Detecting urban commercial patterns using a latent semantic information model: A case study of spatial-temporal evolution in Guangzhou, China

Shili Chen¹,², Haiyan Tao¹,³,⁴*, Xuliang Li¹,³,⁴, Li Zhuo¹,³,⁴

¹ School of Geography and Planning, Sun Yat-sen University, Guangzhou, Guangdong, China, ² Urbanization Institute of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, Guangdong, China, ³ Guangdong Provincial Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, Guangdong, China, ⁴ Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou, Guangdong, China

* taohy@mail.sysu.edu.cn

Abstract

With rapid economic growth since the 21st century, cities in China have experienced considerable economic and social reconstruction. Driven by rapid industrialization, urban spatial structures are undergoing evolution and change. Therefore, this paper analyzes the processes and mechanisms associated with the evolution of the commercial spatial structure in Guangzhou after the financial crisis in 2008 based on both theoretical and empirical analyses. We use a Dirichlet multinomial regression (DMR) model to extract latent semantic information and determine urban functional areas from global positioning system (GPS) and point-of-interest (POI) data collected in Guangzhou in 2009 and 2013. In addition, we use movement patterns and POI data to identify the evolution of Guangzhou’s commercial zones from 2009 to 2013. The results show that the urban commercial structure in Guangzhou gradually changed from a single-center model to a multi-center model with dispersed clusters and that the distribution of the entire spatial structure changed. Meanwhile, Guangzhou’s commercial structure not only varied over time but also exhibited specific geographical features. This paper demonstrates that the proposed method can clearly identify the boundary of the commercial area in Guangzhou and provides a valid spatial-temporal model of change in the city. Moreover, this study not only expounds the future development trends of the urban spatial structure in Guangzhou from a microcosmic perspective but also provides a scientific basis for clarifying the spatial locations and development advantages of urban functions within the city.

Introduction

We travel from our residences to our workplaces every weekday. If our workplaces are close to our residences, then we need to travel only a short time and distance to reach our destination.
However, if our workplaces are far away, we must use transportation. Therefore, urban traffic peaks occur during the morning and evening, producing a pendulum traffic pattern. Additionally, as a result of pendulum traffic, some regions have a much larger population during the day than during the night, a phenomenon known as the "hollowing out" of a city center.

This phenomenon occurs as urbanization increases, the urban population surges, and urban built-up areas expand, causing a change in the urban spatial structure. According to statistics, the urban population currently accounts for over 50% of the total population and will likely reach 69.6% by 2050 [1]. The global built-up area is also continually expanding and changing the spatial structure of the urban interior from the concentric circle structure [2], fan-shaped structure [3], and multi-core structure [4] models proposed by the Chicago School to the multiple centers and multi-center networks [5,6] proposed by the Los Angeles School. Such changes in urban spaces are based on changes in the spatial distribution of urban commerce.

Urban businesses serve as the main functions of cities and emerge with the establishment of cities [7]. The commercial space is one of the most active areas in cities and has a significant impact on urban development and the evolution of the urban spatial structure. Therefore, research in this area constitutes an important part of urban geography studies. As business contacts between countries and regions strengthen and more distribution centers and exchange markets emerge, research on commercial structures is improving, especially with the application of a series of scientific theories and methods, such as quantitative geography, behavioral geography and operational research approaches.

Previous studies of commercial space have mainly focused on two research areas involving the intra-urban and spatial structures of urban systems: the commercial carrier and the commercial subject. Most studies of commercial carriers have employed rent theory [8], central place theory [9–11], measurement methods [12–14], etc. to research the commercial space structure. In studies of business subjects, some scholars have proposed research models of consumer behaviors, business circles, etc. [15–25], while others have suggested more diverse models to describe urban commercial space from the perspective of the market structure and size of the urban center [13,24]. Some recent studies have introduced specific hypotheses to create more dynamic and realistic models, especially considering consumer factors [15,16,19] and the changes in commercial center systems based on central place theory [17,18,20,21,23,25]. However, almost all the data used in these studies were derived from traditional surveys, which are expensive and require considerable time and effort. Due to data limitations, these studies did not analyze the structure and evolution of commercial spaces over any given period.

