Exploratory Analysis of Inter-City Human Activity Interactions for Shrinking Cities in Northeast China with flowAMOEBA

Zhaoya Gong[*1], Ran Tao[2], Qiwei Ma[3] and Changcheng Kan[4]

[1]School of Geography, Earth and Environmental Sciences, University of Birmingham
[2]Spatial Sciences Institute, University of Southern California
[3]School of Architecture, Tsinghua University
[4] Baidu Online Network Technology (Beijing) Co., Ltd

January 12, 2018

Summary

In this study, an exploratory flow clustering method, flowAMOEBA, is employed to study human activity based interactions between cities in Northeast China and in the rest of China on a short-term basis. This method allows us to discover new patterns of interaction from empirical trip data at the individual level. Our research is supported by big data of human movement extracted from mobile device positioning requests. Population effects are examined with selective trip clusters detected to demonstrate the potentials of our approach to unveil the true preference of trip making and to help form new hypotheses with discovered patterns.

KEYWORDS: inter-city interaction, human movement, flowAMOEBA, Northeast China, shrinking cities

1. Introduction

Northeast China (or Dongbei) provinces, including Liaoning, Jilin, and Heilongjiang, have topped the chart in the number of "shrinking cities", which experience population decline and lagged economic growth in the recent decades of Chinese urbanization. Urban shrinkage can be attributed to factors internal to cities including demographic change (e.g. ageing population) and suburbanization; it may also be attributed to economic decline as a mixture of internal and external aspects such as resource depletion, deindustrialization, or a lack of success in inter-city competition for investment (Haase et al, 2014). The latter generally leads to net out migration from the city region in search of jobs in the long run and the change of socioeconomic interactions between the shrinking city and other urban areas at a daily basis. In this sense, urban shrinkage can be conceptualized as an empirical phenomenon resulting from the interplay of changing factors of shrinkage at different spatial scales (from intra-city to inter-city or from regional to national) that produces a decline in population at the city level (Haase et al, 2016).

This study takes a perspective at the inter-city scale and investigate the pattern of socioeconomic interactions on a short-term regular basis (e.g. daily or weekly) between cities in Northeast China provinces and in the rest of China. We aim at gaining further insights into the mechanisms of shrinking cities through the analysis of human movement. Benefitted from the recent advancement of big data technologies, human movement data at the individual level become widely available to represent a considerable proportion of population due to the proliferation of mobile devices (e.g. smart phone) with positioning capability enabled. Inter-city human movement on a short-term basis, usually excluding migration or relocation movements, is a strong indication of socioeconomic interactions, such as

[*] z.gong@bham.ac.uk

commuting, business travels, and tourism, between places. This paper focuses on exploring the various patterns of human activity based interactions between the shrinking cities in Northeast China and the urban areas outside that region.

2. Methodology

We carried out the analysis with the method called flowAMOEBA (Tao and Thill 2018), which is a data-driven and bottom-up spatial statistic method for identifying spatial flow clusters of extremely high- or low-value, for instance, anomalously large volume of human movements between two regions. The method upgrades the classical spatial clustering method AMOEBA (Aldstadt and Getis 2006) from areal data to flow data by properly defining spatial flow neighborhood. As shown in Figure 1, regarding flow *a*, flow *a'* shares the same origin and destination; the two flow *bs* have the same origin (destination) and a contiguous destination (origin) with flow *a*; flow *c* has both origin and destination contiguous to flow *a*; and neither origin nor destination of flow *d* are the same nor contiguous to flow *a*. Therefore, flow *a'* should be aggregated with flow *a*; flow b and flow c are considered as neighbors of flow *a* while flow d is not.

