Privacy Preserving Social Tie Discovery Based on Cloaked Human Trajectories

Qinli Kou
shuxi_kql@163.com

Ye Tian
humanty@163.com

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications
Beijing 100876, China

Zheng Song
Computer Science Department, Virginia Tech
Blacksburg, U.S.
sonyyt@gmail.com

Edith Ngai
Department of Information Technology, Uppsala University
Sweden
edith.ngai@it.uu.se

Wendong Wang
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications
Beijing 100876, China
wdwang@bupt.edu.cn

ABSTRACT

Discovering social connections of people has become a flourishing research topic considering the rich social information inferable from human trajectories. Existing social tie detection methods often require exact locations of users, which cause serious privacy concerns. Although cloaking is a common technique for location anonymization, it has rarely been applied in social tie detection due to the potential loss of significant location information. In this paper, we propose a semantic tree model for social tie detection, which supports different levels of privacy preserving and allows better understanding of location content of the cloaking regions. We propose a novel algorithm that can infer the social ties between users using only their cloaked trajectories without exposing their exact locations. We model the obscured regions generated by the cloaking algorithms in a semantic region tree and infer the similarity between two users based on their temporal and spatial relations in the tree. We evaluate our proposed approach using real trajectory dataset and show that our algorithm can identify social ties successfully with 15% higher accuracy compared with existing approach.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: Metrics—*Privacy*

Keywords

Human Trajectory; Privacy Preserving; Social Tie Discovery

HotPOST'15, June 22, 2015, Hangzhou, China.

DOI: http://dx.doi.org/10.1145/2757513.2757522.

1. INTRODUCTION

Location-based social network applications have become highly popular over the world. Increasing number of people are using GPS-enabled devices to log their outdoor locations and activities [13]. It is also getting very common for people to share information about their current locations and activities with their friends. This shared information is expected to give significant impact in social networks [12]. Recent research has shown that the mobility patterns of individuals may be shaped by their social relationships. Likewise, human trajectories could be used to infer social ties of people.

Inferring social ties has become an important topic in social network analysis. It has been proven beneficial in many different ways [5], including for link prediction [15], product recommendation and community discovery [8]. Here, social ties are usually inferred by the similarity of individuals in both spatial and temporal dimensions according to their location history. Although location information is very useful, an untrusted server may save users' location data and leak them to third parties [7] that create privacy risks. To address this problem, different approaches have been proposed to support location-based services while protecting the location privacy of mobile users [7],[6],[9].

K-anonymity is an important measure for preventing the disclosure of personal data [6], which means that the user's location is indistinguishable from at least k-1 other users. To achieve k-anonymity, a location-based service (LBS) related query is submitted to the LBS server via a centralized location anonymizer, which enlarges the queried location into a bigger region covering at least k-1 other users geographically. This process is also called "cloaking" as it constructs a spatial cloak around the user's actual location [6]. The region covering k users is called *cloaking region* and a trajectory formed by a sequence of cloaking regions is called a *cloaked trajectory*. Fig.1 demonstrated the procedure of generating cloaked trajectories.

As cloaking is a common technique for privacy protection, discovering social ties from cloaked trajectories becomes necessary to understand the connections of people under privacy preservation. Tan et al. [11] have predicted social ties based

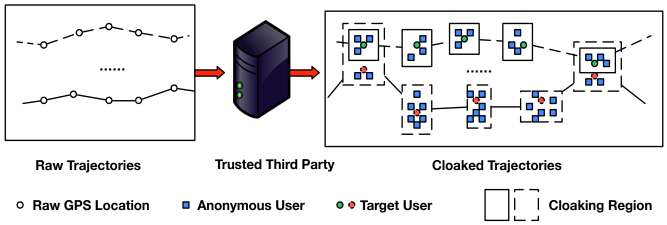


Figure 1: Cloaked Trajectories.

