight. For example, in Facebook, node A and node B are friends, so they have an edge in online layer. However, the information propagation is not immediate between them, because they could not be active in Facebook all the time, if A wants to disseminate the information to B, B will get the information in a probabilistic method. If B browses the Facebook for three times in 1000 time slots, then for node B, $P_{AB}^{on} = 0.003$.

In conclusion, if a pair of nodes are friends, they share an edge in online layer, and the weight of the edge is the communication probability between them in online social network. It is worth noting that, in different network environments, the calculation of $P_{n_i n_j}^{on}$ could be different. For example, Wang et al. [34] treats users differently based on their distances from the promoted location. In [32], they formally model the dynamic independent Cascade model and introduce the concept of adaptive seeding strategy. So we just propose a propagation scheme, which could use different methods to calculate the specific probability.

4.2. Offline mobile network layer

In offline mobile network layer, there is an edge between a pair of nodes if and only if they encounter each other at least once in time cycle T. The weight of the edge between n_i and n_j is defined as their contact probability. We use $P_{n_i n_j}^{off}$ to express the weight. For example, in real world, node A and node B encounter each other for $t_{n_i n_j}$ times, so they share an edge in offline mobile network layer. If A wants to disseminate the information to B through the encounter in real world, the contact probability for them is $P_{AB}^{off} = \frac{t_{n_i n_j}}{T}$.

In conclusion, if a pair of nodes ever encounter each other, they have an edge in offline layer, and the weight of the edge is the contact probability between them in real world. It is also worth noting that, in different network environments, the calculation of $P_{n_i n_j}^{off}$ could be different. For example, Swain et al. [30] conducts a feasibility study to highlight the impact of multi-hop D2D communication in increasing the network coverage and average rate of a Machine Type Communication device. In [14], from the local balance equation of the designed Markov chain, the transition probabilities are derived for distributed implementation. So we just propose a propagation scheme in offline mobile network layer.

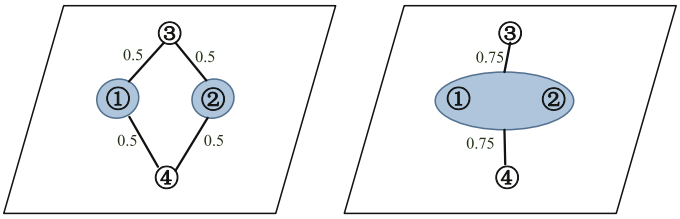

Fig. 4. The explanations for a node's utility and the utility for a set of nodes. For the left subgraph, the utility between node 1 and 3 is 0.5, similarly, the utility for node 2 and node 3 is also 0.5. However, if we select nodes 1 and 2 as the node set S_1, then the utility between S_1 and node 3 is 1-(1-0.5)(1-0.5)=0.75, which is shown in right subgraph.

4.3. Compressed single-layer

We attempt to discuss the information propagation in the double-layer scheme. However, the information propagation way in the double-layer scheme is really complex, which leads us to make the problem easy to solve. Due to the reason that, the weights of edges in both online and offline layers are the communication probability, therefore, we try to compress the double-layer networks into a single-layer, which is easier for us to work on.

The compress process is shown in Fig. 3. For nodes n_i and n_j, they could communicate with each other either in online sharing or in offline encountering. Therefore, the information could not be disseminated between n_i and n_j, if and only if they do not contact each other in offline mobile network, meanwhile, they also do not share the information in online social network. Hence, in the compressed single-layer propagation scheme, each pair of nodes has an edge, whose weight is defined as $P_{n_i n_j}$:

$$P_{n_i n_j} = 1 - (1 - P^{on}_{n_i n_j})(1 - P^{off}_{n_i n_j}) \tag{2}$$

In conclusion, we propose the online and offline double-layer propagation scheme, where each pair of nodes has an edge. The weight of an edge is defined as the communication probability, which considers both the online and offline influences. Then the double-layer propagation scheme is compressed into a single-layer propagation scheme.

5. Top *K* nodes selection algorithm

In the previous section, we compress the double-layer into a single-layer graph. The graph consists of N nodes and the edges among nodes. Every edge has a weight, which is a probability. The purpose of selecting top k nodes is to regard the k nodes as seeds, which have the largest influence. However, the influence of the k nodes is not the brief sum of every node, it is the utility of the node set. The node's utility is defined as the total weights of its outline edges. For example, as shown in Fig. 4, the utility of node 1 is 1 (0.5+0.5).

However, if we put nodes 1 and 2 together, and make it node set S_1. Then, the edge between S_1 and node 3 is 0.75 $(1-(1-0.5)(1-0.5))$. Similarly, the edge between S_1 and node 4 is also 0.75 $(1-(1-0.5)(1-0.5))$. Hence, the utility for S_1 is 1.5 (0.75+0.75). The purpose of this paper is to find the top k nodes, which forms optimal set S, in order to maximize the utility of S: U_S, the above problem is called top k nodes selection problem.

5.1. Top k *node selection*

Before the solution, we first prove that the k nodes selection problem is NP-hard, as shown in the following theorem.

Theorem 1. *The k nodes selection problem is NP-hard.*

Proof. We consider a special case of the k nodes selection problem: all the edges' weights of the undirected graph is equal to 1, and in the single-layer graph, if nodes n_i and n_j both have an edge with n_k, then the set $S_1 = \{n_i, n_j\}$ has an edge with n_k, whose weight is equal to 1. Actually, this special k selection problem is to select the k number of nodes who can cover as many as possible other nodes. this problem can be seen as a k set cover problem, a well known NP-hard problem: given a task set S, a collection of subset $\{S_i | 1 \le i \le n\}$, find a k size of subcollection of $\{S_i | 1 \le i \le n\}$ that covers as many as possible tasks in S. That is to say, the special k selection problem is NP-hard. Consequently, the general k selection problem is also at least NP-hard. The theorem holds. □

Since the k nodes selection problem is NP-hard, we propose a greedy algorithm to solve it. An easy solution is that, we use the greedy criterion that the node who has the largest utility in the single-layer graph will be recruited as seeds and added to the set S first. The same action is repeated for k times. The detailed algorithm is shown in Algorithm 1.

However, the Largest selection algorithm is not the optimal method obviously. For example, in Fig. 5, if we use algorithm 1 to select the top two nodes, in the first round, node 5 is selected because its utility is 3. In the second round, node 6 is selected, because its utility is higher than any other node. However, nodes 5 and 6 are not the optimal top 2 nodes, because

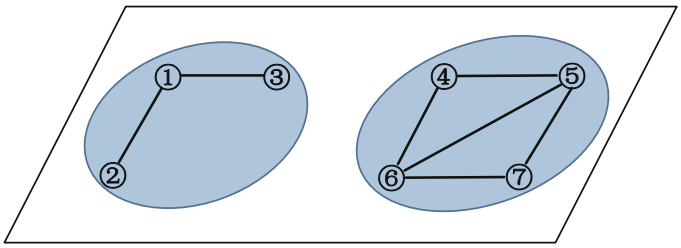

Fig. 5. The different node selection strategies. We assume that the weights of all the edges are 1, then if we select the nodes with the highest utilities, then nodes 5 and 6 will be selected as the seeds. However, in our selection IMOOP, nodes 5 and 1 will be selected. Obviously, our selection is better than selecting the nodes with the highest utilities.

nodes 1 and 5 are obviously better than the set selected by algorithm 1. Nodes 5 and 6 will influence the node set {4, 5, 6, 7}, but nodes 1 and 5 will influence the node set {1, 2, 3, 4, 5, 6, 7}.

