SRC: Discovering Human Activity Community in A City

Tianyuan Huang
South China University of Technology
Guangzhou, China
tianyuanhuang@outlook.com

ABSTRACT

This study investigates human activity community in a city by conceptualizing it as a network embedding problem. In order to learn the latent representations of activity-travel patterns from individual daily trajectories, network embedding learns a vector space representation for each type of activity place as a node connected by movement links to preserve the structure of individual activities. The proposed approach is applied to mobile positioning data at the individual level obtained for a weekday from volunteers at Guangzhou City. Assessments are conducted to validate individual decision making for several types of activities by a field survey. This study contributes to a general framework for discovering individual activity-travel patterns from human movement trajectories.

CCS CONCEPTS

• Information systems → Geographic information systems

KEYWORDS

Network embedding, Graph representation, Geographic knowledge discovery, Urban planning

1 INTRODUCTION

Discovering communities of places by human activities can enable a variety of valuable applications. First, it can provide a quick understanding of a complex city. Trajectory data can identify people's activity-travel patterns and have a better interpretation of complex activities in urban areas. Second, given that one user's trajectory data reflects that user's behavioral preferences, the massive amount of trajectory data of a city can reflect the general behavioral characteristics of its urban residents and the general demand for urban space. In this way, extracting behavioral characteristics from user trajectory data can effectively evaluate the rationality of urban functional layout, in order to assess urban planning strategies.

There has been recent research [3] that used human trajectory and remote-sensor data to discover regions of different functions in a city. There is also research [1] that characterize the relationship among the POIs based on how people access them. In this paper, we introduce a network embedding (Node2Vec) [2] technique, which has been proven successful on multi-label classification in several real-world networks from diverse domains, into individual trajectories analysis. By constructing a network of places for human activities as nodes connected by movement links, activity-travel patterns can be extracted as latent features of the nodes. The latent features will capture the neighborhood similarity between different places and encode people's daily interactions with different places in a continuous vector space with a relatively small number of dimensions.

2 METHODOLOGY

The input data are individual movement trajectories represented as sequences of locations that are attached to Point of Interests (POIs). We seek to construct a network from the trajectory data and then apply the Node2Vec Algorithm to learn an embedding representation for activity places, based on which we analyze communities of human activity places.

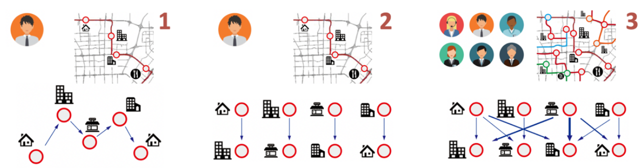

Figure 1: Overview of Network Construction Process

2.1 Construction of Activity Networks

A network is formally defined as $G = \langle V, E \rangle$ in which entities (the nodes in V) are linked by ties (the edges in E), representing any sort of connection, similarity or interaction [1]. We propose to construct an activity network with a set of nodes which correspond to the set of activity places, represented by POIs, where users stay to perform some activities. The overall process is shown in Figure 1. Given a universal set V of n places x_k in a city, one visits a subset of m places within V and it forms one's trajectory that can be represented as a series of place, for example,

$$traj_i = [x_1, x_2, \dots, x_k, \dots, x_m], x_k \in V \tag{1}$$

which is demonstrated in Step 1 of Figure 1. In step 2, a user 's trajectory can be decomposed into a set of Origin-Destination links; each link connects two consecutive places in the series. We denote this set as one's activity set (2).

$$act_i = \{[traj_i[j], traj_i[j+1]] \mid j \in [0, m-1]; j \in \mathbb{Z}\} \tag{2}$$

As (2) is a multiset, it can be written as:

$$act_i = \{[x_k, x_l]^{f([x_k, x_l])} \mid [x_k, x_l] \in A(act_i); x_k, x_l \in traj_i; x_k, x_l \in V\}$$

where $A(\cdot)$ gives the underlying set of act_i, which only contains unique links. $f(\cdot)$ is a function mapping a link to its number of occurrences in the multiset. In Step 3, we construct an activity network G by aggregating all individuals' (the set I) activity sets to form a directed weighted graph.

$$G = \langle V, E \rangle, E = \textstyle\sum_{i \in I}(act_i) \tag{3}$$

where $\sum(\cdot)$ is the sum operation of multiset for all act_i. In other words, E contains all the unique links from a combination of all individuals' activity sets; the weight for a unique link is obtained from the total occurrences of this link in the combination.

