Detecting Location-centric Communities using Social-Spatial Links with Temporal Constraints

Kwan Hui Lim[1,2], Jeffrey Chan[1], Christopher Leckie[1,2], and Shanika Karunasekera[1]

[1] Department of Computing and Information Systems, The University of Melbourne, Parkville, VIC 3010, Australia
[2] Victoria Research Laboratory, National ICT Australia
{limk2@student., jeffrey.chan@, caleckie@, karus@}unimelb.edu.au

Abstract. Community detection on social networks typically aims to cluster users into different communities based on their social links. The increasing popularity of Location-based Social Networks offers the opportunity to augment these social links with spatial information, for detecting location-centric communities that frequently visit similar places. Such location-centric communities are important to companies for their location-based and mobile advertising efforts. We propose an approach to detect location-centric communities by augmenting social links with both spatial and temporal information, and demonstrate its effectiveness using two Foursquare datasets. In addition, we study the effects of social, spatial and temporal information on communities and observe the following: (i) augmenting social links with spatial and temporal information results in location-centric communities with high levels of check-in and locality similarity; (ii) using spatial and temporal information without social links however leads to communities that are less location-centric.

Keywords: Community Detection, Clustering Algorithms, Foursquare, Location-based Social Networks, Social Networks

1 Introduction

The study of communities on social networks typically involves using community detection algorithms to cluster users into different communities based on their friendships on the social network (i.e., social links). With the rising popularity of Location-based Social Networks (LBSN), it is now possible to add a spatial aspect to these traditional social links for the purpose of community detection. Many researchers have used such social-spatial links to detect location-centric communities on LBSNs [2, 3]. The detection of these location-centric communities is especially important for companies embarking on location-based and mobile advertising, which are increasingly crucial to any company's marketing efforts [5]. We posit that the detection of such location-centric communities can be further improved by adding a temporal constraint to such social-spatial links, and demonstrate the effectiveness of this approach using two LBSN datasets.

Table 1. Types of Links

Link Type	Description
Social (SOC)	Links based on explicitly declared *friendships* (i.e., topological links)
Social-Spatial-Temporal (SST)	*Social links* where two users share a *common check-in*, on the *same day*
Social-Spatial (SS)	*Social links* where two users share a *common check-in*, regardless of time
Spatial-Temporal (ST)	Links based on two users sharing a *common check-in*, on the *same day*

In addition, we study the effects of social, spatial and temporal links on the resulting communities, in terms of various location-based measures.

Related Work. The spatial aspects of LBSNs have been used in applications ranging from friendship prediction to detecting location-centric communities. For example, [4] used spatial-temporal links (photos taken at the same place and time) to infer friendships on Flickr, while [13] used spatial links (tweets sent from the same location) and tweet content similarity to predict friendships on Twitter. Similarly, [2] used social-spatial links (friends with common check-ins) to detect location-centric communities on Twitter and Gowalla. Brown et al. [3] also used social-spatial links to study the topological and spatial characteristics of city-based social networks, and [9] found that communities with common interest tend to comprise users who are geographically located in the same city.

Most of these earlier works consider the spatial aspect of check-ins and co-location without the temporal aspect (e.g., visiting the same place over any span of time), while [4] considers this temporal aspect for the purpose of friendship prediction. Our research extends these earlier works by adding a temporal constraint to social and spatial links, for the purpose of detecting location-centric communities. Using two LBSN datasets, we demonstrate the effectiveness of our proposed approach in detecting location-centric communities that display high levels of check-in and locality similarity.

Contributions. We make a two-fold contribution in this paper by: (i) enhancing existing community detection algorithms by augmenting traditional social links with both a spatial aspect and temporal constraint; (ii) demonstrating how these links result in location-centric communities comprising users that are more similar in terms of both their visited locations and residential hometown.