In order to decide the efficient top k node selection strategy and make the influence maximization, we propose the greedy heuristic for IMOOP node selection, which selects the best node to be added to set S, in order to maximize the U_S, rather than selecting the node of maximum utility. As shown in Fig. 5, in IMOOP, we will select the nodes 1 and 5, rather than nodes 5 and 6. The detailed algorithm of the IMOOP is shown in Algorithm 2. Besides, Algorithm 3 shows an easy selection strategy, which randomly selects a node in the remain node set as the seed in each round. It is not difficult to find that, the time complexities of the above three algorithms (Largest, IMOOP and Random) are $O(Nlog_2^N)$, $O(N^2)$ and $O(N)$, while the space complexities are $O(N)$, $O(N)$ and $O(1)$.

5.2. Approximation ratio

Theorem 2. *U_S is a submodular function. More specifically, for two arbitrary node sets S_1 and S_2, $S_1 \subseteq S_2$, and $\forall n_k \in N \backslash S$, the submodular property holds, i.e., $U_{S_1 \cup n_k} - U_{S_1} \geq U_{S_2 \cup n_k} - U_{S_2}$.*

Proof. We first prove that when $|S_2| - |S_1| = 1$, $U_{S_1 \cup n_k} - U_{S_1} \geq U_{S_2 \cup n_k} - U_{S_2}$. Then, we extend it to the general case where $|S_2| - |S_1| = \omega > 1$.

First, without loss of generality, we let $S_2 \backslash S_1 = \{n_h\}$ according to the assumption $S_1 \subseteq S_2$, and $|S_2| - |S_1| = 1$. To prove the submodular property of U_S, we consider the joint successful processing probability of $\forall n_j \in N$, which can be divided into the following three cases:

Case 1: n_k has no communication probability with n_j. For this case, $P_{n_k n_j} = 0$. Then, we have $U_{S_1 \cup n_k} = U_{S_1}$ and $U_{S_2 \cup n_k} = U_{S_2}$. As a result, $U_{S_1 \cup n_k} - U_{S_1} = U_{S_2 \cup n_k} - U_{S_2} = 0$.

Case 2: n_k has a communication probability with n_j, but n_h has no communication probability with n_j. For this case, $P_{n_h n_j} = 0$. According to Eq. (2), $U_{S_2} = U_{S_1 \cup u_h} = U_{S_1}$, and $U_{S_2 \cup n_k} = U_{S_1 \cup n_k \cup n_h} = U_{S_1 \cup n_k}$. Consequently, we can get $U_{S_1 \cup n_k} - U_{S_1} = U_{S_2 \cup n_k} - U_{S_2}$.

Case 3: Both n_k and n_h have a communication probability with n_j. Then for all the nodes n_i in S_1, the communication probability with n_j is defined as P_{1j}, similarly, for S_2, the communication probability with n_j is defined as P_{2j}. Obviously, $P_{1j} \leq P_{2j}$, then $U_{S_1 \cup n_k} - U_{S_1} = 1 - (1 - P_{1j})(1 - P_{kj}) - P_{1j}$. Similarly, $U_{S_2 \cup n_k} - U_{S_2} = 1 - (1 - P_{2j})(1 - P_{kj}) - P_{2j}$. Therefore, we have

Theorem 3. *For a non-negative, monotone submodular function f, let S be a set of size k obtained by selecting elements one at a time, each time choosing an element that provides the largest marginal increase in the function value. Let S^* be a set that maximizes the value of f over all k-element sets. Then $f(S) \geq (1 - 1/e) \cdot f(S^*)$; in other words, S provides a $(1 - 1/e)$-approximation.*

$$
\begin{aligned}
&(U_{S_2 \cup n_k} - U_{S_2}) - (U_{S_1 \cup n_k} - U_{S_1}) \\
&= (1 - (1 - P_{2j})(1 - P_{kj}) - P_{2j}) - (1 - (1 - P_{1j})(1 - P_{kj}) - P_{1j}) \\
&= (P_{1j} - P_{2j}) P_{kj} \leq 0
\end{aligned}
$$

Therefore, $U_{S_1 \cup n_k} - U_{S_1} \geq U_{S_2 \cup n_k} - U_{S_2}$.

In summary, $U_{S_1 \cup n_k} - U_{S_1} \geq U_{S_2 \cup n_k} - U_{S_2}$ holds for $\forall n_j \in N$ in all cases. Now, we consider the case of $|S_2| - |S_1| = \omega \geq 1$. Without loss of generality, we assume that $S_2 \backslash S_1 = \{n_h, n_{h+1}, \cdots, n_{h+\omega-1}\}$. Then, we have $U_{S_1 \cup n_k} - U_{S_1} \geq U_{S_1 \cup n_k \cup n_h} - U_{S_1 \cup n_h} \geq U_{S_1 \cup n_k \cup n_h \cup n_{h+1}} - U_{S_1 \cup n_h \cup n_{h+1}} \geq \cdots \geq U_{S_1 \cup n_k \cup n_h \cup \cdots \cup n_{h+\omega-1}} - U_{S_1 \cup n_h \cup \cdots \cup n_{h+\omega-1}} = U_{S_2 \cup n_k} - U_{S_2}$.

Therefore, U_S is a submodular function. Theorem 2 is proved. □

Submodular functions have a number of very nice tractability properties; one is shown as follows. Suppose we have a function f that is submodular, takes only nonnegative values, and is monotone in the sense that adding an element to a

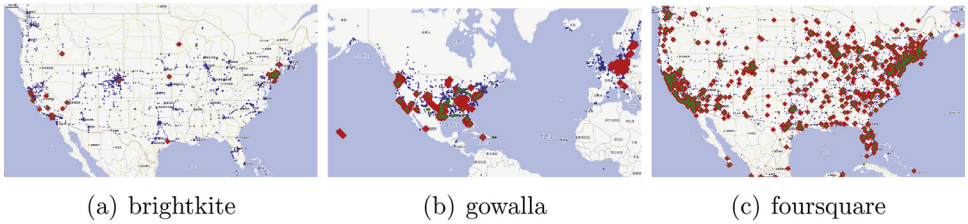

Fig. 6. The traces of the selected nodes in Baidu map of the three real-world data sets.

Table 2
Simulation parameters.

Parameter	Traces		
	Brightkite	Gowalla	Foursquare
Simulation Time	10–100	40–130	10–100
Number of nodes	2241	1537	241
k		30–80	
p^{on}		0.002–0.004	
Communication Range		100–300	

set cannot cause f to decrease: $f(S \cup v) \geq f(S)$ for all elements v and sets S. We wish to find a k-element set S for which $f(S)$ is maximized. This is an NP-hard optimization problem (it can be shown to contain the Hitting Set problem as a simple special case), but a result of Nemhauser, Wolsey, and Fisher. Comuejols et al. [2] shows that the following greedy hill-climbing algorithm approximates the optimum within a factor of $1 - 1/e$ (where e is the base of the natural logarithm): start with the empty set, and repeatedly add an element that gives the maximum marginal gain.

