Title: UNVEILING URBAN COMMUTING STRUCTURE FROM MOBILE PHONE DATA: A CASE STUDY IN SHANGHAI, CHINA

TRB Paper number: 18-00784

Qing Yu
Master Student
Key Laboratory of Road and Traffic Engineering of the Ministry of Education, College of Transportation Engineering, Tongji University
4800 Cao’an Road, Shanghai 201804, China
Tel: +86-21-69583775; E-mail: 714727644@qq.com

Weifeng Li
PhD Student
Key Laboratory of Road and Traffic Engineering of the Ministry of Education, College of Transportation Engineering, Tongji University
4800 Cao’an Road, Shanghai 201804, China
Tel: +86-21-69583775; E-mail: lwf0103@gmail.com

Zhengyu Duan
Assistant Professor, PhD
Key Laboratory of Road and Traffic Engineering of the Ministry of Education, College of Transportation Engineering, Tongji University
4800 Cao’an Road, Shanghai, 201804, China
Tel: +86 13801945109; Email: d_zy@163.com

Dongyuan Yang (Corresponding Author)
Professor, PhD
Key Laboratory of Road and Traffic Engineering of the Ministry of Education, College of Transportation Engineering, Tongji University
4800 Cao’an Road, Shanghai 201804, China
Tel: +86-21-69583775; E-mail: yangdyk@tongji.edu.cn

Keywords: Commuting behavior; Urban commuting structure; Mobile phone data; Transportation Analysis Zone (TAZ); Community detection; Complex network

Acknowledgements: This work is supported by the National Natural Science Foundation of China (Grant No. 51478350).
INTRODUCTION
Traditionally, our knowledge of the city structure mostly comes from urban designs and plans, which only depict the city structure from morphological aspect. However, in recent discussion of urban change, cities are no longer seen as mere morphological entities with clear and detectable borders. In the concept of ‘space of flows’ (1), the urban form of most urban regions are constructed by functional network of communities which may be physically separate but connected through dense flows of commuting trips and other forms of daily mobility (2).

Dwelling and employment are the two fundamental elements of a city. The structure of a city is closely related to the commuting behaviors of residents. The allocation of job centers and residential communities generates commuting demand and travel flows. The travel flows connect discrete places into an integrated system, forming the commuting structure of a city. Therefore, studying the residential commuting behavior can help us understand their commute demand, unveil the city commuting structure, and even further, to support the urban transportation planning.

The emerging of pervasive, geospatial data generated by individuals have recently triggered an opportunity for studying individual mobility patterns (3), city dynamics (4) and city structure (5). As a new travel survey tool, mobile phone data is more pervasive and accurate, which allow us to fully track the trip chain of individual in both temporal and spatial dimensions. It offers a new way to study urban structure from the functional aspect.

The purpose of this study is to explore urban structure from the job-housing function of the city. We introduce community detection method to analyze job-housing relationships extracted from mobile phone data of Shanghai, China. Our study unveils a two-level hierarchical commuting structure of Shanghai and we further exploring the reason for forming this commuting structure.

DATA
Anonymous mobile phone data used in this paper was collected for billing and operational purposes from September 1 to September 15, 2011 in Shanghai, China. It includes the information of encrypted mobile phone identifier, service time, service type, location of base transceiver station (BTS) and location area (LA), etc. A record of mobile phone data will be generated when a call is placed or received, a text message is sent or received, the phone is switched on or switched off, or the phone signal is handed over from one BTS to the other BTS, etc. The average number of records was 1 billion per day, covering 25 million active users. The coverage radius of BTS is 500 to 800 meters.

METHODOLOGY
Data preprocessing
Once a user can be captured by more than one BTSs at the same time, its signal will be hand over frequently between these BTSs and generate a great number of records in a very short time even if there is no position changing of the user. The frequent handover leads to data noise and waste of system resources. Therefore, a binning method (6) was used to cope with this problem and reduce the volume of data.
Identifying residents and their homes
Due to the large data volume, we choose a simple method proposed by Li et al. (6) to identify the home locations of residents from mobile phone data. The method includes two rules: (a) In the period from 9 p.m. to 9 a.m. the next day, the user stays at one place for no less than six hours. (b) In our observation periods, the user stays at the place satisfied rule (a) in more than 2/3 of the days (10 days).
If a user satisfied both two rules, the user can be considered as a resident and the place will be considered as the location of home.

Identifying commuters
A threshold $\delta$ is defined to identify commuters. If a resident stay more than $\delta$ of time outside home in weekday, he can be regarded as a commuter, otherwise, he is a non-commuter. Here we choose the mean of stay time outside home of all residents as the threshold $\delta$, which will split residents into two equal parts — commuters and non-commuters.

Extracting job anchor points
Anchor points are the important control points in individuals’ daily life (7). In this study, we mainly consider home and job anchor points. The location where an individual stays over 30 minutes is defined as an activity point. And the activity points at work time (9a.m. – 18p.m. on weekdays) are regard as job anchor points.