2 Methodology

Our proposed approach to detecting location-centric communities involves first building a social network graph $G = (N, E_t)$, where N refers to the set of users and E_t refers to the set of links of type t (as defined in Table 1). SOC links are essentially topological links that are used in traditional community detection tasks, while SS links were used in [2] to detect location-focused communities with great success. Our work extends [2] by adding a temporal constraint to these links, resulting in our SST links.[3] Furthermore, we also use ST links to determine

[3] While SST links can also be defined as two friends who share a common check-in within D days, our experiments show that a value of D=1 offers the best results,

the effects of adding this temporal constraint solely to the spatial aspects of links (i.e., without considering social information). While there are many definitions of links, these four types of links allow us to best investigate the effects of social, spatial and temporal information on location-centric communities.

Then, we apply a standard community detection algorithm on graph G, resulting in a set of communities. Thus, the different types of links (SOC, SST, SS and ST) used to construct the graph G will result in the different types of communities that we evaluate in this paper. We denote the detected communities as Com_{SOC}, Com_{SST}, Com_{SS} and Com_{ST}, corresponding to the types of links used. In this experiment, Com_{SST} are the communities detected by our proposed approach, while Com_{SOC}, Com_{SS} and Com_{ST} serve as baselines.

For the choice of community detection algorithms, we choose the Louvain [1], Infomap [12] and LabelProp [11] algorithms. Louvain is a greedy approach that aims to iteratively optimize modularity and results in a hierarchical community structure, while Infomap is a compression-based approach that uses random walkers to identify the key structures (i.e., communities) in the network. LabelProp first assigns labels to individual nodes and iteratively re-assigns these labels according to the most frequent label of neighbouring nodes, until reaching a consensus where the propagated labels denote the different communities. In principle, any other community detection algorithms can be utilized but we chose these community detection algorithms for their superior performance [6], and also to show that our obtained results are independent of any particular community detection algorithm.

3 Experiments and Results

Datasets. Our experiments were conducted on two Foursquare datasets, which are publicly available at [8] and [7]. Foursquare dataset 1 comprises 2.29M check-ins and 47k friendship links among 11k users, while dataset 2 comprises 2.07M check-ins and 115k friendship links among 18k users. Each check-in is tagged with a timestamp and latitude/longitude coordinates, which is associated with a specific location. In addition, dataset 1 provides the hometown locations that are explicitly provided by the users. We split these datasets into training and validation sets, using 70% and 30% of the check-in data respectively. The training set is used to construct the set of SST, SS and ST links, which will subsequently be used for community detection as described in Section 2.

Evaluation Metrics. Using the validation set, we evaluate the check-in activities and locality similarity of users within each Com_{SOC}, Com_{SST}, Com_{SS} and Com_{ST} community. Specifically, we use the following evaluation metrics:

1. **Average check-ins**: The mean number of check-ins to all locations, performed by all users in a community.

---

hence the current definition of SST links. More importantly, using higher values of D days converges SST links towards SS links, which we also investigate in this work.