6. Performance evaluation

We conduct extensive simulations to evaluate the performances of the proposed top k selection strategies. The traces that we used, the simulation settings, the compared selection strategies, the measured performances, and the simulation results are presented as follows.

6.1. The traces used and settings

We adopt the three widely-used real-world traces, *Brightkite trace set, Gowalla trace set*, and *Foursquare trace set* [17,41], to test the performances of the proposed top k selection strategies. The *Brightkite trace set* was once a location-based social networking service provider where users shared their locations by checking-in. The friendship network was collected using their public API, the network is originally directed but we have constructed a network with undirected edges when there is a friendship in both ways. The *Gowalla trace set* is a location-based social networking website where users share their locations by checking-in. The friendship network is undirected and was collected using their public API, we have collected a total of 6,442,890 check-ins of these users over the period of Feb. 2009 – Oct. 2010. The *Foursquare trace set* is a location-based social networking web, software for mobile devices. This service is available to users with GPS enabled mobile devices, such as iPhones and Blackberries. This contains the friendship network crawled in December 2010 by Fred Morstatter.

We first address the above three real-world data sets by filtering some abnormal user traces (discontinuous records or remote areas) node traces. Then, we put the traces into a Baidu map according to the GPS. Through invoking the JavaScript API in the Baidu map, we draw the traces of the selected nodes (Fig. 6). The nodes selected by the proposed IMOOP are colored as red, and the nodes selected by the Largest strategy are colored with green. The unselected nodes are colored with blue. The detailed simulation parameters in this network environment are listed in Table 2.

6.2. Strategies and performances in comparison

To demonstrate the performance of the proposed top k selection strategy, we carry out simulations focusing on the following two parts: (1) nodes could only influence the neighbors, and the influenced neighbors could not influence other nodes (cover ratio). (2) nodes could only influence the other nodes as an Epidemic manner (cover ratio).

For the first part, we mainly consider the following three top k selection strategies: (1) IMOOP, which is proposed in this paper. IMOOP adds the node into node set S in each round, which could maximize the utility of the selected node set S. (2) Largest, which selects the node with the highest utility as seed in each round. (3) Random, which randomly selects k nodes in all the N nodes as seeds. We attempt to prove that IMOOP achieves the highest cover ratio compared with the other two top k selection strategies. However, in this part simulations, only the selected k seeds could influence the neighbors (one hop), the cover ratio is calculated through the number of influenced neighbors.

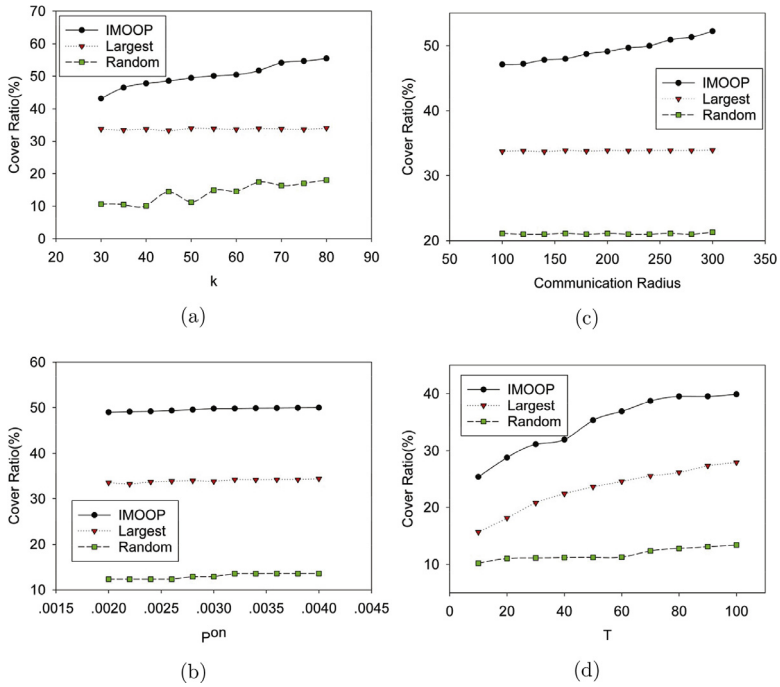

Fig. 7. Performance comparisons on the real world trace Brightkite: cover ratio in Basic Influence Propagation.

For the second part, we turn our attention to the influence manner. The simulation settings are similar to that of the first part. However, the seeds could spread the information in the Epidemic manner, the influence level will be higher than that of the first part simulations.

While a range of data is gathered from the simulations, we take the following one main performance metrics into consideration:

(1) Cover Ratio, which is the ratio between the node number covered by the information and the total number of nodes.

6.3. Simulation results

We use three datasets (Brightkite(2241), Gowalla(1573) and Foursquare(241)) to evaluate our proposed scheme IMOOP.

Our experiment consists of two parts: Basic Influence Propagation (one hop) and Extended Influence Propagation (multi hops). In Basic Influence Propagation, only selected k nodes can influence their neighbors. While in Extended Influence Propagation, the nodes which are influenced can also influence the other neighbors. There are four key parameters in two-layer influence propagation: selected node number k, communication radius in offline mobile network r, communication probability in online social network P^{on} and propagation time T. Each time we fixed three of them, and evaluate the other parameter's effect on cover ratio in both Basic and Extended Influence Propagation.

6.3.1. Performances of brightkite

Figs. 7 and 8 show experiment results of influence propagation in Brightkite. For Brightkite, the initialized value of these parameters are k=50, r=200, P^{on}=0.003 and T=300. In Figs. 7-a and 8-a, we change selected node number k from 30 to 80 with an increment of 5. In both Basic and Extended Influence Propagation, we find that our proposed IMOOP outperforms the other baseline methods for Brightkite. Random achieves the worst performance because it picks source nodes without any heuristic information. IMOOP outperforms Largest about 10% of cover ratio in Basic Influence Propagation and 5% in Extended Influence Propagation. With the node number growing, the distance between IMOOP and Largest becomes larger. Note that in Fig. 8-a, when we pick 30 nodes, Largest performs similar IMOOP. The reason is that Largest pick the most popular node in networks ignoring the influence between two nodes. So if we pick a small set of nodes in a small scale network, Largest can achieve a good cover ratio as IMOOP can do. However, the real-world dataset is huge and large-scale. IMOOP totally outperforms Largest when deployed in a large scale network and select a big set of source nodes, as shown in Fig. 8-a.

