Context-Aware Communities and Their Impact on Information Influence in Mobile Social Networks

Na Yu and Qi Han
Department of Electrical Engineering and Computer Science
Colorado School of Mines, Golden, CO 80401
{nyu, qhan}@mines.edu

Abstract—When mobile device users meet at a certain place, they may share information obtained from some other places. Since people are more likely to share information if they can benefit from the sharing or if they think the information is of interest to other people, there might exist communities where people share information more often with community members. The communities in mobile social networks represent real social groups where connections are built when people encounter. In this paper, we consider the location and time related to the shared information as the social context in mobile social networks. We propose context-aware community structure that groups people who are more likely to influence each other (i.e., share information with each other in certain contexts) into the same communities. Further, we provide a context-aware community-based user participation strategy in information influence that can reduce unnecessary influence cost. Our evaluation results show that the context-aware community structure is constructed with high internal pairwise similarity, reasonable average community size, and reasonable number of communities. Further, the community-based user participation strategy provides both high average influences and high influence efficiency.

I. INTRODUCTION

With the increasing popularity of mobile devices, mobile device users who are within the mobile device’s transmission range (e.g., via the WiFi or Bluetooth interfaces) have the potential of sharing information they have. This information sharing most likely occurs in mobile social networks where people meet each other at certain places despite their different daily mobility patterns. When they encounter, they can exchange information that they have obtained from other places. For example, in the campus scenario, students in the library might be informed of library events such as the arrival of new books or magazines. When they go out of the library and meet other students, they may share the library events with those students. Meanwhile, those students might come from other places such as recreation center or food court where they have been informed of a basketball game or free food, thus these events going on in the recreation center or food court can also be shared with those students just coming from the library. However, due to limited energy and storage of mobile devices, not all people are always willing to store the information they obtain and further share with others. Moreover, there might also be lots of redundant information sharing if people always exchange information with everyone they meet. Since people are more likely to share information if they can benefit from the information received from others or if they can provide useful information for others, there might exist special interest groups where people often share information with group members. In this following, “information influence” is used interchangeably to refer to information sharing among people in certain location and time contexts from the perspective of information owners. We use “information diffusion” to encompass information influence and also other users pulling information from information owners.

The special interest groups mentioned above are often referred to as “communities”. A community is a densely connected group of nodes within a social network such that connections between communities are sparse. Previous work has investigated different ways to discover communities and studied their patterns [1]. It has also been shown that joining a community provides an individual with tremendous benefits, thus individuals shall have incentives to join certain communities [2]. In mobile social networks, communities represent real social groups where connections are built when people encounter. The fact that members inside a community are more likely to influence each other has been exploited to improve information dissemination and query [3]. Leveraging these previous findings and insights, in this work we group people that are more likely to influence each other into the same communities, with the hypothesis that people will participate more efficiently in these communities in their future encounters. To validate this hypothesis, we then study how this grouping affects information influence among people in order to find efficient information diffusion strategies.

Existing work as detailed in section VI cluster mobile users into communities merely based on contacts that occur when users are close enough, more specifically when they are within WiFi or Bluetooth ranges. Contacts based communities only show that people inside the same communities are in contact more often or have longer contact durations. They are incognizant of where and when the contacts occur, hence losing critical information in mobile social networks: the social context (i.e., location and time) of these contacts. This is because if we can incorporate the history of people’s mobilities (i.e., where they have been and what information they have obtained before the contacts) into the contacts, we can construct more informed and meaningful communities than using pure contacts. Therefore, it is more reasonable to integrate social context with contacts to form context-aware communities. This work aims to demonstrate i) context-aware communities may
be formed to truly reflect how people interact in mobile social networks, and ii) context-aware communities are effective in information influence when user participations are taken into account. To the best of our knowledge, this work is the first that studies the impact of context-aware communities on information influence in mobile social networks.

Our performance evaluation shows that the constructed context-aware community structure has high average internal pairwise similarity, reasonable average community size, and reasonable number of communities. In addition, the community-based information diffusion provides high average influenced users, high average influences to users, and high influence efficiency. We use information and event interchangeably in the following.

II. Problem Description

In order to represent the social context with location and time, we adopt the concept Point of Interests (POIs) [4]. POIs are popular locations with frequent user stays.