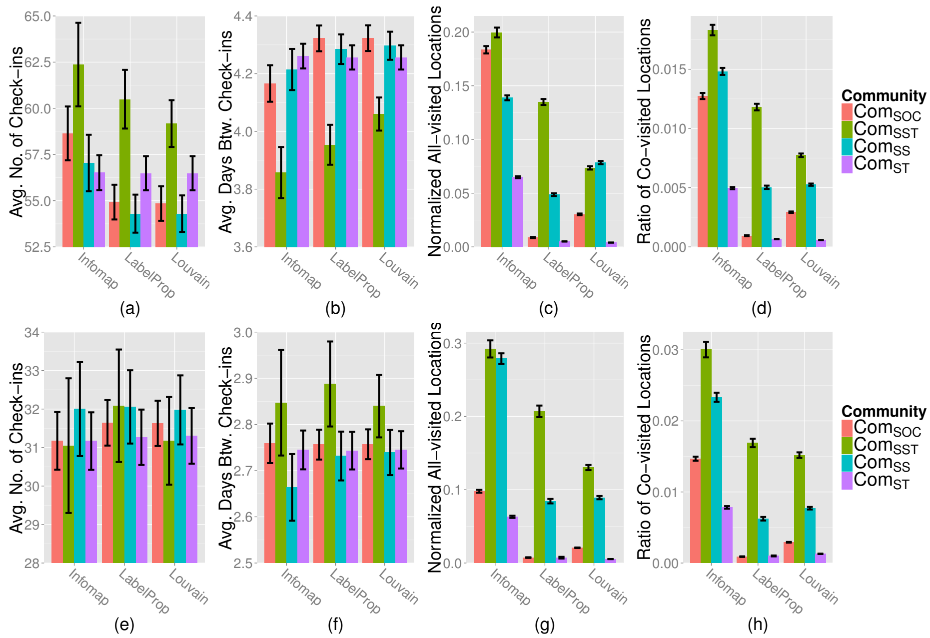

Fig. 1. Average number of check-ins, average days between check-ins, normalized number of all-visited locations and ratio of co-visited locations for Foursquare dataset 1 (top row) and dataset 2 (bottom row). For better readability, the y-axis for Fig. 1a/b/e/f do not start from zero. Error bars indicate one standard deviation. Best viewed in colour.

2. **Average unique check-ins**: The mean number of check-ins to unique locations, performed by all users in a community.
3. **Average days between check-ins**: The mean number of days between consecutive check-ins, performed by all users in a community.
4. **Normalized all-visited locations**: The number of times when all users of a community visited a unique location, normalized by the community size.
5. **Ratio of co-visited locations**: Defined as $\frac{1}{|C|}\sum_{i \in C} \frac{|L_i \cap L_C|}{|L_C|}$, where L_i is the set of unique locations visited by user i, and L_C is the set of unique locations visited by all users in a community C.
6. **Ratio of common hometown**: The largest proportion of users within a community that share the same hometown location.

Evaluation metrics 1 to 3 measure the level of user check-in activity, while metrics 4 to 6 measure the user locality (check-in and hometown) similarity within each community. Ideally, we want to detect communities with high levels of check-in activity and locality similarity. As Metrics 1 to 3 are self-explanatory, we elaborate on Metrics 4 to 6. Metric 4 (normalized all-visits) determines how location-centric the entire community is based on how often the entire community visits the same locations. We normalize this metric by the number of users in a community to remove the effect of community sizes (i.e., it is more likely for a community of 50 users to visit the same location than for a community of 500 users). Metric 5 (co-visit ratio) measures the similarity of users in a community (in terms of check-in locations) and a value of 1 indicates that all users visit the exact set of locations, while a value closer to 0 indicates otherwise. Similarly, a value of 1 for Metric 6 (hometown ratio) indicates that all users in a community reside in the same location, while a value of 0 indicates otherwise.

Results. We focus on communities with >30 users as larger communities are more useful for a company's location-based and mobile advertising efforts. Furthermore, there has been various research that investigated the geographic properties of communities with ≤30 users [2, 10]. In particular, [10] found that communities with >30 users tend to be more geographically distributed than smaller communities. Instead of repeating these early studies, we investigate the check-in activities and locality similarity of communities with >30 users.

In terms of the average number of check-ins (Fig. 1a/e), unique check-ins (not shown due to space constraints) and days between check-ins (Fig. 1b/f), Com_{SST} outperforms Com_{SOC}, Com_{SS} and Com_{ST} on dataset 1, regardless of which community detection algorithm used. However for dataset 2, the performance of Com_{SST} is largely indistinguishable from that of Com_{SOC}, Com_{SS} and Com_{ST}.[4] For both datasets, there is no clear difference among Com_{SOC}, Com_{SS} and Com_{ST} in terms of the average number of check-ins, unique check-ins or days between check-ins. These results show that our proposed SST links can be used to effectively detect communities that are more active in terms of check-in activity (for dataset 1), and such communities serve as a good target audience for a company's location-based and mobile advertising efforts. There is no clear difference among using SOC, SS and ST links (for both datasets). For the detection of location-centric communities, the locality similarity of these communities is a more important consideration, which we investigate next.