In Figs. 7-b and 8-b, we change communication radius in offline mobile network r from 100 to 300 with increment of 20. It's obvious that the larger r is, the better cover ratio it can achieve. Also, IMOOP performs best while Random performs worst. IMOOP outperforms Largest about 15% of cover ratio in Basic Influence Propagation, which is a great improvement. While in Extended Influence Propagation, the gap between IMOOP and Largest becomes smaller, but IMOOP is still better

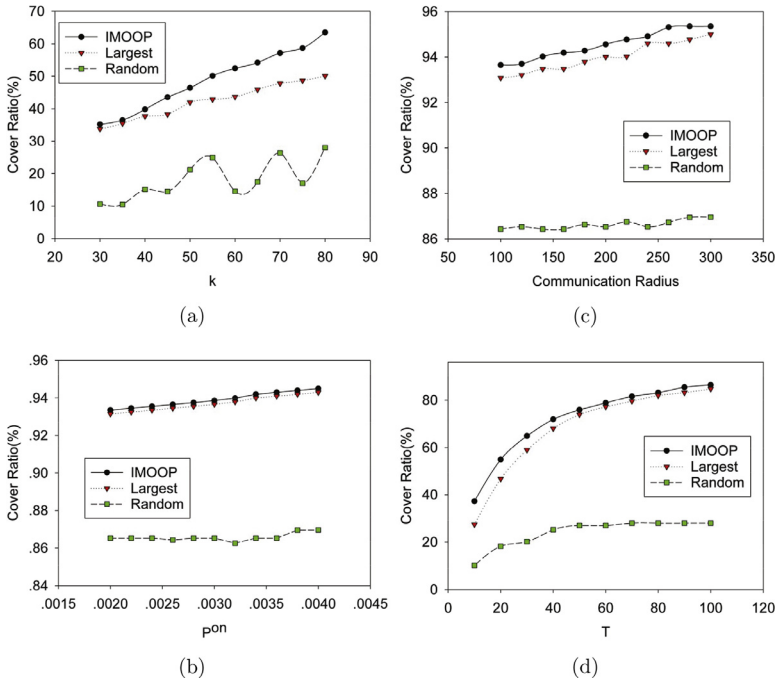

Fig. 8. Performance comparisons on the real world trace Brightkite: cover ratio in Extended Influence Propagation.

than Largest. The reason is that in this experiment, each point presents the convergence of cover ratio in different r. And our proposed method uses more heuristic knowledge. So we can select proper nodes as source nodes.

In Figs. 7-c and 8-c, we change communication probability in online social network P^{on} from 0.002 to 0.004 with increment of 0.0002. We can see in Basic Influence Propagation and Extended Influence Propagation, P^{on} has a little effect on cover ratio for each method. The reason is that in Brightkite, number of user's social neighbors is little compared with neighbours' in real-world. And IMOOP still performs best. In Basic Influence Propagation, IMOOP outperforms Largest with 10% and Random with 30%. That's a great improvement. While in Extended Influence Propagation, IMOOP and Largest are in the same level and IMOOP is a little better.

In Figs. 7-d and 8-d, propagation time T is the parameter we want to evaluate, which ranges from 10 to 100 with increment of 10 time intervals. In Basic Influence Propagation, IMOOP outperforms Largest about 10% in cover ratio, and outperforms Random about 30%. While in Extended Influence Propagation, we can see that IMOOP can achieve a better cover ratio faster than Largest can do and improve the performance. If we set the threshold of cover ratio as 50% in Extended Influence Propagation, IMOOP achieves at 14th time intervals, Largest achieves at 23rd time intervals and Random cannot achieve it. In a word, IMOOP performs better than two baseline methods Largest and Random, and Random performs worst in Brightkite.

Random is the baseline method which gets the poorest performance in Brightkite dataset. Note that IMOOP performs much better than Largest with different parameters. The reason is that Largest method only picks nodes with highest utility every round, ignoring the common situations that the nodes with highest utility are always neighbors, which weakens the spread of information. However, IMOOP finds the max utility of selected nodes set and achieves a global optimization. So our proposed method achieves the best performance among baselines.

6.3.2. Performances of gowalla

Figs. 9 and 10 show experiment results of influence propagation in Gowalla. In Gowalla, the initialized value of these parameters are k=50, r=200, P^{on}=0.003 and T=300. We design experiments as we did in Brightkite.

Figs. 9-a and 10-a show that IMOOP can achieve better cover ratio when changing the scale of selected node number N compared with Largest and Random, as in Brightkite. But we find that in Basic Influence Propagation, Random performs better than Largest. (shown in Fig. 9-a). Also, it happens in the other three experiment of Basic Influence Propagation (changing communication radius in offline mobile network r, communication probability in online social network P^{on} and propagation time T, shown in Fig. 9 b–d). This is very interesting because Largest uses the most popular k nodes as source nodes. With this heuristic knowledge, it should be able to achieve a better performance than Random does. However, when we analyze the nodes Largest and Random selected, and the construction of networks, we find the source nodes that Largest selected have formed a group that each node is the neighbor of the other nodes in this group. So these k nodes selected by Largest are most popular in networks because they have much more neighbors in online and offline world than others, but cover

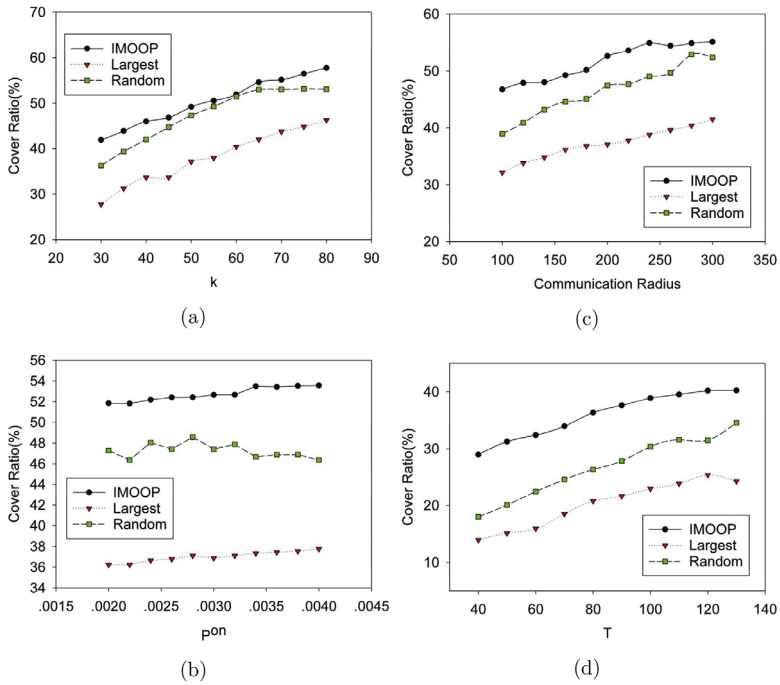

Fig. 9. Performance comparisons on the real world trace Gowalla: cover ratio in Basic Influence Propagation.