We assume that each PoI only has one type of event, and the event is updated periodically and broadcasted within the PoI region via a central server. In the area of interest, there are multiple POIs, each has a center location, a region size, an event update period $T_e$, and a set of event updates (i.e., pieces of information). Each event update is associated with a unique update id, its PoI location, and its update time. There also exist a number of mobile users, each has a unique user id, a mobility profile with a series of contacts and location visits, a set of event updates with the time when the user was influenced, and a user influence lifetime $T_l$. For instance, in the campus scenario, the recreation center is broadcasting the event updates of games every two hours, and a student who has received an event update about a basketball game is willing to store this event update for five hours in order to share with other students he might meet. In this case, the event update period for the recreation center is $T_e = 2$ hours, and the user influence lifetime for the student is $T_l = 5$ hours. Without loss of generality, we assume that $T_e$ is the same for all POIs, and $T_l$ is the same for all users. As shown in Fig. 1, there are four POIs (represented using small rectangles) and 20 mobile users (represented using small circles). Mobile users visit the POIs at different times and are influenced by the PoI events during their stays in the POIs, then they may choose to further influence other mobile users with the PoI events if they meet other users outside the POIs anytime before the expiration of $T_l$. Once an event reaches a user outside its PoI, the event may be further forwarded to other users within $T_l$. So the event influence is continued in a ripple carry fashion.

The problem at hand is to form communities that will facilitate information influence with the goal of reaching more relevant users without introducing significant overhead. This problem is broken down into two issues: (1) constructing context-aware communities; (2) studying the impact of context-aware communities on information influence.

III. Context-Aware Community Structure

In this section, we present how to construct context-aware communities based on users’ previous mobility patterns and contact history. We first introduce directed weighted influence graph. This graph embeds all the contacts users have as well as context information of these contacts. We then apply the Directed Clique Percolation Method (CPMd) [5] to construct communities regrading each PoI.

A. Influence Graph

An influence graph is represented by $G = (V, E, T_l)$, where $V$ is the set of users, $E$ is the set of directed weighted influences between any pair of users, and $T_l$ is the user influence lifetime. Since a user has the potential to influence another user with the already influenced PoI events within $T_l$ via contacts, there is a directed edge in the influence graph from one vertex to another if there exists a potential influence. The weight of an edge from vertex $i$ to vertex $j$ is $w_{i,j} = (I_{P_1}, I_{P_2}, ..., I_{P_l})$, where $P_i$ is the $i$th PoI, and $I_{P_l}$ is the number of potential influences regarding the $i$th PoI within $T_l$ at each contact. If the contacts are recorded periodically, the number of contacts can also reflect the total contact duration, then we do not need to include the contact duration in the weight. Although the contact duration is not explicitly reflected in the influence graph, it does affect the total number of the information influences over time. In other words, its effects will certainly be reflected in the constructed community structure.

User mobility profiles: In order to construct the influence graph, we need to generate user mobility profiles first. A typical mobility profile of a user includes when the user is inside which PoI and outside of any PoIs and when the user is in contact with other users. Fig. 2 is an example mobility profile for “User 1”, where the solid boxes tagged with POIs show the durations that “User 1” is inside the PoIs, and the connections with other users show the contacts happen with those users at that time.

Constructing the influence graph from user mobility profiles: Assume there are $m$ users in total and $n$ of them have their mobility profiles available. Also assume that there are $s$ PoI visits in total for all the users, and there are $c$
contacts in total among all the users. Then, the influence graph construction follows these steps: (1) Do a merge sort on the $n$ user mobility profiles based on time, so the integrated user mobility profile has $2s + c$ moments of user actions. As shown in Fig. 3, at each moment, the user action is represented by "$U_i P_i S$" meaning User $i$ enter PoI $l$, "$U_i P_i E$" meaning User $i$ exit PoI $l$, or "$CU_i i_j$" meaning User $i$ and User $j$ are in contact. (2) Scan the integrated user mobility profile to find all the bidirectional potential influences between any pair of users. When a contact occurs, the potential influence from one user to another is added by the times of each PoI the user has visited (i.e., count the number of moments with user action "$U_i P_i E$") within $T_i$. However, if the contact occurs inside a PoI (i.e., the users have not yet exited the PoI), then the users are considered self-influenced (i.e., the influence regrading the current PoI is only used for future influences to other users). (3) put all the $m$ users as the vertices and put the count of potential influences per each PoI from one user to another as the weight of the directed edge on the influence graph.

