Physica A 461 (2016) 694–707

Contents lists available at ScienceDirect

Physica A

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

Statistical properties of urban mobility from location-based travel networks

Jinjun Tang [a,b], Shen Zhang [a], Wenhui Zhang [c], Fang Liu [d], Weibin Zhang [b,*], Yinhai Wang [b]

[a] School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150001, China
[b] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195-2700, USA
[c] Traffic School, Northeast Forestry University, Harbin, 150040, China
[d] School of Energy and Transportation Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China

HIGHLIGHTS

- Travel trips are extracted from raw data and operation efficiency is analyzed.
- Travel networks based on occupied and vacant trips are constructed.
- Some statistical properties of traveling network are used to explore mobility.
- Two traditional community detection algorithms are used to recognize traffic zones.
- Spatial structure and hotspots are analyzed from local density of locations.

ARTICLE INFO

Article history:
Received 2 December 2015
Received in revised form 14 April 2016
Available online 11 June 2016

Keywords:
Human mobility
Taxi GPS data
Network properties
Community detection
Hotspots estimation

ABSTRACT

This study explores urban mobility from a network-based perspective. The data samples used in study were collected from more than 1100 taxi drivers during a half year period in the city of Harbin in China. We extract trips from the original dataset and analyze operational efficiency. Then, by constructing travel networks based on occupied and vacant taxi trips, we calculate some statistical properties of the network such as degree, strength, edge weight, betweenness, clustering coefficient and network structure entropy. Analysis of such properties allows for a deep exploration of travel mobility. We also analyze the correlation between strength and betweenness to evaluate the importance of nodes in the network. Furthermore, two traditional community detection algorithms: the Louvain method and the visualization of similarities (VOS) method are applied to divide traffic zones in the mainland area of Harbin city. Two indices, the Rand index (RI) and the Fowkles–Mallows index (FMI) are adopted to evaluate recognition performance, which shows the similarity between administrative division and results from the algorithms. Finally, a dilatation index based on the weighted average distance among trips is applied to analyze the spatial structure of an urban city. Furthermore, hotspots are identified from local density of locations with different thresholds as determined by the Lorenz curve.

* Corresponding author.
E-mail address: wbin.zhang@outlook.com (W. Zhang).

http://dx.doi.org/10.1016/j.physa.2016.06.031

1. Introduction

Traffic problems in urban systems such as traffic congestion, accidents and air pollution have been attracting a large amount of consideration from citizens. The movement of citizens (by different modes: private car, public bus, taxi, bike, etc.) generates traffic flow on the road networks which if not managed properly can lead to the aforementioned problems. Thus, how to deeply understand human mobility in an urban city, which can be affected by a number of factors such as the spatial structure of city, land use and layout road networks, is a key step in the establishment of reasonable transportation planning and management strategies. Recently, researchers from a variety of different fields such as sociology [1], ecology [2], disease spread [3], urban planning [4] and traffic systems [5] have begun focusing on human mobility. However, many earlier studies used data collected from conventional methods which may lead to analysis issues caused by small sample sizes and low spatial coverage rate. For example, as a questionnaire based approach is constrained by the amount of data samples, it is difficult to use this traditional method to explore human mobility in depth and accurately. The fast development of information and communication technology makes it possible to understand travel behavior of people by providing access to large-scale and granular data obtained from recording individual information chronologically. Various datasets including wireless network traces [6], global positioning system (GPS) traces from probe vehicle data [7–10], mobile phone data [11–20], social media data [21] and banking notes [22] have been collected in earlier studies to investigate the spatiotemporal features of human movement.

As a main part of public transportation systems in cities, taxis account for a massive proportion of citizens' travel due to their accessibility and flexibility. Furthermore, when one considers GPS-equipped taxis as probe vehicles, these mobile sensors provide us with new tools to discover spatiotemporal patterns of people's movements and even origin and destination distributions. Thus, compared to data sources from cell phones, taxi location data can reflect travel characteristics more precisely as passengers who use taxis have clearly defined origins and destinations. Recently, several studies have focused on human activity recognition, hotspot discovery, urban planning and transportation planning [23–31]. Liu et al. [23] introduced a new method to explore intra-urban human mobility and land use variation based on taxi trajectory data from the city of Shanghai. Chang et al. [24] proposed a model to predict context-aware demand distributions for taxi management systems; the model includes four steps: data filtering, clustering, semantic annotation, and hotness calculation. Qi et al. [25] established the relationship between the loading/unloading characteristics of taxi passengers and the social function of urban regions. They conclude that the amount of customer loading/unloading in a region is able to reflect the social activity dynamics. Chang et al. [26] used a spatial statistical analysis, data mining and a clustering algorithm on historical data to predict potential hotspots of taxi requests. Castro et al. [27] proposed an overview of mechanisms for using taxi GPS data to analyze peoples' movements and activities, which includes three main categories: social dynamics, traffic dynamics and operational dynamics. Liang et al. [28] found that taxis' traveling displacements and trip durations follow an exponential distribution. Liu et al. [29] used a two-level hierarchical polycentric city structure to study the spatial interaction perspective in Shanghai city with large scale taxi data. Pan et al. [30] discussed a new method to classify urban land-use features by using taxi traces. Wang et al. [31] studied human mobility from trips displacement, duration and time interval based on taxi trace data from five cities in two countries.

Converting the main urban area in a city into a grid or series of cells and assuming the travel of entities can be represented as movements between grid cells has been regarded as an effective approach to analyze human mobility based on large-scale, location-based data [21,32–34]. As urban human mobility reflects overall travel characteristics based on statistical results from large scale distribution data, this simplified method can also obtain satisfactory results when analyzing urban mobility in a network. A trip extracted from taxi GPS trace can be stored as a transition matrix based on grids, and large amounts of trips can then be combined to construct a location-based network, namely a travel network. The purpose of this study is to explore human mobility from the network perspective. In this study, we use taxi GPS data collected from more than 1100 drivers during a six month period in Harbin city to characterize travel mobility. Travel efficiency is analyzed for occupied and vacant trips, and results a statistical analysis of travel distance and time are then used to uncover patterns in human mobility. In order to construct the travel network, we consider grid cells as nodes and their connections as edges. The weights of edges are determined by the number of trips between grid cells or nodes. Thus, human mobility is explored by using statistical properties of the network: degree, weights, strength, betweenness, clustering coefficients, and the network structure entropy. Finally, we analyze spatial structure and identify hotspots in urban city.

The rest of this paper is organized as follows. Section 2 introduces the dataset used in the paper, travel efficiency and the distribution of trips with respect to distance and time. Results of network-based methods are discussed in Section 3. Section 4 describes spatiotemporal properties of urban travel. The conclusion is provided in the final section.

2. Data collection and analysis

2.1. Data source

The taxi GPS data we used in this study were collected from approximately 1100 drivers in Harbin city, which is located in the northeastern part of China. The data collection period took place from July to December in 2012 and the recording rate was 30 s; hence, there is a total of 2880 samples per day. Each data point in the sample contains not only location information