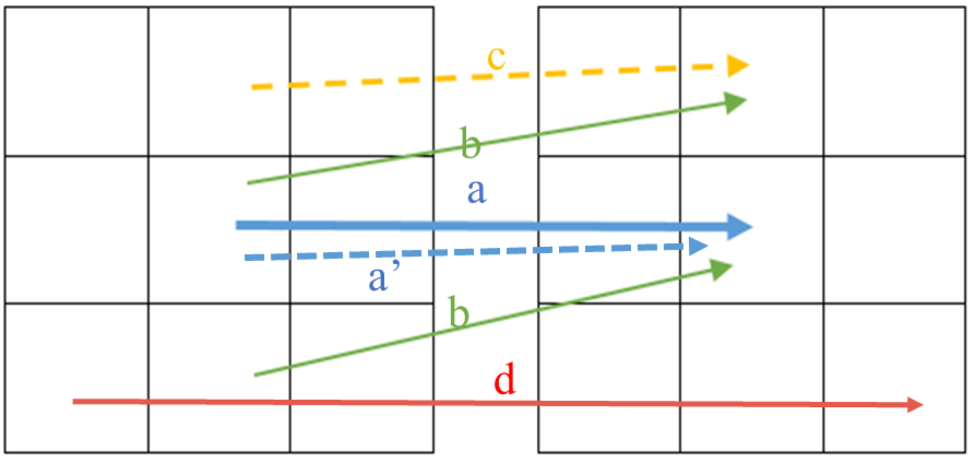


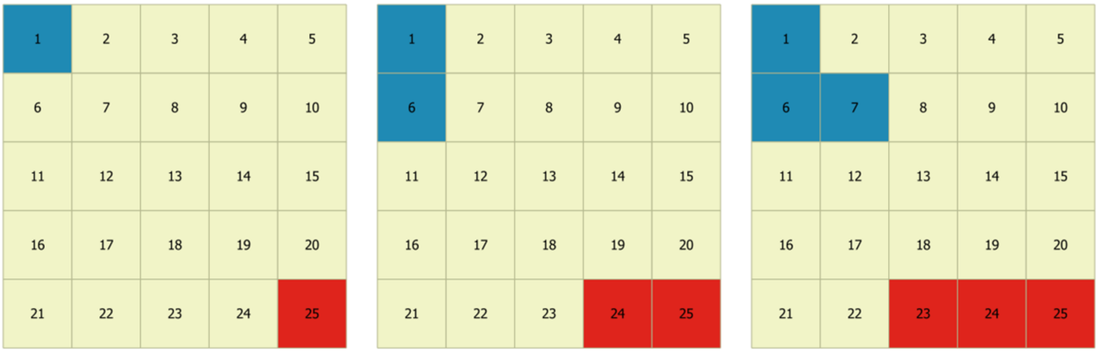
Figure 1 Different cases of flow neighbor relationship (Tao and Thill 2018)

To identify flow clusters, the algorithm starts from an arbitrary seed flow and attempts to iteratively expand the cluster towards the neighboring flows with the aim of maximizing or minimizing the local G_i^*statistic (Getis and Ord 1992; Ord and Getis 1995), defined as follow:

$$G_i^* = \left(\sum_{j=1}^{N} w_{ij}x_j - \bar{x}\sum_{j1}^{N} w_{ij}\right) / S\sqrt{\frac{N\sum_{j=1}^{N} w_{ij}^2 - {\sum_{j1}^{N} w_{ij}}^2}{N-1}}, \text{ where } S = \sqrt{\frac{\sum_{j=1}^{N} x_j^2}{N} - \bar{x}^2}. \quad (1)$$

In this study we set the spatial weight w_{ij} equals 1 if flow *j* neighbors flow *i*, otherwise 0. *N* is the total number of flows. x_j is the value of flow *j*. $\bar{x}$ is the mean value of all flows.

We use Figure 2 to illustrate how flow clusters grow in space. Setting flow $F_{1,25}$ (originates from cell 1 and ends at cell 25) as the seed flow, the algorithm first calculates its G_i^* statistic, namely $G_{1,25}^*$. A positive (negative) $G_{1,25}^*$ sets the goal as searching for high (low)-value flow clusters. Next, the algorithm integrates one of the neighbor flows with flow $F_{1,25}$ and calculates G_i^*statistic again. If the absolute value of G_i^* increases after integration, the neighbor flow is successfully included in the cluster. After all first-order neighbor flows are traversed, the algorithm moves on to the second-order neighbors. This expansion ceases when the absolute value of G_i^* stops increasing. By then it obtains a flow cluster pertaining to the seed flow $F_{1,25}$.



(a) (c) (b)

Figure 2 Illustration of flow cluster expansion, with blue denoting origin and red denoting destination. (a) Seed flow; (b) cluster expands to first-order neighbors; (c) cluster expands to second-order neighbors (Tao and Thill 2018)

The steps above are repeated until every flow has served as a seed. Then, all the flow clusters are sorted by their absolute G_i^* values. The ones with the highest values while without any overlaps in space are preserved. Lastly, a Monte-Carlo simulation is carried out to select the flow clusters that pass statistical significance test as the final results.