---

**B. Context-Aware Communities**

Once we have constructed the influence graph, we can construct communities. There are various ways to discover communities, most of which are based on either centralized modularity optimization approach [6] or local clique percolation approach [7]. We choose to apply the CPMd algorithm [5] to find communities. The CPMd algorithm is the directed graph version of the well known Clique Percolation Method (CPM) [7] where modules are defined as $k$-cliques (complete subgraphs of size $k$). The extension of CPMd to CPM is that it uses directed $k$-cliques to discover modules with directed links. Similarly, CPMd can also use a weight threshold to indicate the existence of directed links as CPM does. The advantages of using a clique percolation based method are i) it is local that it does not have resolution limit as centralized modularity optimization approaches have, and ii) the definition of modules based on $k$-cliques is actually based on link-density, so that connections are more concentrated inside communities, and iii) it allows overlaps between communities.

To make the communities context-aware, we find communities regarding each PoI by running the CPMd algorithm on the sub graphs of the influence graph. The sub graphs only contain the links whose weight regarding this PoI is greater than the weight threshold used in the CPMd algorithm. In this way, the context-aware community structure is constructed with PoI related communities, represented as $C_s = (PoI_{1} : (C_1, C_2, ...) , PoI_2 : (C_1, C_2, ...))$. The significance of the context-aware community structure is that users that are more likely to influence each other with PoI-specific information are grouped together, so users can determine their communities based on the contexts of information they receive. In addition, communities can overlap so that each user can belong to multiple communities. Note that, the context-aware communities are constructed in a centralized server based on the profiles collected from all the user devices. The constructed communities are then distributed to all the user devices for future usage.

---

**IV. COMMUNITY-BASED INFORMATION DIFFUSION**

A diffusion model in social networks typically mines top-$K$ influential nodes and then targets them as initially influenced nodes to maximize the influences [8]. However, in mobile social networks, the initially influenced nodes (i.e., mobile users) are determined by their locations. If a mobile user is close to the source of an event, then he is initially influenced by this event. Therefore, in this paper, we assume that users who are inside the PoIs are initially influenced by the PoI event updates. Then, the diffusion model boils down to answering the question of how the initially influenced users influence other users with events they received from the PoI. This involves user participations in information influence.

We propose a community-based user participation strategy. Users might have certain selfishness in mobile social networks and they might participate in the information influence process with incentives. Existing work [9] has presented the impact of different distributions of altruism on the throughput and delay in mobile social networks, where both uniform and community-biased traffic patterns are used for evaluation. The observation is that mobile social networks are very robust to the distributions of altruism due to the nature of multiple paths. Based on this observation, we can simply assume that users have the same participation strategy with certain selfishness in the diffusion model.

In our model, users have higher probabilities to influence others within the same community, while have lower probabilities to influence users in other communities. Thus, the user participation incentive can be modeled by two parameters ($\alpha, \beta$) to represent the intra-community and inter-community influence probabilities. For a complete community-based participation strategy, we assume that $\alpha = 1$ and $\beta = 0$. When two users are in contact, they first exchange the information summary (i.e., a list of active event updates with PoIs and update versions that the users have), then they decide which pieces of information to share with each other.

Recall that users stay in the PoI regions are originally influenced by the event updates, and users outside the PoI regions are influenced by users who are already influenced by the event updates via contacts, the diffusion model using the user participation strategy will tell which users are influenced...
by the event updates from each PoI and which event updates are propagated to other users.

V. PERFORMANCE EVALUATION

To evaluate the performance of our proposed context-aware community structure and also its impact on information influence, we use empirical datasets that have both location and contact information. We divide the dataset into two parts: the training set is used to construct community structure and the test set is used to study the impact of the communities on information influence.

A. Empirical Dataset

We evaluate our approach using the UIM dataset [10], the most recently released dataset where we can obtain both meaningful locations (i.e., PoIs) and user contacts. The UIM dataset contains both the WiFi traces and Bluetooth traces for a number of people around UIUC campus. Locations can be obtained from the WiFi APs in the WiFi traces, while contacts can be obtained from the Bluetooth traces which are more precise than WiFi based contacts. In the dataset, there are traces of 27 users that are collected from 03/01/2010 to 03/19/2010. We first analyze the WiFi traces and Bluetooth traces respectively as follows.