In recent years, the development of information and telecommunication technology and the accessibility to big data have given urban geology a new perspective [26–28], including new research methods to study urban spatial structures. Big data has been widely applied to study urban spatial structures. Such studies have analyzed urban spatial structures and urban functional zoning based on mobile data [29,30], bus data [31–33], GPS data [34–38], and location-based service (LBS) data [39–43]. Additionally, a city’s taxi data reflect the travel needs of residents and have low costs, high accuracy, wide coverage, and immediate availability. Compared to traditional research on urban spatial structures, studies that identify a city’s spatial structure through GPS data not only require less time for field research but also avoid the need for survey results. Moreover, the results of these studies are more accurate due to the large sample size. Some scholars have previously used big data to study people’s travel behaviors [40,44–46]. In addition, some researchers have used taxi data to analyze several aspects of land use. For example, Pan used taxi GPS traces to perform land-use classification and found that taxi pick-up/set-down patterns corresponded to the land-use classes of regions [47]. Yang used
GPS data to analyze land-use patterns in Washington, DC, and found that taxi movement were strongly associated with land-use patterns [48]. In addition, Nong used taxi data to delimit the boundaries of retail centers and determine their hierarchical characteristics [49]. However, few scholars have employed big data to analyze customer behaviors and quantitatively evaluate commercial spatial structures. Furthermore, Chinese studies in this research area are very rare. From the perspective of economics, customer behavior plays a significant role in the formation and evolution of commercial structures. The travel behaviors of taxis in a city reflect the customer behaviors of residents during their activities.

Based on floating car data from 2009 and 2013, this paper takes Guangzhou as an example and identifies its commercial structure after the financial crisis using a latent semantic analysis. Specifically, the study analyzes the mobility patterns of Guangzhou residents between different regions and examines point-of-interest (POI) data on related factors that compose a commercial zone, which are used to identify Guangzhou’s commercial structure. The results reveal that the commercial structure of Guangzhou transformed from a single center to a decentralized, polycentric network between 2009 and 2013, and the distribution of the overall spatial structure underwent considerable change. Meanwhile, the commercial structure of Guangzhou not only changed over time but also displayed unique regional characteristics. The results prove that the adopted method can be used to effectively identify commercial areas in Guangzhou and to investigate the evolution of the commercial spatial structure of this city after the financial crisis.

**Study area and data**

**Study area**

The research area of this paper is the downtown area of Guangzhou in China (longitude 112°57′~114°03′ and latitude 22°26′~23°56′). To ensure the representativeness of the sample, we consider both the spatial distribution and temporal evolution of the research objects and select “eight old districts”. Thus, the traditionally defined downtown area before the administrative division adjustment in 2014 is chosen as the research area (among those eight districts, Fangcun District has been incorporated into Liwan District, and Dongshan District has been incorporated into Yuexiu District). Additionally, we divide the study area of Guangzhou into 439 zones based on the distribution of freeways, city expressways, national highways, provincial highways, urban main roads, and secondary main roads (as displayed in Fig 1). However, because the Baiyun district along the upper part of the Huanan Expressway is mainly composed of tourist attractions, farmlands, and mountains, this zone is excluded from the research area.

**Data**

China’s public transportation system collects passengers’ pickup information but does not record the passengers’ drop-off information. Thus, it lacks complete origin-destination information and cannot provide accurate travel information [50,51]. In addition, due to the fixed ranges of subways and buses, some scholars have shown that the active range of conventional public transport is only 400 meters [52]. The range of a subway station service is generally a 500-meter circular area centered at the station [53], which fails to achieve full coverage of a city. Compared to these types of traffic data, taxi data include pickup and drop-off information and cover an entire city; thus, they are more suitable for exhaustive studies of resident travel.

Meanwhile, many scholars have shown that floating car data can accurately represent road traffic scenarios and human behaviors [54,55]. In Guangzhou, the taxi industry also plays an important role. In terms of the number of public transport passengers, 196 million daily
Fig 1. Study area. The figure shows the location of Guangzhou in China and the study area in Guangzhou.

https://doi.org/10.1371/journal.pone.0202162.g001
passenger taxi trips were recorded in 2009, accounting for 18.67% of all public transport. By October 2013, the average daily passenger volume of taxis reached 2.2 million, accounting for 14.16% of all public transport. Therefore, floating car data can be used to represent the travel needs of urban residents. To illustrate city expansion and changes in land use and spatial structure, the present study utilizes GPS trajectory data from taxis and POI data that describe city buildings. The floating car data include GPS data from May 11, 2009 (Monday), to May 17, 2009 (Sunday), and from October 8, 2013 (Tuesday), to October 14, 2013 (Monday). The GPS data were collected at an interval of five seconds and included basic information such as the plate number of the taxi, time, longitude, latitude, speed, position, occupancy status, etc.