3. Preliminary results and discussions

Our individual-level movement data are from the data products of Baidu Map, one of the largest big data venders providing spatial data, mapping and analytics services in China. In 2017, Baidu Map's geolocation data covers 0.6 billion mobile devices with averagely 80 billion times of positioning requests. Specifically, trips were extracted from these requests made during 10 days from April 1st to 10th of 2017 between prefectures in China. For this study, we only focus on two types of trips: 1) trips that originate from Northeast China and end at outside of this region; 2) trips that originate from outside of Northeast China and end at the region.

A set of trip clusters with associated origin and destination cities have been detected. We focus on some of these clusters to demonstrate the usefulness of our method. The first example features the trip-based interaction from cities in Northeast China to Shanghai (Table 1). Table 1a shows the relevant clusters found with the raw number of inter-city trips. One cluster includes cities of Harbin and Changchun as origins, and the other includes cities of Dalian, Shenyang and Anshan as origins. These are all large cities by population in Northeast China, which may contribute to the inflation of their significance in cluster testing simply because more people make more trips. After removing the population effect of the origins, i.e. set flow value as the number of trips divided by population of the origin city, we obtain one large cluster including a number of large-, mid-, and small-size cities that are geographically contiguous (Table 1b). It confirms our expectation that cities with various sizes and relatively high volumes of trips will emerge by controlling for the population effect. Further, population effect may also exist on the destination side. Because, as in this example, Shanghai as the largest city in China, possess its general attractiveness, which may inflate the significance level of interactions and obscure the relative importance of other destinations carrying certain particular utility for trip makers. Table 2b shows two smaller clusters with the account of population effects on both sides. The first cluster includes Harbin, Changchun and adjacent mid- and small-size cities as origins, while the second is a single-pair cluster between Dalian and Shanghai. These patterns are able to reflect the true preferences of trip makers with their choices of destinations that can fulfill some special utilities. For instance, the first cluster indicates business activities and tourist trips, while the second cluster indicates trips by sea in addition to other types of trips because both the origin and destination have ports.

The destination side effect is more evidently illustrated by another example based on trips from

Northeast China to Sanya, a mid-size tourist city. Table 2a shows two single-pair clusters originating from large cities due to the population effect. After accounting for that, clusters disappear. However, by accounting for the effects on both sides, one large cluster re-emerges with varying size of cities in a contiguous region (Table 2b and left part of Figure 3). The real preference is unveiled here with the behind story that many people suffered from the extreme winter in Northeast China have invested in vacation homes in Sanya, which has a warm and nice weather at the southern coastal area of Hainan island. Those people's consumption power during seasonal vacations and holidays attracted more Northeast people to move to Sanya and started local businesses carrying strong Northeast characterises, such as restaurants and retails, to serve the target group of consumers. The cluster pattern for the opposite direction, trips from Sanya to Northeast China, indicates similar implications (right part of Figure 3).

Table 1 Clusters detected for trips from Northeast China cities to Shanghai

Cluster	Origin	Destination
a		
1	Harbin Changchun	Shanghai
2	Shenyang Dalian Anshan	Shanghai
b		
1	Dalian Shenyang Changchun Harbin Yingkou Baishan Anshan Fushun Daqing Liaoyang Panjin Jilin Fuxin Jinzhou Benxi Dandong Mudanjiang Tonghua	Shanghai
c		
1	Changchun Harbin Baishan Daqing Jilin	Shanghai
2	Dalian	Shanghai

Table 2 Cluster detected for trips from Northeast China cities to Sanya

Cluster	Origin	Destination
a		
1	Shenyang	Sanya
2	Harbin	Sanya
b		
1	Harbin Shenyang Changchun Dalian Daqing Jilin Panjin Mudanjiang Benxi Fushun Songyuan Tonghua Jixi Anshan Liaoyuan	Sanya