UIM WiFi Traces: The WiFi traces are collected every 30 minutes. There are 5614 WiFi AP MACs appeared in the WiFi traces of all the 27 users we consider, and the number of valid WiFi AP MACs which are not occasionally occur is 985 (i.e., occur at least 27 times in the entire data set in our evaluation). Then, we apply the star clustering algorithm [11] on the WiFi APs, we get about 200 clusters, each cluster represents a location. After merging the clustered locations and eliminating short duration ones (i.e., duration less than 10 minutes in our evaluation), we get a number of most popular locations, each of which might be considered as a PoI. The exact number of PoIs from the most popular locations will be decided later in forming the context-aware communities.

UIM Bluetooth Traces: The Bluetooth traces are collected every 1 minute. In addition to the Bluetooth MACs of the 27 users we have previously considered, there are another 7671 Bluetooth MACs appeared in the Bluetooth traces of these 27 users, and the number of Bluetooth MACs that have frequent contacts are 123 (i.e., at least 100 contacts in the whole dataset in our evaluation). Therefore, we also take these 123 bluetooth MACs into consideration, and then we have 150 users in total in our evaluation.

By combining the WiFi trace and Bluetooth trace for each user, we can get the user mobilities with both locations and contacts. Then, we construct the user mobility profiles as shown in Fig. 2 with the most popular locations and also user contacts we have discovered from the dataset. We combine the user mobility profiles into an integrated user mobility profile, and divide it into two parts—the training set and the test set. The training set contains the user mobility records from 03/01/2010 to 03/10/2010, and it is used to construct the context-aware communities. The test set contains the user mobility records from 03/11/2010 to 03/19/2010, and it is used to test the information influence based on the context-aware communities constructed from the training set.

B. Quality of the Constructed Community Structure

Since the influence graph we have constructed from the training set has low degrees for most users, we choose the weight threshold as 1 and the clique parameter $k = 3$ in the CPMd algorithm to form the context-aware communities in the contexts of PoIs. In fact, existing work has observed the same $k$ for another dataset [12].

In the literature, there is no standard criteria for evaluating community structures. The most adopted evaluation metrics are modularity Q [6], [13]–[15] and normalized mutual information [16]–[19]. However, the modularity Q has different variations in different community detection algorithms that support overlapping communities. The normalized mutual information comes from information theory uses ground truth as the base line, thus it is not applicable to our work where ground truth is unknown. In order to exploit the local feature of the CPMd algorithm, we adopt a metric—Internal Pairwise Similarity (IPS) used in a greedy local optimization based community detection approach [20]. IPS measures the average similarity between the friendship declarations of pairs within the community. It is applicable to evaluating community detection algorithms supporting overlapping communities without ground truth as the baseline.

We use average IPS, average community size, and average number of communities per PoI to measure the quality of the constructed communities. We specifically evaluate the impact of the user influence lifetime $T_i$ on these metrics. The evaluation results are shown in Fig. 4, Fig. 5, and Fig. 6, respectively. Instead of using a threshold to decide the number of PoIs from the most popular locations, we identify the reasonable PoIs with considerable IPS (>average IPS), community size (>average community size), and number of communities (>average communities per PoI). The impact of $T_i$ on the number of PoIs that have been found is shown in Fig. 7, and only these PoIs will be further considered in the information influence process.

![Fig. 4. Average IPS](image1.png) ![Fig. 5. Average Community Size](image2.png)

Fig. 4 shows that average IPS of context-aware communities constructed from the influence graph ranges from 0.04 to 0.055, while our experiment on the average IPS of randomly connected graph is about 0.009 which is consistent with the result of connected random groups for 150 nodes in [20]. Thus, the average IPS of context-aware communities is much
calculate the number of event updates.

To the calculation. Event update period among users, the original influences from the PoIs are included in the influence graph which represents the potential influences. Based on the influence graph and the user participation strategy, we can calculate the exact total number of information increases, there is more potential information influence. Can be constructed. In other words, when influence lifetime increases, there is more potential information influence.