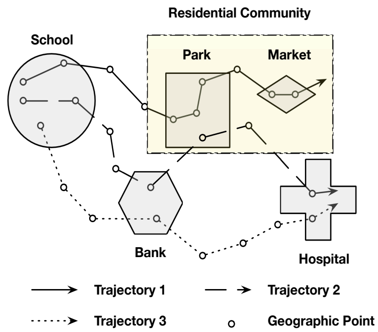


Figure 2: Containment relationship of areas.

on protected location data from k-anonymity cloaking, but the results has suffered from low accuracy. We observe that the existing approach has considered only overlapping area of cloaking regions, but has ignored other significant factors. In addition, different privacy levels (k) of users may lead to cloaking regions with different sizes. In Fig.2, the cloaking region would be marked in the *residential community* for a user with high privacy level, while it would be marked in the *park* (contained by *residential community*) for a user with low privacy level. Hence, it is crucial to understand the hierarchical relations of cloaking regions.

In this paper, we transform the cloaking regions into the *semantic regions*, which imply the semantic meaning of a region. We represent the containment relationship of semantic regions in a *hierarchical semantic tree*, where each of the nodes corresponds to a *semantic region*. Next, we propose a novel algorithm to infer social ties with hierarchical semantic tree. Based on this, the probability of the existence of social ties between users could be measured by a *similarity score*. Finally, we evaluate our proposed algorithm using a real dataset collected from a location-based social network.

The contributions of this paper is summarized as follows. First, we propose a model transform cloaking regions to semantic regions and construct a hierarchical semantic tree. Second, we propose a novel algorithm to discover social ties from cloaked trajectories with hierarchical semantic tree model. Finally, we use real Gowalla dataset [1] to evaluate the performance of our proposed scheme. The results has shown 15% higher accuracy on social tie discovery over existing algorithms.

The rest of this paper is organized as follows. Section II reviews the related researches. Section III introduces the model we proposed. Section IV gives the full details of our algorithm. Section V evaluates our approach on a real-world dataset and reports on evaluation results. Section VI concludes our work.

2. RELATED WORK

The relation between socials ties and human mobility has been widely explored in recent years [13],[12],[2],[10]. Wang et al. [12] have revealed a close correlation between human trajectories and social ties. Similarly, Cho et al.[2] have studied the relation between social ties, and human geographic and temporal dynamics. They found that the similarity of human trajectories is a strong indication of a tie in the social network. In addition, a number of studies on semantic trajectory data mining have appeared in the literature [14]. For example, Xiao et al. have used semantic location histories (SLH) to estimate the users' similarity [13]. A semantic trajectory consists of a sequence of locations labelled with semantic tags (called semantic locations) to capture the landmarks on the trajectory. These semantic locations often imply the activities carried out by human beings. For example, Alvares et al. have explored the semantic trajectory patterns based on the mobility histories of mobile users. Firstly, they identified the stops of each trajectory and mapped these stops to the semantic landmarks. Then, they applied a sequential pattern mining algorithm to obtain the frequent semantic behaviours of the mobile users. Ying et al. further proposed a method to predict the next location of the user based on the geographic and semantic features of his trajectory [14]. Different from previous work, we extend the semantic trajectory concept and explore the hierarchical relationships of semantic regions to achieve more accurate social tie discovery results in this work.

Although location information is very useful, user's location privacy has received considerable concerns such as in Location-based Services (LBSs). Different location privacy protection mechanisms have been proposed. Among them, location perturbation and obfuscation have been used most frequently [7],[3]. The concept of k-anonymity for location privacy was introduced by Gruteser and Grunwald [4]. It guarantees that the target object is indistinguishable from the other k-1 objects in a set (a group of mobile users in our case). The authors have designed an adaptive interval cloaking algorithm, which generates spatio-temporal cloaking areas containing at least k_{min} users. In k-anonymity cloaking, a larger k provides higher level of location privacy, and vice versa. Different from existing work, we support social tie detection of individuals based on their cloaked trajectories resulted from the k-anonymity cloaking method. This approach can preserve different levels of privacy for users and support social tie detection without invading their location privacy.

3. THE PROPOSED MODEL

This section shows the procedure of modeling the trajectories of individuals as hierarchical semantic tree in order to explore their social ties.