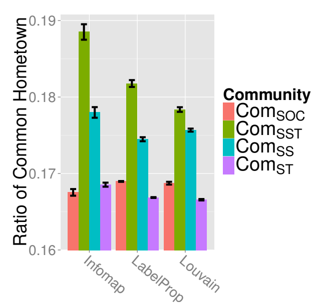

Fig. 2. Common hometown ratio for dataset 1.

We examine locality similarity of the four communities in terms of the normalized number of all-visited locations (Fig. 1c/g), ratio of co-visited locations (Fig. 1d/h) and ratio of common hometown (Fig. 2). We only compare the ratio of common hometown for dataset 1 as this information is not available for dataset 2. For both datasets, Com_{SST} offers the best overall performance in terms of these three locality similarity metrics, while Com_{SS} offers the second best overall performance.[5] On the other hand, Com_{ST} resulted in the worst performance for both datasets. These results show that using our proposed SST links results in

[4] With an exception in Fig. 1f where Com_{SST} marginally underperforms Com_{SOC}, Com_{SS} and Com_{ST}.

[5] With exceptions in Fig. 1c where Com_{SS} (using Louvain) outperforms Com_{SST}, and Com_{SOC} (using Infomap) outperforms Com_{SS}.

communities comprising users who tend to frequently visit similar locations and reside in the same geographic area. Such location-centric communities are useful for the purposes of providing meaningful location-relevant recommendations and to better understand LBSN user behavior.

4 Discussions and Conclusion

We demonstrate how standard community detection algorithms can be used to detect location-centric communities by augmenting traditional social links with spatial information and a temporal constraint. Our evaluations on two Foursquare LBSN datasets show that: (i) augmenting social links with spatial information allows us to detect location-centric communities (ii) however, using spatial/temporal information (without considering social links) results in communities that are less location-centric than communities based solely on social links, thus spatial/temporal information should not be used independently; and (iii) our proposed approach of augmenting social links with both spatial and temporal information offers the best performance and results in location-centric communities, which display high levels of check-in and locality similarity.

Acknowledgments. This work was supported by National ICT Australia (NICTA).

References

1. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. of Statistical Mechanics 2008(10), P10008 (2008)
2. Brown, C., Nicosia, V., et. al.: The importance of being placefriends: discovering location-focused online communities. In: Proc. of WOSN. pp. 31–36 (2012)
3. Brown, C., Noulas, A., Mascolo, C., Blondel, V.: A place-focused model for social networks in cities. In: Proc. of SocialCom. pp. 75–80 (2013)
4. Crandall, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., Kleinberg, J.: Inferring social ties from geographic coincidences. PNAS 107(52) (2010)
5. Dhar, S., Varshney, U.: Challenges and business models for mobile location-based services and advertising. Communications of the ACM 54(5), 121–128 (2011)
6. Fortunato, S.: Community detection in graphs. Physics Reports 486(3) (2010)
7. Gao, H., Tang, J., Liu, H.: Exploring social-historical ties on location-based social networks. In: Proc. of ICWSM. pp. 114–121 (2012)
8. Gao, H., Tang, J., Liu, H.: gSCorr: modeling geo-social correlations for new check-ins on location-based social networks. In: Proc. of CIKM. pp. 1582–1586 (2012)
9. Lim, K.H., Datta, A.: Tweets beget propinquity: Detecting highly interactive communities on twitter using tweeting links. In: Proc. of WI-IAT. pp. 214–221 (2012)
10. Onnela, J.P., Arbesman, S., González, M.C., Barabási, A.L., Christakis, N.A.: Geographic constraints on social network groups. PLoS one 6(4), e16939 (2011)
11. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Phy. Review E 76(3), 036106 (2007)
12. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. PNAS 105(4), 1118–1123 (2008)
13. Sadilek, A., Kautz, H., Bigham, J.P.: Finding your friends and following them to where you are. In: Proc. of WSDM. pp. 723–732 (2012)