Table 1 shows a sample of taxi data from 2009.

| License Plate | GPS Time (2009/5/11) | Longitude | Latitude | Speed | Direction | Car State |
|---------------|----------------------|-----------|----------|--------|-----------|-----------|
| Yue A000S9    | 09:00                | 113.2825  | 23.14107 | 19     | 360       | 4         |
| Yue A000S9    | 09:01                | 113.2827  | 23.14199 | 14     | 28        | 4         |
| Yue A000S9    | 09:02                | 113.2865  | 23.14364 | 21     | 42        | 4         |
| Yue A000S9    | 09:03                | 113.2903  | 23.14471 | 37     | 67        | 4         |
| Yue A000S9    | 09:04                | 113.2917  | 23.14177 | 18     | 156       | 4         |
| Yue A000S9    | 09:05                | 113.2879  | 23.13731 | 24     | 159       | 4         |
| Yue A000S9    | 09:06                | 113.284   | 23.1344  | 27     | 258       | 4         |
| Yue A000S9    | 09:07                | 113.2831  | 23.1351  | 39     | 344       | 4         |
| Yue A000S9    | 09:08                | 113.2821  | 23.1432  | 0      | 74        | 4         |
| Yue A000S9    | 09:08                | 113.2826  | 23.14195 | 21     | 352       | 5         |

POIs are an essential data type used in navigation, smart transportation, and other LBSs. The POI data in this paper can be divided into 15 categories and 65 types. Each POI is numbered based on the “category name + type name + serial number”, including information such as ID, type, longitude, and latitude. Table 2 shows some sample POI data. According to the objectives and requirements of this study, we combine some of the POI data and divide the data into 29 categories. It is worth noting that this paper does not consider the impacts of different classes of POIs on the results.

Table 2. Sample POI data.

| ID     | Type    | RDID     | Longitude | Latitude |
|--------|---------|----------|-----------|----------|
| 20000001 | 81406024 | 20131990 | 113.604   | 22.6478  |
| 20000002 | 81102065 | 20123966 | 113.608   | 22.6464  |
| 20000003 | 81402034 | 20131990 | 113.604   | 22.6474  |
| 20000004 | 80601054 | 20131990 | 113.604   | 22.6473  |
| 20000005 | 81506044 | 20156064 | 113.566   | 22.6554  |
| 20000006 | 80101035 | 20119062 | 113.593   | 22.6292  |
| 20000007 | 81004035 | 20046330 | 113.575   | 22.6407  |
| 20000008 | 81406005 | 20021084 | 113.609   | 22.6431  |
| 20000010 | 80101062 | 20046302 | 113.606   | 22.6449  |
| 20000011 | 81201015 | 20046305 | 113.605   | 22.644   |
| 20000012 | 80701054 | 20150230 | 113.623   | 22.6288  |
| 20000013 | 81102005 | 20046311 | 113.607   | 22.6589  |
| 20000018 | 80603045 | 20021131 | 113.607   | 22.6447  |
| 20000019 | 81001005 | 20000251 | 113.568   | 22.6531  |
Methodology

The Dirichlet multinomial regression (DMR) model is a topic model that was proposed by Blei in 2003 [56]. As a widely recognized format for text processing, it can build a model to determine the hidden topics in documents. This approach not only improves upon traditional methods of text similarity calculation but also caters to searches of semantic topics based on large corpora or even massive internet data [56–58]. The model can determine the probability of multiple topics in each document from a corpus and can fully extract the semantic information from a word or sentence. Thus, we use DMR model to analysis the GPS data.

According to the DMR model, each document in the corpus can be regarded as a combination of multiple topics. Each word in the document is associated with a topic. Therefore, if all the words in a document are available, the distribution of topics can be obtained through mathematical derivation.

Urban functional zoning and text representation studies are similar to a certain extent. If we regard a region as a document and a function as a topic, then every trip in a region can be considered a word, and a functional zone can be characterized by its agglomeration of activities, intraregional transport infrastructure, resident mobility, and inputs within its borders of interaction [58]. The DMR model [45] is displayed in Fig 2.