Construction of complex network
For the residents in TAZ $i$, the number of job anchor points they have per hour in TAZ $j$ can be define as the number of connections from TAZ $i$ to TAZ $j$, which is denoted as $Q_{ij}$ here. By aggregating all job and home anchor points, we can obtain a matrix $V$:

$$
V = \begin{bmatrix}
Q_{11} & Q_{12} & \cdots & Q_{1j} \\
Q_{21} & Q_{22} & \cdots & Q_{2j} \\
\vdots & \vdots & \ddots & \vdots \\
Q_{i1} & Q_{i2} & \cdots & Q_{ij}
\end{bmatrix}
$$

To build a network, each TAZ is represented as a node. Between every node $i$ and $j$, we constructed two directed edge $E_{ij}$ and $E_{ji}$ with the weight $Q_{ij}$ and $Q_{ji}$.

Community Detection
Community detection can be implemented in many algorithms (8). Here, the fast unfolding algorithm is adopted to decompose our network (9). This algorithm is based on modularity optimization. The modularity of a partition is a scalar value between -1 and 1 that measures the density of links inside communities as compared to links between communities (10).
This algorithm includes the following two steps which are repeated iteratively until no increase of modularity is possible: (a) Modularity optimization: optimized modularity by allowing only local changes of communities; (b) Community aggregation: the identified communities are aggregated in order to build a new network of communities.
FINDINGS

By the identification methods, we identified 9.86 million residents, accounting for 42% of the total population of over 23.47 million in Shanghai by the end of 2011 (11). From the 9.86 million residents identified, we firstly eliminate the mobile phone users who never move during our observation period (1.13 million users in total). Then, for the remaining 8.73 million residents, the mean of stay time outside home on weekday for all 8.73 million residents is 7.93 hours. Choosing this value as threshold $\delta$ can divide the residents into two equal parts – commuters and non-commuters.

After extracting the job anchor points for commuters, we aggregate them into TAZs level and construct the complex network. The community detection algorithm iterates twice and find a two-level hierarchical structure. The results are shown in FIGURE 1(a) (level 1 structure) and (b) (level 2 structure). The hierarchical sub-regional structure provides insights into how the city could be properly divided into closely related sub-regions on the basis of job-housing relationship. Communities in the network represent regions with intense job-housing connection.

One of the interesting findings is that in both two structures, the boundaries of communities perfectly consist with administrative boundaries. In suburban district, each community is an administrative unit. But in central urban area, communities are often cross several administrative units. This finding indicates that residential commuting behavior is highly restricted by the administrative boundaries, especially in suburban area. The finding also proves the rationality of the city commuting structure we uncovered.

In order to depict the commuting pattern between communities, we calculate the number of imported and exported job anchor points for each community, from which we can easily classify communities into three types (see TABLE 1).

<table>
<thead>
<tr>
<th>Community type</th>
<th>Communities in level 1</th>
<th>Communities in level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Job center communities</td>
<td>1,8,10</td>
<td>1,4</td>
</tr>
<tr>
<td>2-Residential communities</td>
<td>2,3,4,9,15,17</td>
<td>2,3,8</td>
</tr>
<tr>
<td>3-Isolated communities</td>
<td>5,6,7,11,12,13,14,16</td>
<td>5,6,7,9,10,11,12</td>
</tr>
</tbody>
</table>

Concentric, sector and multiple nuclei structure are the three generalizations of urban structure (12; 13). From the result of classification, we can simplify the commuting structure of Shanghai into a combination of these three structures. On the city scale, we can see a multiple nuclei structure. Central urban area is the largest center and there are several centers of isolated communities in suburban area. The central urban area is a concentric structure, with community 1 in level 1 structure as job center and several similar residential communities on the periphery (community 2,3,4,15,17 in level 1 structure). In the level 2 structure, we can clearly see that the communities finally merging into a sector structure. The central urban district can be considered as a circle with four parts of areas (community 1-4 in the level 2 structure) as sectors radiating out from the center of the circle.

Further exploring the reason for forming the commuting structure, we compare the level 2 structure with the layout of metro network (see FIGURE 2). In the level 2 community structure,
communities in central urban area are all extend outward along with radiating metro lines, with averagely three metro lines in one community. Communities are also formed at the end of metro lines. From this structure, we can infer that the commuting behavior of residents living along the metro lines depend heavily on the metro line, and their work places aggregate along the metro line. For residents living at the end of metro lines, their work places are aggregated in sub-urban communities. This result demonstrates that, as the major traffic corridors, metro lines are playing important roles in forming city commuting structure.

![Community detection results](image)

**FIGURE 1** Results of community detection
CONCLUSIONS

This study proposes methodologies to identify commuters, extract job anchor points and unveil city commuting structure from mobile phone data. In the case study of Shanghai, we find a two-level hierarchical commuting structure.

By analyzing the hierarchical structure, it is found that:

1. Residential commuting behavior is highly restricted by the administrative boundaries, especially in suburban area;
2. The commuting structure in Shanghai can be simplify into a combination of concentric, sector and multiple nuclei structure;
3. Metro line network is playing an important role in forming city commuting structure.

REFERENCES