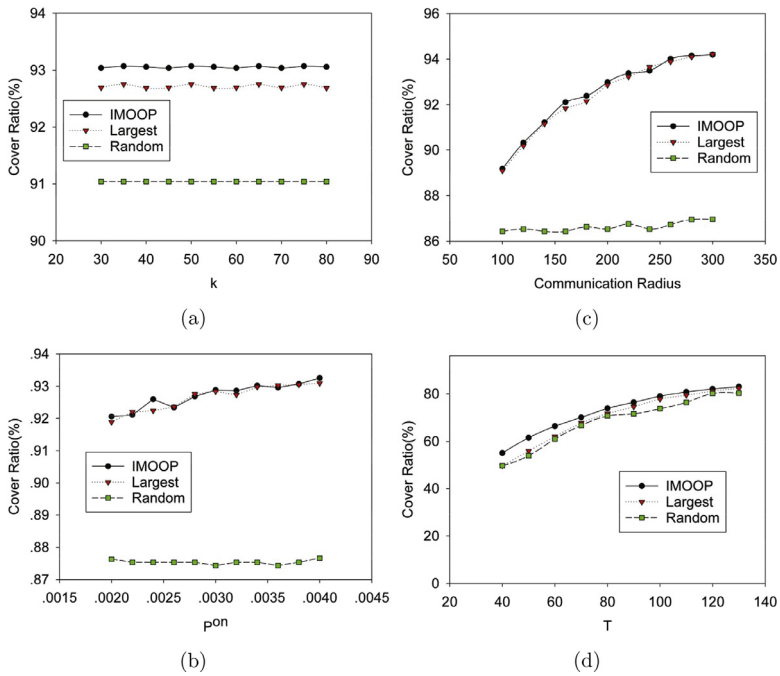

Fig. 10. Performance comparisons on the real world trace Gowalla: cover ratio in Extended Influence Propagation.

ratio is not as good as Random because there are so many similar neighbors in these k nodes. Our proposed IMOOP has advantage on this situation by receding the contribution of same neighbors on utility of candidate nodes.

From Fig. 9 a–d, we find that IMOOP achieves better performance than Largest and Random in Basic Influence Propagation. When changing selected node number k, k=60, Random achieves a good performance as IMOOP does shown in Fig. 9-a, while the gap becomes larger when k becomes smaller or larger. And the performance of Random is not stable, so we ensure that our proposed IMOOP is better than Random in Basic Influence Propagation. In Fig. 10-a, the performance of

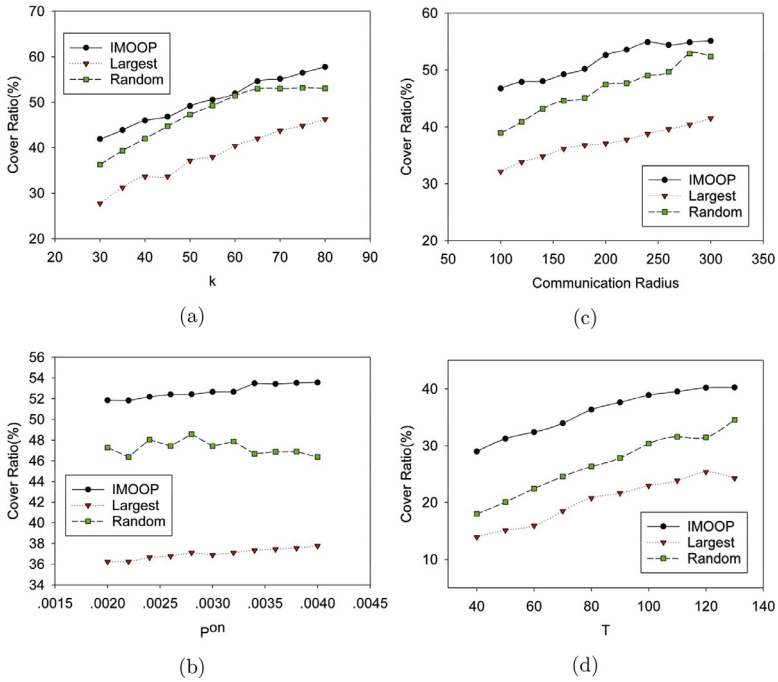

Fig. 11. Performance comparisons on the real world trace Foursquare: cover ratio in Basic Influence Propagation.

all three methods are stable because the cover ratio has become restrained. And IMOOP outperforms the other two baseline methods when cover ratio restrained (0.3% over Largest and 2% over Random).

In Figs. 9-b and 10-b, we change offline communication radius r from 100 to 300 with increment of 20. When changing offline communication radius r, IMOOP performs best and Random performs worst. In Basic Influence Propagation, as shown in Fig. 9-b, IMOOP can achieve a better cover ratio with 3% improvement over Random and almost 15% over Largest. And the performance of IMOOP is more stable than the others. In Extend Influence Propagation, IMOOP is better than Largest in a little extent, while both of them are much better than Random (average 5%).

In Figs. 9-c and 10-c, communication probability in online social network P^{on} changes from 0.002 to 0.004 with increment of 0.0002. Note that in Basic Influence Propagation and Extended Influence Propagation, P^{on} has a little effect on cover ratio for Largest and IMOOP. And IMOOP performs better and more stable. In Basic Influence Propagation, IMOOP outperforms Random with 7% and Random with 15%. While in Extended Influence Propagation, IMOOP and Largest are in the same level and IMOOP is a little better.

We vary propagation time T from 40 to 130 with increment of 10, as shown in Figs. 9-d and 10-d. In Basic Influence Propagation, our proposed IMOOP outperforms Random over 10%, and Largest over 15%. In Extended Influence Propagation, three methods perform at the same level but IMOOP is faster and more stable.

When our method is applied on little-scale dataset, IMOOP consider global utility of selected item set, so it achieves a much better performance in basic influence propagation. But in extended influence propagation, because of the sparse network structure and small neighbor node set for every node, the selected node set is almost similar for IMOOP and Largest. So the performance of both methods are in the same level.

6.3.3. Performances of foursquare

Figs. 11 and 12 show experiment results of influence propagation in Foursquare (LA). In Foursquare (LA), the initialized values of these parameters are k=20, r=200, P^{on}=0.003 and T=100. The scale of Foursquare (LA) is small, convergence rate for cover ratio is fast, so the initialized value in Foursquare (LA) is different from those in Brightkite and Gowalla. In Figs. 11 and 12, we find that Random performs more unstable than in Brightkite and Gowalla, especially in Extended Influence Propagation because the small scale of dataset will increase the randomness of performance.

In Figs. 11-a and 12-a, we change selected node number k from 30 to 80 with increment of 5, and find the relationship between selected node number k and Cover Ratio. Note that our proposed method is better than baseline methods in both Basic Influence Propagation and Extended Influence Propagation on Foursquare (LA) dataset. In Basic Influence Propagation (Fig. 11-a), IMOOP can improve the cover ratio about 10% over Largest, and 25% over Random. In Extended Influence Propagation, we find that IMOOP is over Largest and Random all the time. But when the number of selected nodes k increases, cover ratio will be in convergence. The convergence of cover ratio with three different methods are close to 96%. However, as shown in Fig. 12-a, IMOOP has a better performance compared with baseline method when scale of source nodes k is

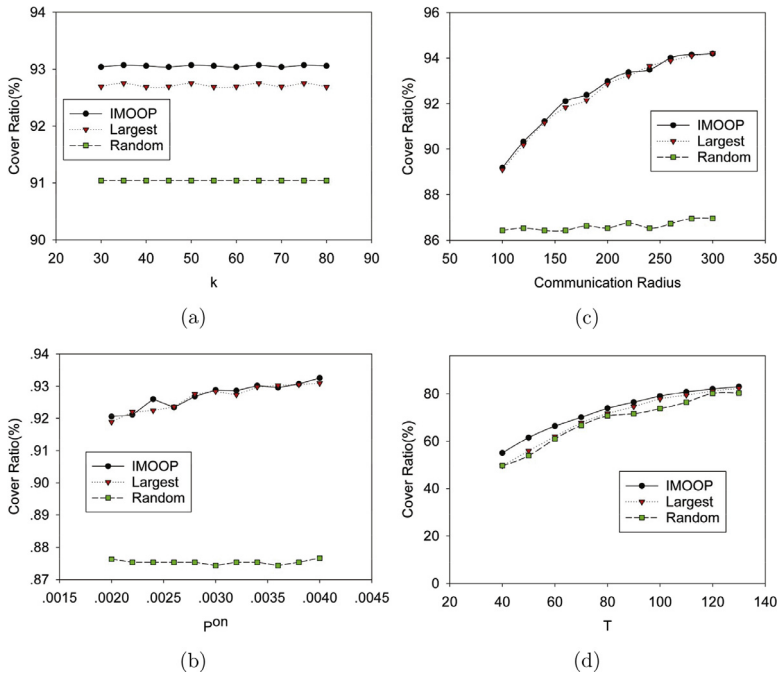

Fig. 12. Performance comparisons on the real world trace Foursquare: cover ratio in Extended Influence Propagation.