Nodes represent random variables and are denoted using different colors and shapes according to their roles in the generative process: white represents latent variables, gray represents observable variables, rectangles indicate that the variables require loop computation, and arrows represent parameters and the directions of variables.

In Fig 3, α and η represent the input parameter of the distribution of the Dirichlet region functional zone and the distribution of the residential travel functional zone, respectively. Assuming that a certain region has K functions, β is a matrix of K * M (M represents the number of trips of all residents in city D). Each βk is a distribution throughout the city. The proportion of a functional zone in the dth region is θd, and θd,k is the proportion of functional zone K in city d. The functional zone distribution of city d is Zd, and Zd,n represents the functional distribution of resident n in city d. The number of observed trips of residents in city d is represented by Wd, and Wd,n stands for the number of trips of resident n in city d.

Moreover, the α value from the DMR model accounts for the POI feature vector of each region αr. For example, \( \alpha_r = \exp(X_r^T \lambda_k) \). Therefore, different combinations of POI category distributions will yield distinct α values. Hence, the activity distribution is the sum of the POI features and mobility patterns. Finally, by applying DMR and inputting the mobility patterns
and POI features, we obtain the activity distribution of each region and the mobility pattern distribution of each activity.

Mobility pattern $M$ is recorded as a tuple:

$$M = < O_i, T_i, D_j, T_j >$$  

where $O_i$, $T_i$, $D_j$, and $T_j$ represent the departure zone, departure time from the departure zone, destination zone, and arrival time in the target zone, respectively. If one departs from zone $S$ at time $T$ and arrives in destination zone $X$, then "O_X_T" is the mobility pattern of the departure zone, where "O" stands for the departure zone $S$, and "S_D_T" is the mobility pattern of the destination zone, where "D" represents the destination zone $X$.

First, we extract the origin-destination (O/D) pairs from the massive floating car data set and link the O/D pair with the research area to determine the area of each trip. Then, Formula 1 is used to perform one-to-one mapping from the O/D pairs to the mobility pattern. An example of mobility pattern is 305_D_21, which represents a taxi trip starting from area 305 and ending in area 21. Next, we build a DMR model based on the Mallet platform [59] to process the mobility patterns. After processing the DMR model, we obtain the probability distribution of each topic of each functional region and employ ordering points to identify clustering structure (OPTICS) to determine the commercial zone. Finally, we identify and analyze the commercial zone.

The OPTICS method is a density-based spatial clustering method. This method overcomes the limitation of global parameters set by the density-based spatial clustering of applications with the noise (DBSCAN) algorithm, which cannot perform clustering operations under non-uniform spatial data distributions. The OPTICS method is an improved DBSCAN algorithm. To illustrate, Fig 3 shows four different density clusters, A, B, C and D; if the value of $\varepsilon$ is large, the A cluster can be identified, but neither A and B nor C and D can be separated. Furthermore, when the value of $\varepsilon$ is small, cluster A will be difficult to recognize.

To solve the problem of spatial data density distribution, the OPTICS algorithm is added to the density-based clustering sorting.
Here, sort represents the density structure of spatial data, and the information expressed is equivalent to the spatial clustering structure obtained with different parameter values (e.g., neighborhood radius). Unlike the DBSCAN algorithm, the OPTICS algorithm stores the order in which spatial entities are processed, providing a series of neighborhood radius parameter values and allowing spatial entities to be clustered from high density to low density at one time. To achieve this operation, the OPTICS algorithm introduces the concepts of core distance and reachable distance, calculating the kernel distance and a suitable reachable distance for each spatial entity. Using this information, we can identify several clusters of data with different densities.

Results
Following the abovementioned procedures and OPTICS clustering, we obtain the resulting commercial structure of Guangzhou city (as displayed in Figs 4 and 5).

Discovering commercial space
After constructing a model to determine the zone clustering of areas with the same functions, it is essential to identify functional zones based on their actual functions. There are various criteria for urban functional zoning, and the partitioning results are distinct. Identification of all of the functional areas was performed in a previous paper [34] and is not repeated here.

![Fig 4. The functional zones in Guangzhou in 2009.](https://doi.org/10.1371/journal.pone.0202162.g004)
This paper partitions the study region according to social functions, residents’ needs, and other factors to identify the commercial zone. In our opinion, a mature commercial zone requires a long-term development strategy that provides a balanced combination of shopping, restaurants, activities, facilities, and other attractions near the area. The travel pattern not only includes travel peaks in the morning and evening but also appeals to the city’s residents. Therefore, we calculate the distribution of POIs in the research area and the flow of taxis to the commercial zone each week, day, and hour to identify the commercial structure. Because the taxi flow in 2009 is similar to that in 2013, we use only the taxi flow information from 2013 in further analyses due to the manuscript length limitations.