3.1 System Overview

In the system, the trusted server uses rectangles to indicate cloaking regions for convenience of calculation. Since different users may expect different levels of privacy (indicated by k), the size of the cloaking regions are usually different. We denote the set of users by $\mathcal{U} = \{U_i : i = 1, 2, ..., U\}$ and their trajectories by $\mathcal{T}_u$, where $u \in \mathcal{U}$. Each trajectory $t_u \in \mathcal{T}_u$ is composed of a set of rectangular regions and user's time of entering and leaving this region. Let $\{R_p, E_p, L_p\}$ be a specific spatial-temporal portion p of a trajectory t_u,

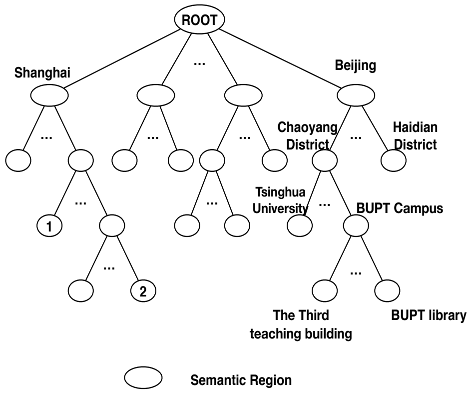


Figure 3: Hierarchical Semantic Tree.

where E_p and L_p are the time when the user arrived and left region R_p. Let $\mathcal{R} = \{R_i : i = 1, 2, ..., R\}$ be the set of cloaking regions. The location of a cloaking region can be indicated by the x,y-coordinates of its top-left and bottom-right corners.

3.2 Semantic Regions

Since the semantic-based trajectories imply individuals' personal interests and preferences to a great extent, we transform the cloaking regions on trajectories to semantic regions. Some existing applications are available to provide the semantic information of accurate GPS locations. However, a cloaking region could not be regarded as an accurate location with exact latitude and longitude.Thus, we cannot transform a cloaking region to a semantic region by using Google Maps API directly.

To transform a cloaking region to a semantic region, we select several locations in the region as a sample set $\mathcal{L}$, and $\mathcal{S}$ be the set that represents all the semantic meanings associated to the locations in $\mathcal{L}$. Next, we compare the semantic meanings of all the sub-regions in $\mathcal{S}$. There are two cases: 1) all of the locations in $\mathcal{L}$ carry the same semantic meaning, denoted by **A**; 2) there exists some differences among semantic meanings in $\mathcal{S}$. In the first case, the semantic meaning of the cloaking region is **A**. The second case shows that there are more than one semantic meaning in the cloaking region, to say, none of the semantic meanings could satisfy the anonymity level of this user at the moment. So a bigger region covering all of the semantic meanings is selected as the semantic region.

3.3 Hierarchical Semantic Tree

As the anonymity levels of individuals are different, the size of semantic regions could be different too, which leads to exist containment relationship among semantic regions. Intuitively, a person with a larger k is more likely to have a larger semantic region, and it is possible to contain a semantic region belongs to a person holding a smaller k. Due to the containment relationship, which is not only semantic related but also located related, hierarchical structure is produced. We propose a hierarchical semantic tree to express the containment relationship, and each of the nodes is associated to one of the semantic regions. The structure of the hierarchical semantic tree is shown in Fig.3. For example, the region *BUPT Campus* in the tree contains two children nodes, *BUPT library* and *The third teaching building*. Also, *BUPT Campus* and *Tsinghua University* may be located in adjacent nodes included in the node of *HaiDian District*. Due to the changing population density of semantic regions, a user with a constant privacy level may not stay in regions at the same level all the time. For example, a user in region 1 may move to region 2 in the next timestamp (see Fig.3).

We consider that it is more meaningful to compare the regions with similar time when calculating the similarity of two trajectories, thus we define the *pair regions*.

DEFINITION 3.1. (**Pair regions**) *For different users $\mathcal{U}_1$ and $\mathcal{U}_2$, two trajectories t_1 and t_2 are from $\mathcal{U}_1$ and $\mathcal{U}_2$ respectively. Two spatial-temporal semantic portions p_i and p_j belong t_1 and t_2 respectively, they could be defined as pair regions if and only if they satisfy temporal constraint $TimeDiff$. We define that the time difference between the two regions in* pair regions *should meet the requirements $TimeDiff(E_{p_i}, E_{p_j}) \leqslant \delta_t$ and $TimeDiff(L_{p_i}, L_{p_j}) \leqslant \delta_t$, where δ_t is a given temporal constraint implying the maximum time difference.*

We set δ_t is equal to one hour in our work. When defining $TimeDiff$, we should notice two issues:1) We should guarantee that the trajectories are meaningful within a reasonable time span. It means that the time span between the compared trajectories should not be too long;2) On the premise of satisfying the time span, we consider that the $TimeDiff$ is not related to the date, and just referred to the time of one day. It is used to ensure that the regions with similar pattern on different days should still be considered as pair regions.