small, and it will achieve cover ratio convergence with smaller k. IMOOP achieves convergence when k=60, less than Largest, k=70, and Random, k=75. This experiment proves that our proposed IMOOP can get to the best cover ratio with smaller scale of source nodes, which can save compute ability and storage cost.

In Figs. 11-b and 12-b, we change communication radius in offline mobile network r from 100 to 300 with increment of 20. When changing offline communication radius r, IMOOP performs best and Random performs worst. In Basic Influence Propagation, as shown in Fig. 11-b, IMOOP can achieve a better cover ratio with 6% improvement over Largest and almost 20% over Random. In Extend Influence Propagation, the cover ratio restrains to 98% when r=300 with three methods. And IMOOP can get a better performance when r is small. Note that the larger offline mobile network r is, the more chance two people can exchange information with each other. So our proposed method can achieve satisfied cover ratio when offline communication is changing.

Then it comes to communication probability in online social network P^{on}. In Figs. 11-c and 12-c, P^{on} is the parameter we want to evaluate, which ranges from 0.002 to 0.004 with increment of 0.0002. Note that in Basic Influence Propagation, different P^{on} makes a little sense on cover ratio, and IMOOP is better than Largest about 9%, Random about 60%. The reason is that in Basic Influence Propagation, only selected k nodes can influence the other nodes. And their neighbors online are fixed according to the dataset. So the performance of changing P^{on} is stable with three methods. But in Extended Influence Propagation, the performance of Largest becomes more unstable than IMOOP. The reason is that Largest only picks up most popular nodes as source nodes and ignores the influence of similar neighbors. And randomness of Random will be more obvious in small scale of dataset. Our proposed method IMOOP can get a stable performance on cover ratio when changing P^{on}.

Figs. 11-d and 12-d show the effect of propagation time T on cover ratio, we change T from 10 to 100 with the increment of 10 time intervals. In Basic Influence Propagation (Fig. 11-d), IMOOP can get a higher cover ratio than two baseline methods. In Extended Influence Propagation, the cover ratio of three different methods restrains on 67%. And IMOOP achieves the convergence faster than other methods.

We use three different scales of datasets (Brightkite, Gowalla and Foursquare (LA)) to evaluate our proposed methods with two baseline methods: Largest and Random. We study four important parameters in online-offline influence propagation networks. The results show that no matter in huge or small datasets, our proposed IMOOP can achieve a stable, satisfying performance, which outperforms the baseline methods.

Then, in order to show the influence in the dissemination of online social network. We test the cover ratios of IMOOP-offline and IMOOP in the three real world data sets, respectively. We also set two groups of simulations in one hop and multi hops situations. IMOOP-offline just propagates the information in offline mobile network, while IMOOP propagates the information in both online and offline networks. It is not difficult to find that, IMOOP achieves a better cover ratio compared with that of IMOOP-offline, due to the reason that, online propagation assists in information dissemination Fig. 13.

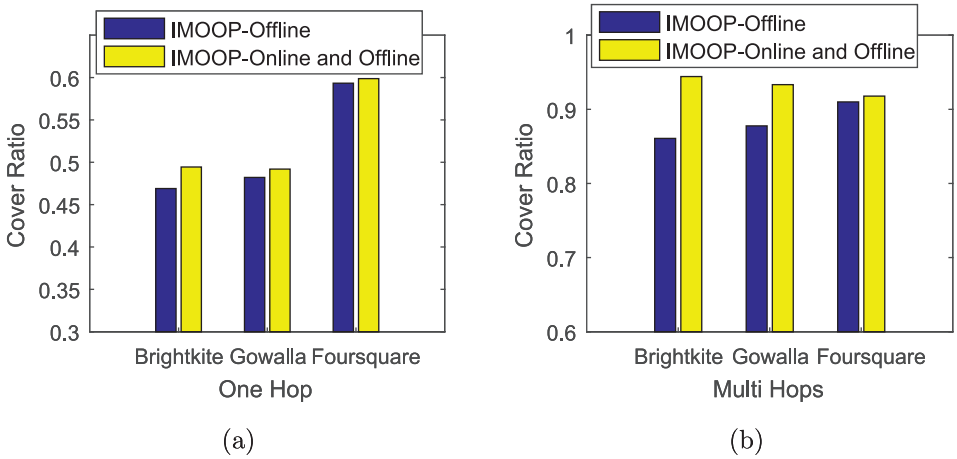

Fig. 13. The compare between IMOOP-offline and IMOOP in three real world data sets.

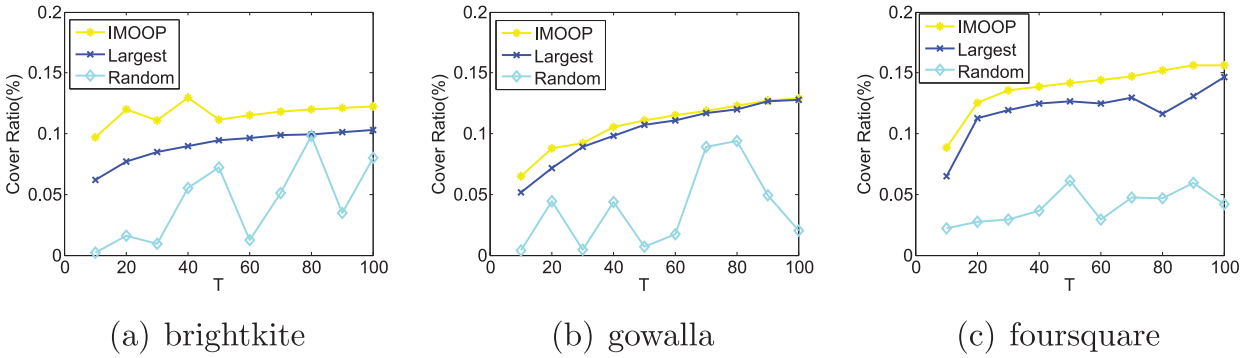

Fig. 14. The experiment results on datasets with a larger number of nodes for one-hop setting.