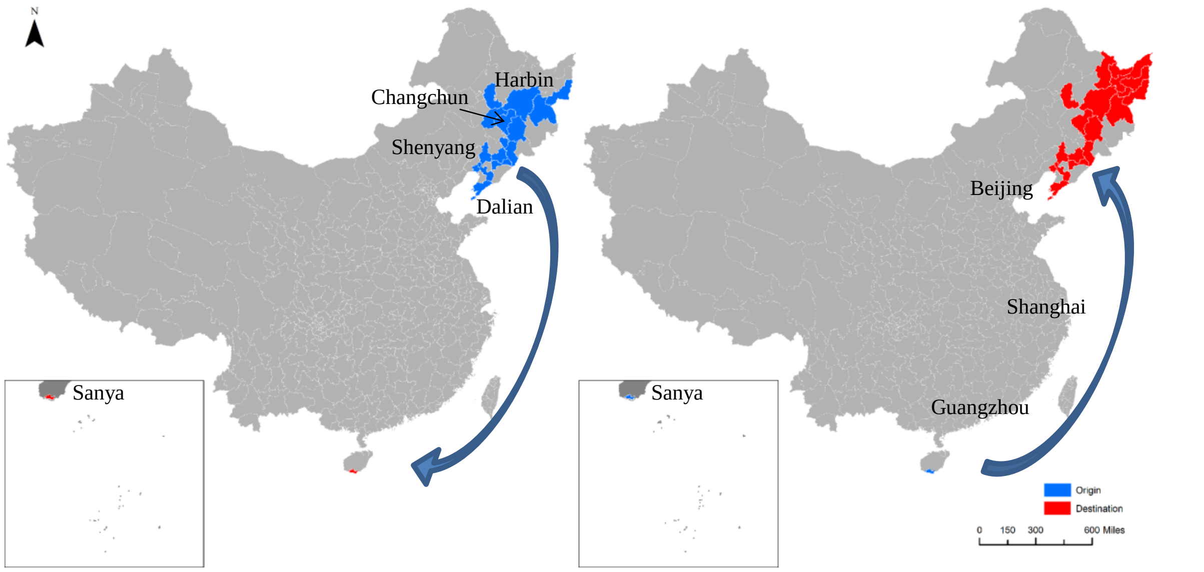


Figure 3 Clusters detected from trips between Northeast China cities and Sanya with population effects on both origin and destination sides removed

4. Conclusions

We employ an exploratory flow clustering method, flowAMOEBA, to the study of human activity based interactions between Northeast China cities and the urban areas outside that region on a short-term basis. This method, considering both spatial heterogeneity and dependency between neighboring flows based on an adjacency setting for their origins and destinations, allows us to discover new patterns of interaction from empirical trip data at the individual level. This research is supported by big data of human movement extracted from mobile device positioning requests. Population effects are examined with selective trip clusters detected to demonstrate the potentials of our approach to unveil the true preference of trip making and to help form new hypotheses with discovered patterns.

5. Biography

Zhaoya Gong is a Lecturer in Human Geography at the School of Geography, Earth and Environmental Sciences at the University of Birmingham. His research centers on leveraging computational and data sciences to advance GIScience, and dynamic processes of urbanisation and complex modalities of spatial structures generated at various scales.

Ran Tao is a Lecturer with the Spatial Sciences Institute in the University of Southern California. His research centers on GIScience methods and applications in spatial interaction, transportation, migration, crime, and civic conflicts.

Qiwei Ma is a PhD candidate at Tsinghua University and a senior urban planner at China’s Academy of Urban Planning and Design.

Changcheng Kan is a senior data analyst at Baidu.com Times Technology (Beijing) Co., Ltd with over six years’ experience in database, data products, machine learning, data mining and other big data-related fields.

References

Aldstadt, J., and Getis, A. (2006). “Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters.” *Geographical Analysis* 38 (4): 327–343.

Haase A, Bernt M, Großmann K, Mykhnenko V, and Rink D (2016). Varieties of shrinkage in European cities. European Urban and Regional Studies, 23(1), 86–102.

Haase A, Rink D, Grossmann K, Bernt, M, & Mykhnenko V (2014). Conceptualizing urban shrinkage. Environment and Planning A, 46(7), 1519–1534.

Getis, A., Ord, J. (1992). “The Analysis of Spatial Association by Use of Distance Statistics.” *Geographical Analysis* 24 (3): 189–206.

Ord, J., and Getis, A. (1995). “Local Spatial Autocorrelation Statistics: Distributional Issues and an Application.” *Geographical Analysis* 27 (4): 286–306.

Ran Tao, Jean-Claude Thill. (Forthcoming). flowAMOEBA: Identifying Regions of Anomalous Spatial Interactions. *Geographical Analysis*.