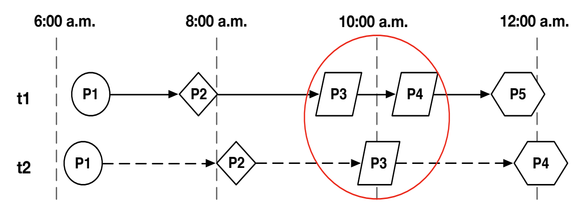


Figure 4: Pair regions in two trajectories.

Fig.4 illustrates the meaning of *pair regions*, and pair regions are denoted with the same shape. We notice that it is possible for a region on a trajectory to have more than one paired regions with another trajectory. It is because several regions may satisfy the temporal constraint δ_t, just like regions p_4, p_5 and p_6, which are circled in Fig.4. When calculate the similarity of two trajectories, we measure all of the available pair regions on them.

3.4 The Relationship Between Trajectories' Similarity and Social Ties

In this paper, we consider that the similarity of trajectories could reflect the similarity of individuals, which means that the similarity implies the behaviours and mobility patterns of individuals. The similarity is depicted by a *similarity score*. Intuitively, social ties may exist between users with high similarity score. We suggest hierarchical semantic tree as a feasible model to infer social ties. The reason is twofold. First, it is more likely for people to have social ties if their mobility patterns are matching well. They may share same semantic region patterns in the hierarchical semantic tree. Second, even though their semantic regions are not the same, there could still be social ties between them if their semantic regions are close in the hierarchical semantic tree.

4. THE PROPOSED SOCIAL TIE DISCOVERY ALGORITHM

The possibility of existing social ties mainly depends on the similarity of their trajectories. We calculate the similarity based on the pair regions on two trajectories selected before. Then, some important factors, such as the lowest common ancestor node and the length of the shortest path, tied to the similarity based on the nodes of the pair regions are obtained by searching the tree. The similarity ties to three aspects: 1)First, we should consider the level of the lowest common ancestor node of the pair regions in the hierarchical semantic tree. Lower level may lead to greater impact on similarity; 2)Second, the shortest length path between two semantic regions in the hierarchical semantic tree should also be taken into consideration. Intuitively, two nodes are closer in the tree if they are closer at physical distance; 3)At last, we consider that it is more influential in the prediction if a user's semantic region belongs to a lower level in the hierarchical semantic tree, that is to say, has a more accurate region.

We define the following metrics to measure the properties and relation of the semantic regions.

1. $len(R_i, R_j)$: the length of the shortest path from region R_i to region R_j in the hierarchial semantic tree.
2. $lca(R_i, R_j)$: the lowest common ancestor node of R_i and R_j.
3. $depth(R_i)$: the length of the path to the region R_i from the root node in the tree, and $depth(root) = 1$.
4. $deep\_max$: the max $depth$ among the nodes in the tree.
5. $sim(R_i, R_j)$: the similarity between cloaking regions R_i and R_j.
6. $inf(R_i)$: the influence of R_i, which is decided by the level of region R_i. Intuitively, $inf(R_i)$ will monotonically increase with respect to $depth(R_i)$.

According to the discussion above, the approach we proposed to measure the similarity of trajectories could be illustrated in the following equation:

$$sim(R_i, R_j) = e^{-\alpha \times len(R_i,R_j)} \times \{inf(R_i) \times inf(R_j) \times \frac{e^{\beta \times depth(lca(R_i,R_j))} - e^{-\beta \times depth(lca(R_i,R_j))}}{e^{\beta \times depth(lca(R_i,Rj))} + e^{-\beta \times depth(lca(R_i,Rj))}}\} \quad (1)$$

As shown in Eq.1 above, we know that formula 1 will monotonically increase with respect to $depth(lca(R_i, R_j))$, and then decrease with $len(R_i, R_j)$. Besides, α and β are parameters scaling the contribution of the length of the shortest path and the level of the lowest common ancestor, respectively. The settings of these parameters in the proposed algorithm should be adapted according to the performance with different combinations of α and β. Also, the $inf(R_i) \times inf(R_j)$ represents the mutual influence contributed to the similarity. The procedure of calculating the similarity of two trajectories is illustrated in **Algorithm 1**.