As displayed in Figs 4 and 5, we obtain the spatial distribution of nine zones (F0~F8) through clustering. Figs 6 and 7 demonstrate the travel flows on weekdays and weekends in different zones, i.e., the temporal distributions of departures and arrivals. Zone F3 maintains the highest flow. Moreover, the flow in this zone increases significantly on weekends, which implies that there are significantly more visits to this zone on weekends.

In addition, we select POIs for banking, insurance, restaurants, accommodations, recreation, and other services as the distribution types in the commercial zone and obtain the results shown in Fig 8.

By further analyzing the departure flows on weekdays/weekends, we find that the flows in zones F3 and F4 are essentially identical. Nevertheless, the overall flow in zone F4 is smaller than that in zone F3 but undergoes more changes. Meanwhile, based on the distribution types of POIs, zone F3 has the widest coverage of different types of POIs, among which restaurants,
Fig 6. The different functional zones of the (a) departures and (b) arrivals of taxis in 2013.

https://doi.org/10.1371/journal.pone.0202162.g006

Fig 7. (a) Departure flows in different zones on weekdays (b) Departure flows in different zones on weekends.

https://doi.org/10.1371/journal.pone.0202162.g007

Fig 8. The distribution of commercial points of interest. The figure shows the business districts from Fig 5 and commercial POIs in the research area.

https://doi.org/10.1371/journal.pone.0202162.g008
shopping, and offices account for relatively large proportions and are associated with integrated service attributes. In contrast, zone F4 features residential buildings and living services associated with living space attributes. Therefore, we believe that zone F3 is a commercial zone, whereas zone F4 is a mature residential zone.

In addition, zones F3 and F4 have similar “pendulum traffic” patterns because the commercial areas and residential areas in Guangzhou are highly mixed. Because the commercial zone is more attractive than the residential zone, it has a larger flow of people. In terms of the flow variations, the residential zone displays more regularity in travel behaviors in the morning and evening, resulting in more variations in the flow. However, the commercial zone attracts both regular residents and irregular visitors and thus has a flow that slightly varies. This variation constitutes the major criterion for us to distinguish between the commercial zone and the mature residential zone.

**Human mobility based on taxi data in commercial regions/zones**

By examining the times of flows in the commercial zone based on the departure and arrival flow data, we find that the flow in this zone displays well-defined regularity on weekdays and weekends. As shown in Figs 9 and 10, the departure and arrival flows in the commercial zone reach peaks in the morning and evening on weekdays. The arrival flows peak at 9:00–10:00 am, reflecting the need of commuters to go to work during that period. The evening flow begins to increase at 7:00 pm, and its peak flow surpasses that in the morning, indicating that many residents outside the functional zone also travel to the commercial zone for work,
shopping, recreation, etc. In addition, although the weekend flow is higher than that on weekdays, it does not change significantly over time. Thus, a more balanced distribution of flows in the commercial zone is observed on weekends, which conforms to the typical mobility pattern in the commercial zone. Moreover, there is a distinct growth in the flow of arrivals to the commercial zone in the early hours of Saturday, which can be explained by the unique weekend nightlife in Guangzhou, a first-tier city.

**The evolution of the urban commercial structure after the financial crisis and result verification**

The 2008 global financial crisis caused considerable changes to the development of many countries and cities, and Guangzhou is no exception. Using taxi trajectory data, we can analyze some changes to the commercial structure of Guangzhou, which are shown in Figs 4 and 5. The results reveal that the commercial space in 2009 was distributed in fragments around the city center. By 2013, the number of business fragments decreased, and many fragments combined and spread, covering the entire city of Guangzhou (as displayed in Fig 11). This result was likely caused by the enhancement in residents’ quality of life and the diverse economic development led by the service industry, which promoted the change in Guangzhou’s commercial structure from a single center to a polycentric network.