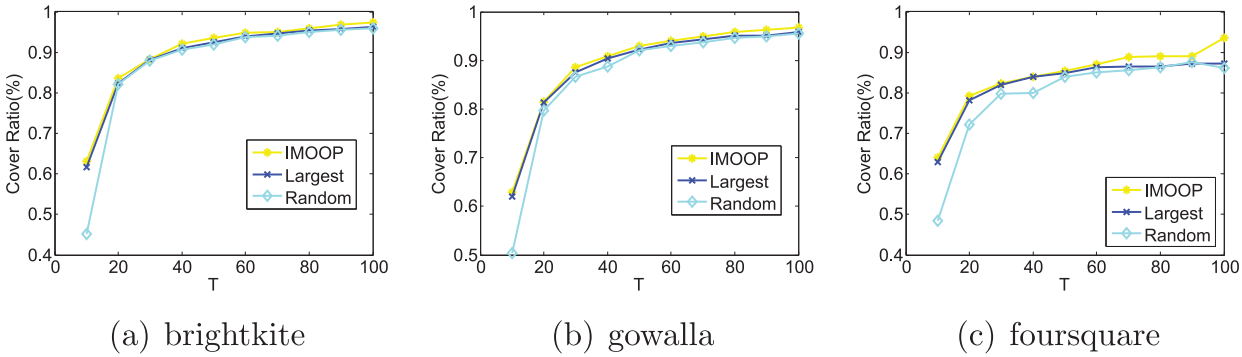

Fig. 15. The experiment results on datasets with a larger number of nodes for multi-hops setting.

Finally, in order to conduct experiments on datasets with a larger number of nodes to test the scalability of compared algorithms, we test the cover ratio along with *TTL*, in the original real-world datasets: *Brightkite trace set* (58,228 nodes), *Gowalla trace set* (196,591 nodes), and *Foursquare trace set* (106,218 nodes). The simulation results are shown in Figs. 14 and 15. As shown in Figs. 14 and 15, IMOOP can achieve a better cover ratio when changing the *TTL* compared with Largest and Random, even in large-scale datasets.

7. Conclusion

We have looked into the problem of information propagation in both online social network and offline mobile network. First, we propose the online and offline double-layer propagation scheme. In this scheme, aiming at spread the information as widely as possible, we further propose a top k nodes selection algorithm. The selected nodes are regarded as seed nodes,

Algorithm 1 Greedy heuristic for Largest node selection.

Input:
Number of nodes: N
Set of nodes: S
Set S' total communication probability: U_S
Output:
Top k nodes set: S
$S \longleftarrow \varnothing$; $U_S = 0$
for i =1 to k **do**
$P \longleftarrow \arg\max_{P \in N \backslash S} U_P$
$S = S \cup P$; update U_S
return S

Algorithm 2 Greedy heuristic for IMOOP node selection.

Input:
Number of nodes: N
Set of nodes: S
Set S' total communication probability: U_S
Output:
Top k nodes set: S
$S \longleftarrow \varnothing$; $U_S = 0$
for i =1 to k **do**
$P \longleftarrow \arg\max_{P \in N \backslash S} U_{S \cup P}$
$S = S \cup P$; update U_S
return S

Algorithm 3 Greedy heuristic for Random node selection.

Input:
Number of nodes: N
Set of nodes: S
Set S' total communication probability: U_S
Output:
Top k nodes set: S
$S \longleftarrow \varnothing$; $U_S = 0$
for i =1 to k **do**
$P \longleftarrow$ random P in $N \backslash S$
$S = S \cup P$; update U_S
return S

which spread the information through both online and offline networks, in order to maximize the influence. Then, the Information Maximization strategy in Online and Offline double-layer Propagation scheme (IMOOP) is proposed. We prove that the influence maximization in double-layer propagation scheme is NP-hard, then we describe practical greedy heuristics for the resulting NP-hard problems and compute their approximation ratios. We conduct extensive simulations based on real mobility datasets (Brightkite, Gowalla and Foursquar). The results show that, the proposed propagation scheme achieves a higher information cover ratio, compared with the other propagation methods.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant No. 61772230 and Natural Science Foundation of China for Young Scholars No. 61702215, China Postdoctoral Science Foundation No. 2017M611322.

References

[1] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proc. of ACM SIGKDD, 2009.
[2] G. Comuejols, M. Fisher, G. Nemhauser, Location of bank accounts to optimize float., Manage. Sci. (23) (1977).
[3] L. Fotia, F. Messina, D. Rosaci, G.M.L. Sarne, Using local trust for forming cohesive social structures in virtual communities, Comput. J. 60 (11) (2017) 1717–1727.
[4] X. Gan, Y. Li, W. Wang, L. Fu, X. Wang, Social crowdsourcing to friends: an incentive mechanism for multi-resource sharing, IEEE J. Sel. Areas Commun. PP (99) (2017) 1–13.