2.2 Network Embedding

We apply Node2Vec [2] to the networks of activity places and learn a mapping of places to a low-dimensional space of features that maximizes the likelihood of preserving the network structure of nodes. We then discover the communities of human activity places based on the learned embedding presentations.

2.3 Community Discovery

In this study, communities are groups of highly interactive and densely connected places that are frequently visited by individual trajectories. These places will be mapped as points close to each other in the embedding space. In order to determine the similarity between places, we calculate cosine similarity between the embeddings. Formally, *similarity* between embedding A and B could be defined as:

$$similarity = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{4}$$

For a given place, a rank of other places can be found based on the calculated cosine similarity (4). A community can be delineated by choosing the top-k ranked places surrounding the target place. Alternatively, community detection can be done with clustering techniques [1].

3 EXPERIMENTAL RESULTS

The set of collected observations in a weekday in Guangzhou is composed of 48,178 points for 2,575 volunteers in the Guangzhou city. We compute the network of 1,852 nodes and 21,742 edges and learn continuous feature representations for the nodes. Figure 2 shows the visualization of our network.

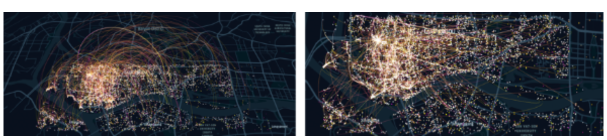

Figure 2: Visualization of the Trajectory Network

Ranking of Similarity	Xinfeng Residential District	Hualin International Plaza	Shisanhang Apparel Market	Qingping Medicine Market
1	Liwanhu Park	Hualin Temple	Hengbao Shopping Center	Sun-Yat-San Memorial Hospital
2	Xinglong District	7 Days Inn	Xinya Commercial Center	Guangzhou Hospital of Chinese Medicine
3	Duobao Market	Minghui Plaza	Yutian Center	Xinfeng District

Table 1: Selected Training Results

Each point is associated with a POI where user stays to perform some activities. In order to discover places having high interactions, we choose four target POIs with a relatively high frequency of visits in the dataset and find the top 3 most similar POIs of each target POI based on the embeddings and the cosine similarity measure. Table 1 shows the above selected results. To validate our network embedding approach, we complete a field investigation. Specifically, we interviewed 50 people in those four different places and requested them to label two common destinations in their daily activities. We then aggregated the results and found the top 3 most visited POIs for each place. It shows that over 50 percent of their choices are assigned into the same POIs with high similarity in Table 1.

4 DISCUSSION AND FUTURE WORKS

We proposed an exploratory study on human activity community based on POI data and a network embedding approach. The novelty of this work is two-fold. First, we presented an algorithm to build a complex network that synthesizes trajectories of places that people visit to conduct daily activities. Second, we employ a network embedding technique to learn a representation for activity places as nodes connected by movement links to preserve the structural human activities. Some potential future works include discovering socioeconomic groups from individual trajectories and simulation of human travel patterns to assist urban-transport planning.

ACKNOWLEDGMENTS

The author would like to thank Dr. Zhaoya Gong from the University of Birmingham and Dr. Qiwei Ma from Tsinghua University, for their extensive guidance.

REFERENCES

[1] I. R. Brilhante, M. Berlingerio, R. Trasarti, C. Renso, J. A. F. d. Macedo and M. A. Casanova, "ComeTogether: Discovering Communities of Places in Mobility Data," 2012 IEEE 13th International Conference on Mobile Data Management, Bengaluru, Karnataka, 2012, pp. 268-273.

[2] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY, USA, 855-864.

[3] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). ACM, 186-194.