Algorithm 1 The algorithm of calculating two trajectories' similarity.

Input:
The hierarchical semantic tree, HST;
The set of pair regions of trajectories t_1 and t_2, $\mathcal{PR}$;
Output:
The similarity of trajectories t_1 and t_2, Sim;

```
Sim = 0;
Num = |PR|;
for all (R_i, R_j) ∈ PR do
    LCANode = searchLCA(R_i, R_j, HST);
    SP = ShortestPathLength(R_i, R_j, HST);
    LCALevel = LocatedLevel(LCANode, HST);
    Level1 = LocatedLevel(R_i, HST);
    Level2 = LocatedLevel(R_j, HST);
    Inf1 = Infulence(Level1);
    Inf2 = Infulence(Level2);
    SimPair = e^(−α×SP) × Inf1 × Inf2 × (e^(β×LCALevel) − e^(−β×LCALevel)) / (e^(β×LCALevel) + e^(−β×LCALevel));
    Sim = Sim + SimPair;
end for
Sim = Sim/Num;
return Sim;
```

Using Fig.5 as an example, hypothetically, the semantic region of a user ($Jack$) is covered by region A and the semantic region of another user ($John$) is in region B at the timestamp m_1, while region C in the second level is their lowest common ancestor, and the length of shortest path is equal to 5. Then we can get the similarity of the first pair of regions. In the next timestamp m_2, $Jack$ and $John$ arrive at region D and E respectively. The second score can also be calculated based on Eq.1 above. After calculating all pairs of regions in the trajectories, we set the vector $\overrightarrow{\mathbf{s}}$ as the similarity vector to record the similarity of each pair of regions, denoted by $\overrightarrow{\mathbf{s}} = (simPair_1, ..., simPair_n)$. Finally,

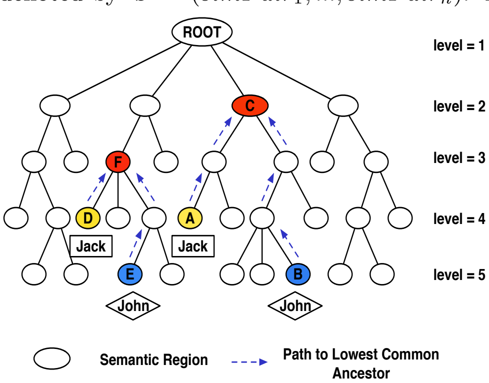


Figure 5: Calculate the similarity of regions.

the similarity score of the two trajectories (t_1, t_2) can be calculated by:

$$Sim(t_1, t_2) = \frac{1}{|\overrightarrow{s}|} \times \sum_{i=1}^{n} simPair_i. \quad (2)$$

If there are more than one trajectories in individuals, we compare any two trajectories from two individuals each time. And a threshold δ_s is set to decide whether there exists social ties. We select δ_s by comparing different performances under different thresholds, and the one achieving the best performance in terms of F-measure is selected as the proper threshold δ_s.

5. PERFORMANCE EVALUATION

5.1 Setup

We assess the proposed algorithm by using a real-world dataset from Gowalla [1]. The friendship network consists of 196,591 nodes and 950,327 edges. The dataset has collected a total of 6,442,890 check-ins from February 2009 to October 2010. Users, whose trajectory contains too few records are ignored. For those retained users, we randomly set k for each user to represent his privacy protection level.