[5] N. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, D. Song, Evolution of social-attribute networks: measurements, modeling, and implications using google+, in: Proc. of IMC, 2012, pp. 131–144.
[6] X. Gong, X. Chen, K. Xing, D. Shin, M. Zhang, J. Zhang, From social group utility maximization to personalized location privacy in mobile networks, IEEE/ACM Trans. Networking PP (99) (2017) 1–14.
[7] X. Gong, L. Duany, X. Chen, J. Zhang, When social network effect meets congestion effect in wireless networks: data usage equilibrium and optimal pricing, IEEE J. Sel. Areas Commun. PP (99) (2017) 1–13.
[8] Y. Gu, J. Song, W. Liu, L. Zou, Hlgps: a home location global positioning system in location-based social networks, in: Proc. of IEEE ICDM, 2016.
[9] Y. Gu, Y. Yao, W. Liu, J. Song, We know where you are: home location identification in location-based social networks, in: Proc. of IEEE ICCCN, 2016.
[10] L. Guo, D. Zhang, G. Cong, W. Wu, K. Tane, Influence maximization in trajectory databases, IEEE Trans. Knowl. Data Eng. 29 (3) (2017) 627–641.
[11] M. Han, M. Yan, Z. Cai, Y. Li, X. Cai, J. Yu, Influence maximization by probing partial communities in dynamic online social networks, Trans. Emerging Telecommun.Technol. 28 (2016).
[12] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, Y. Li, Cost-efficient strategies for restraining rumor spreading in mobile social networks, IEEE Trans. Veh. Technol. PP (99) (2016) 1–12.
[13] H. Hu, Y. Wen, D. Niyato, Spectrum allocation and bitrate adjustment for mobile social video sharing: a potential game with online qos learning approach, IEEE J. Sel. Areas Commun. PP (99) (2017) 1–14.
[14] S.M.A. Kazmi, N.H. Tran, W. Saad, Z. Han, T.M. Ho, T.Z. Oo, C.S. Hong, Mode selection and resource allocation in device-to-device communications: a matching game approach, IEEE Trans. Mob. Comput. PP (99) (2017) 1–14.
[15] D. Kempe, J.M. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proc. of ACM KDD, 2003.
[16] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? in: Proc. of WWW, 2010, pp. 591–600.
[17] J. Leskovec, A. Krevl, Snap datasets: stanford large network dataset collection Jun. (2016).
[18] H. Li, P. Li, P. Wu, Dominated competitive influence maximization with time-critical and time-delayed diffusion in social networks, J. Comput. Sci. (2018).
[19] J. Li, Z. Cai, M. Yan, Y. Liu, Using crowdsourced data in location-based social networks to explore influence maximization, in: Proc. of IEEE INFOCOM, 2016.
[20] Y. Li, C. Courcoubetis, L. Duan, Dynamic routing for social information sharing, IEEE J. Sel. Areas Commun. PP (99) (2017) 1–15.
[21] D. Lian, Y. Ge, F. Zhang, N.J. Yuan, X. Xie, T. Zhou, Y. Rui, Content-aware collaborative filtering for location recommendation based on human mobility data, in: Proc. of IEEE ICDM, 2015.
[22] B. Liu, H. Xiong, S. Papadimitriou, Y. Fu, Z. Yao, A general geographical probabilistic factor model for point of interest recommendation, IEEE Trans. Knowl. Data Eng. 27 (5) (2015) 1167–1179.
[23] W. Lu, C. Zhou, J. Wu, Big social network influence maximization via recursively estimating influence spread, Knowl. Based Syst. 113 (2016) (2016) 143–154.
[24] Z. Lu, Y. Wen, W. Zhang, Q. Zheng, G. Cao, Towards information diffusion in mobile social networks, IEEE Trans. Mob. Comput. 15 (5) (2016) 1292–1304.
[25] P.D. Meo, F. Messina, D. Rosaci, G.M.L. Sarne, 2d-social networks: a way to virally distribute popular information avoiding spam, in: Proc. of IDC, 2015.
[26] P.D. Meo, K. Musial-Gabrys, D. Rosaci, G.M.L. Sarne, L. Aroyo, Using centrality measures to predict helpfulness-based reputation in trust networks, ACM Trans. Internet Technol. 17 (1) (2017).
[27] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of online social networks, in: Proc. of IMC, 2007, pp. 29–42.
[28] D.K. Mohammad, R. Alireza, B. Negin, M. Mohammad, Minimum positive influence dominating set and its application in influence maximization: a learning automata approach, Appl. Intell. (2017) 1–24.
[29] J. Shang, S. Zhou, X.Li, L. Liu, H. Wu, Cofim: a community-based framework for influence maximization on large-scale networks, Knowl. Based Syst. (2016) 88–100.
[30] S.N. Swain, R. Thakur, S.R.M. Chebiyyam, Coverage and rate analysis for facilitating machine-to-machine communication in lte-a networks using device-to-device communication, IEEE Trans. Mob. Comput. PP (99) (2017) 1–15.
[31] K. Thilakarathna, A.C. Viana, A. Seneviratne, H. Petander, Design and analysis of an efficient friend-to-friend content dissemination system, IEEE Trans. Mob. Comput. 16 (3) (2017) 702–715.
[32] G. Tong, W. Wu, S. Tang, D. Du, Adaptive influence maximization in dynamic social networks, IEEE/ACM Trans. Networking 25 (1) (2017) 112–125.
[33] B. Viswanath, A. Mislove, M. Cha, K. Gummadi, On the evolution of user interaction in facebook, in: Proc. of WOSN, 2009, pp. 37–42.
[34] X. Wang, Y. Zhang, W. Zhang, X. Lin, Efficient distance-aware influence maximization in geo-social network, IEEE Trans. Knowl. Data Eng. 29 (3) (2017) 599–612.
[35] Z. Wang, L. Sun, M. Zhang, H. Pang, E. Tian, W. Zhu, Propagation- and mobility-aware d2d social content replication, IEEE Trans. Mob. Comput. 16 (4) (2017) 1107–1120.
[36] Y. Wen, P. Lei, W. Peng, X. Zhou, Exploring social influence on location-based social networks, in: Proc. of IEEE ICDM, 2014.
[37] M. Xiao, J. Wu, L. Huang, R. Cheng, Y. Wang, Online task assignment for crowdsensing in predictable mobile social networks, IEEE Trans. Mob. Comput. PP (99) (2016) 1–14.
[38] D. Yang, D. Zhang, V.W. Zheng, Z. Yu, Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns, IEEE Trans. Syst. Man, Cybern. 45 (1) (2015) 129–142.
[39] Y. Yang, X. Mao, J. Pei, X. He, Continuous influence maximization: what discounts should we offer to social network users? in: Proc. of ACM SIGMOD, 2016.
[40] Z. Yao, Y. Fu, B. Liu, Y. Liu, H. Xiong, Poi recommendation: a temporal matching between poi popularity and user regularity, in: Proc. of IEEE ICDM, 2016.
[41] Q. Yuan, G. Cong, Z. Ma, A. Sun, Nadia magnenat-thalmann: time-aware point-of-interest recommendation, SIGIR, 2013.
[42] J. Zhang, C. Chow, Y. Li, Igeorec: a personalized and efficient geographical location recommendation framework, IEEE Trans. Serv. Comput. 8 (5) (2015) 701–714.
[43] J. Zhang, F. Wang, K. Wang, W. Lin, X. Xu, C. Chen, Data-driven intelligent transportation systems: a survey, IEEE Trans. Intell. Transp. Syst. 12 (4) (2011) 1624–1639.
[44] S. Zhang, Y. Wu, X. Cai, Wechat traffic profile and impact on mobile networks, in: Proc. of 2014 ICCC, 2014, pp. 334–338.
[45] T. Zhou, J. Cao, B. Liu, S. Xu, Z. Zhu, J. Luo, Location-based influence maximization in social networks, in: Proc. of ACM CIKM, 2015.

Yongjian Yang received his B.E. degree in automatization from Jilin University of Technology, Changchun, Jilin, China, in 1983; his M.E. degree in computer communication from Beijing University of Post and Telecommunications, Beijing, China, in 1991; and his Ph.D. in software and theory of computer from Jilin University, Changchun, Jilin, China, in 2005. He is currently a professor and a PhD supervisor at Jilin University, the Vice Dean of the Software College of Jilin University, Director of Key lab under the Ministry of Information Industry, Standing Director of the Communication Academy, and a member of the Computer Science Academy of Jilin Province. His research interests include: network intelligence management, wireless mobile communication and services, and wireless mobile communication.

Yuanbo Xu received his B.Sc. degree and M.S. degree in the College of Computer Science and Technology, Jilin University, Changchun, China. He is pursuing his Ph.D. degree in Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Jilin University, Changchun, China. His research interests include applications of data mining, recommender system, and mobile computing.

En Wang received his B.E. degree in software engineering from Jilin University, Changchun, in 2011, his M.E. degree in computer science and technology from Jilin University, Changchun, in 2013, and his Ph.D. in computer science and technology from Jilin University, Changchun, in 2016. He is currently an associate professor in the Department of Computer Science and Technology at Jilin University, Changchun. He is also a visiting scholar in the Department of Computer and Information Sciences at Temple University in Philadelphia. His current research focuses on the efficient utilization of network resources, scheduling and drop strategy in terms of buffer-management, energy-efficient communication between human-carried devices, and mobile crowdsensing.

Kaihao Lou received his B.E.degree in Software Engineering from Jilin University, Changchun, Jilin,China, in 2017. He is currently a postgraduate student in Computer System Architecture from Jilin University, Changchun, Jilin, China. His current research focuses on Mobile Crowdsensing.

Dongming Luan received his B.E.degree in Software Engineering from Jilin University, Changchun, Jilin, China, in 2017. He is currently a postgraduate student in Computer System Architecture from Jilin University , Changchun, Jilin, China. His current research interests arise in the area of Mobile Crowdsensing .