**C. Performance of Information Influence**

Based on the influence graph and the user participation strategy, we can calculate the exact total number of information influences in the entire diffusion process. In addition to the influence graph which represents the potential influences among users, the original influences from the PoIs are included in the calculation. Event update period $T_e$ is also used to calculate the number of event updates.

We compare our community-based information influence with the following two strategies.

- **Barter-based participation [21]**: Each user provides as much information for an encountered user as the encountered user provides him. Thus, the incentive is modeled based on pair-wise information exchange among users.
- **Always Influence**: Users always share information with everyone that is in contact. It utilizes every contact opportunity to maximize the information influence, but causes lots of redundant information sharing. We use it as the base line.

We use the context-aware communities constructed from the training set for the community based user participation strategy to show the information influence results in the test set. We compare the three influence strategies using the following metrics: average influenced users per PoI, average unique event updates per PoI received by per user, and influence efficiency (i.e., the ratio of unique event updates received by the users to the total event updates that have been forwarded during the diffusion). We observed that the performance has similar trends when varying the event update period $T_e$. Here we only show the results of varying the user influence lifetime $T_l$, considering $T_e = T_l$. The impacts of $T_l$ are shown in Fig. 8, Fig. 9, and Fig. 10, respectively.

Fig. 8 shows that the average influenced users per PoI of the Community-based user participation strategy is always higher than the Barter-based user participation strategy. Although it is a little lower than the Always Influence strategy, the difference becomes smaller when $T_l$ increases. In addition, it is obvious that the higher $T_l$, the higher average influenced users per PoI.

Similarly, Fig. 9 shows that the average unique event updates per PoI received by per user of the Community-based user participation strategy is also always higher than the Barter-based user participation strategy. Although it is a little lower than the Always Influence strategy, the difference is negligible when $T_l$ is relatively low. It is also obvious that the higher $T_l$, the higher average unique event updates per PoI received by per user.

Finally, Fig. 10 shows that the influence efficiency of the Community-based user participation strategy is always higher than both the Barter-based user participation strategy and the Always Influence strategy. Although the result of the Barter-based user participation strategy increases relatively fast when $T_l$ increases, and it exceeds the Always Influence strategy when $T_l$ becomes larger than 6 hours, it still cannot exceed the Community-based user participation strategy within the range of $T_l$ we consider. Moreover, incorporating the previous results of the average influenced users per PoI and the average unique event updates per PoI received by per user, the Community-based user participation strategy definitely outperforms the Barter-based user participation strategy.

VI. RELATED WORK AND CONCLUDING REMARKS

Studying the detection and evolution of community structures is a popular topic in traditional social networks. For instance, [8] captures and identifies interesting events from non-overlapping snapshots of interaction graphs, and uses these events to characterize complex behavioral patterns of individuals and communities over time. [22] further studies the detection and evolution of communities in a uniform process, where the community structure at a given time step is determined by both the observed networked data and the prior distribution given by historic community structures.

Although community structures play an important role in social network analysis [1], research on communities in mobile social networks is still largely incomplete. Communities in mobile social networks have been formed based on either only contacts among users [23] or only user proximity [25]. The concept of social context for pervasive social computing has been proposed [24] before but not been used in community construction. Although location and context information has
been considered in mobile ad hoc networks [4] [23], no community structure has been studied.

User participation has been used for data dissemination [26] and information query [21] in mobile ad hoc networks without considering social communities. Existing work using game theory based user participation to identify communities in social networks [2] allows users to select communities to maximize their gains, but it is focused on user benefits rather than the location and time intrinsic of mobile social networks.

Influences among mobile users instead of only contacts have been used to detect communities in mobile social networks [3], but user participation is not taken into account.

This work integrates social contexts into contacts in mobile social networks to construct context-aware communities, which have been shown to have good community properties and benefit information influence. The context-aware community structure aims at grouping users that are more likely to influence each other, taking into account both their contacts and social contexts, and it can be further used to design efficient information diffusion services in mobile social networks. Our future work includes improving the algorithm of constructing context-aware communities, providing better user participation strategies incorporating game theory (e.g., reputation-based or credit-based strategies), considering other influencing factors in community-based information influence (e.g., relevancy of the information to specific users based on historical information), using more empirical datasets to evaluate the community structure and information influence, and designing community-based message forwarding and information query protocols.

ACKNOWLEDGEMENT

This project is supported in part by NSF grant CNS-0915574.

REFERENCES