Precision, recall and F-measure are adopted as main evaluation index, which are defined by Eq.3, 4, and 5.

$$Precision = \frac{p^+}{p^+ + p^-}, \tag{3}$$

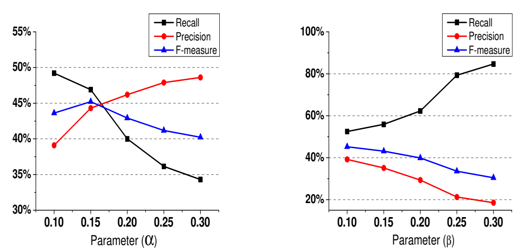
$$Recall = \frac{p^+}{|R|}, \tag{4}$$

$$F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{5}$$

where p^+ and p^- indicate the number of correct predictions and incorrect predictions of the existence of social ties respectively. $|R|$ indicates the total number of social tie records in the social network.

5.2 Sensitivity Tests

The sensitivity tests evaluate our approach under various parameter settings. In order to evaluate the impact of different factors on the performance of our algorithm, we evaluate the performance by varying one parameter and fixing the others.



(a) Performance changing with α (b) Performance changing with β

Figure 6: Performance in various settings.

As shown in Fig.6, the precision rate of our method is improved and the recall of our method is reduced when α is increased. We could also observe that the precision of our method is reduced and the recall is improved when β increases. Besides, we observe that β leads to a greater change of precision rate and recall rate compared with α. As depicted in Eq.1, α is related to the length of the shortest path between two semantic regions, while β is related to the level of the lowest common ancestor node between two semantic regions. It means that the level of the lowest common ancestor has a greater impact on the results. Thus, it is suggested that the containment relationship among semantic regions is more influential on the similarity of individuals.

In order to explore the best combination of α and β, more experimental results with various parameter settings are reported. At the same time, the relationship among precision rate, recall rate and F-measure could be shown via a composite curve (PRF curve) in Fig.7. Although, there is no causal relation between precision and recall, they are regarded as two inter-constraint measures. In this situation, we consider that we will obtain the best performance when F-measure achieves the highest value. It is because that F-measure takes both precision rate and recall rate into consideration. Moreover, we observe that there exists an opposite tendency between precision rate and recall rate, they are getting closer when F-measure increases. From the experi-

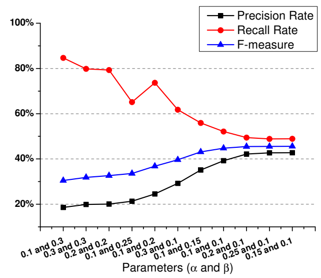


Figure 7: The PRF curve under different parameters.

ment above, we set $\alpha = 0.15$, $\beta = 0.1$ after this study, which achieves the best performance of F-measure. Under this situation, our approach could achieves 42.74% in precision and 48.91% in recall, which reflects the performance considering all anonymity levels of privacy protection.

5.3 Performance Comparisons

The experiment aims to validate that our algorithm could discover social ties at some extent and make a comparison with the previous algorithm. We evaluate our performance from two aspects as follows: 1)To validate performances in terms of different anonymity levels, all of the users are grouped by their values of k. Then we compared the performance of different groups; 2)We compare the performance of our approach with the KSTCM model [11] on our dataset.

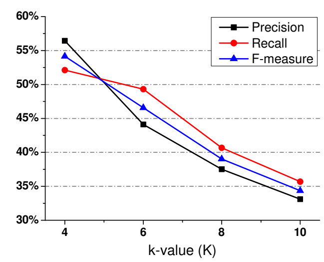
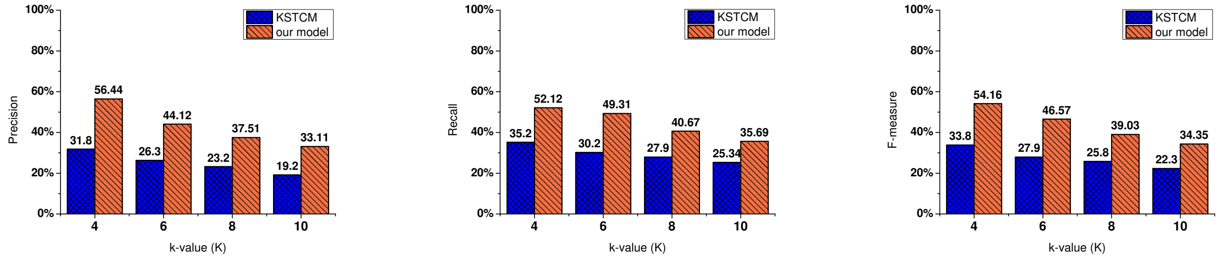


Figure 8: The influence of k on performance.

In our approach, when k increases, the cloaking region should be expanded to ensure at least k-1 anonymous users are in the region. In Fig.8, we observe that the precision rate and recall rate deteriorate as k increases, to say, we could get better performance for users with lower privacy requirement. This observation indicates that the size of a cloaking region has a great impact on the accuracy of social tie detection.

The comparison of our mode and the KSTCM model [11] could be shown in Fig.9. We observe that our approach gets a better performance than KSTCM model and achieves almost 15% higher accuracy on all three metrics. It demonstrates that our approach using semantic regions is more capable than using raw cloaking regions. Besides, taking ac-



(a) Precision Rate comparison (b) Recall Rate comparison (c) F-measure comparison

Figure 9: The comparison of precision rate, recall rate and F-measure under different values of k.

count of the hierarchical containment relationship between semantic regions is probably another factor for the improved performance. The result implies that semantic regions could reveal more individuals' interests and preferences, and individuals usually have closer social ties when sharing similar semantic regions.

6. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a novel approach to discover social ties using cloaked trajectories instead of accurate GPS trajectories. Rather than directly matching historical locations in the geographic space, we transform cloaking regions to semantic regions, and further proposed a hierarchical semantic tree to make the containment relationship visible. The hierarchical semantic tree can be used to calculate the similarity of the trajectories of individuals and predict the existence of social ties between them. The evaluation results demonstrated that our approach could infer social ties and preserve privacy of users successfully. We have compared our algorithm with existing work using k-anonymity cloaking method and demonstrated that our approach can achieve much higher accuracy in social ties detection.

In the future, we would like to make a improvement in the privacy preserving algorithm by changing the shape of cloaking regions.

7. REFERENCES

[1] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. Friendship and mobility: User movement in location-based social networks. In *Proceedings of KDD'11*, pages 1082–1090. ACM, 2011.

[2] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: User movement in location-based social networks. In *Proceedings of KDD'11*, pages 1082–1090, 2011.

[3] Bugra Gedik and Ling Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 7(1):1–18, 2008.

[4] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services*, pages 31–42. ACM, 2003.

[5] Hsun-Ping Hsieh and Cheng-Te Li. Inferring social relationships from mobile sensor data. In *Proceedings of WWW Companion'14*, pages 293–294, 2014.

[6] Jie Gao Khuong Vu, Rong Zheng. Efficient algorithms for k-anonymous location privacy in participatory sensing. In *Proceedings of INFOCOM'12*, pages 2399 – 2407. IEEE, 2012.

[7] Ben Niu, Qinghua Li, and Xiaoyan Zhu. Achieving k-anonymity in privacy-aware location-based services. In *Proceedings of INFOCOM'14*, 2014.

[8] Srinivasan Parthasarathy, Yiye Ruan, and Venu Satuluri. Community discovery in social networks: Applications, methods and emerging trends. In *Social Network Data Analytics*, pages 79–113. Springer, 2011.

[9] Krishna PN Puttaswamy, Shiyuan Wang, and Troy Steinbauer. Preserving location privacy in geosocial applications. *IEEE Transactions on Mobile Computing*, 13(1):159–173, 2014.

[10] Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. Exploiting place features in link prediction on location-based social networks. In *Proceedings of SIGKDD'11*, pages 1046–1054. ACM, 2011.

[11] Rong Tan, Junzhong Gu, Peng Chen, and Zhou Zhong. Link prediction using protected location history. In *Proceedings of ICCIS 2013*, pages 795–798. IEEE, 2013.

[12] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of SIGKDD'11*, pages 1100–1108. ACM, 2011.

[13] Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing*, 5(1):3–19, 2014.

[14] Josh Jia-Ching Ying, Wang-Chien Lee, and Tz-Chiao Weng. Semantic trajectory mining for location prediction. In *Proceedings of SIGSPATIAL'11*, pages 34–43. ACM, 2011.

[15] Jun Zhang, Chaokun Wang, and Jianmin Wang. Who proposed the relationship?: Recovering the hidden directions of undirected social networks. In *Proceedings of WWW'14*, pages 807–818, 2014.