In different regions of a city, commercial spatial structures undergo different changes: some commercial zones remain, some vanish, and some expand. Located on the central axis of old

![Fig 11. Evolution of the commercial structure from 2009 to 2013.](https://doi.org/10.1371/journal.pone.0202162.g011)
town, the two famous business circles, Beijing Road (A) and Shangxiajiu Pedestrian Street (B), have time-honored cultures and business traditions as well as a variety of stores (as shown in Fig 11). They are retail business circles with integrated services. From 2009 to 2013, their coverage area increased slightly. Thus, the two business circles have sprawled outward and developed into new business zones. As two important community business circles, China Plaza (Nonglinxia Road) (C) and Jiangnanxi (D) serve primarily the surrounding local residents. With a static and limited range of target customers, these two business circles did not experience significant changes in their spheres of influence from 2009 to 2013. Fig 11 shows the distribution of the remaining business zones in the region.

Some business zones have lost their original business attributes due to upgrades and transformations. The Xiwan Road-Zengcha Road area (E), which has gradually developed into a hub of wholesale markets and logistic and freight services, is an example. Conversely, some areas experienced increases in residential or cultural and recreational functions and gradually converted to living spaces from commercial spaces during the period 2009–2013. Shipaidong Road-Wushan Road (F) and the northern part of Xianliedong Road to Yanling Road are examples of increased residential function. For example, the Gangding and Shipai area has numerous residential buildings but lacks sufficient offices and warehouses. Additionally, its main industry, IT retail sales, has suffered in recent years due to the economic downturn and the impact of electronic commerce, which have greatly decreased its commercial influence. Examples of growth associated with cultural and recreational functions include the northern part of Zhongshanqi Road to Zhongshanba Road (G) and the areas surrounding Tianhe Park. For example, the Chen Clan Ancestral Hall located on Zhongshanba Road has become a tourism landmark in Guangzhou.

The new business zones reflect the long-term strategic plan in Guangzhou. The city’s commercial spatial structure has changed from a cluster in the city center to a network covering the entire city and its surroundings. The area of Guangbai and Qifu Road (H) around the new town possesses integrated commercial functions and promotes the development of the new urban district in the northern part of the city. Due to the prime location of Fangcun, the west bank of the Pearl River has developed a commercial circle, Huadiwan-Fangcun (I), which connects and reaches to neighboring cities. The development of Huangpu District (J) in the eastern part of the city has proven effective, and new business zones are gradually forming there. Numerous emerging business districts are developing rapidly and are connected to the existing business areas along the central axis, including the high-end office zone of Zhujiang New Town (K) and exhibition centers such as the International Convention and Exhibition Center in Pazhou (L). For instance, Yuancun, an extension of the Zhujiang New Town CBD, was transformed from an old industrial zone into a residential area and then into a financial hub with financial and commercial functions from 2009 to 2013. Thus, it is currently associated with a commercial structure. In summary, the developments of these emerging commercial zones are based on the planning strategy in Guangzhou: “develop the south, optimize the north, move toward the east, and connect the west”. The city is adjusting its spatial structure to enhance its city functions and to convert from a single center to a polycentric network.

We use data from the “Guangzhou Large-scale Retail Business Network Development Plan (2011–2020)” released by the Guangzhou Planning Bureau in 2013 as the standard to verify the commercial functional zoning results produced in our research. The data show that by the end of 2010, Guangzhou completed the construction of commercial functional areas in five metropolitan zones (Beijing Road, Central City East, Agriculture and Forestry Road, Zhongshan Three Road, and Thirteen Line-Shangxiajiu Tianhe Road) and seven district zones (Jiangnan, Xiaogang, Shijing-Xinshi, Dashadi, Shiqiao, Licheng, and Jiekou). Excluding
Shiqiao, Licheng, and Jiekou, which are shown in Figs 4 and 5, these commercial functional areas are outside the study area. Thus, the proposed method can effectively identify the distribution of business districts.

**Discussion**

To validate the results regarding the commercial zones, this research uses POI data to identify commercial areas. The steps are as follows: First, the POI data are divided into 29 types. Second, the probability of the occurrence of each type of POI in each research area is calculated. Third, the statistical results are used as the input data in the DMR model. Finally, the OPTICS clustering method is used to cluster the results of the DMR model and obtain the results for the commercial zones (as displayed in Fig 12). The results based on the POI data and the DMR model are shown in Fig 12(a) and 12(b) presents the commercial zones based on the GPS data, POI data and DMR model.

Land-use maps are very difficult to obtain in China. Therefore, through other related projects, this study obtains a limited land use status map of the research area, which is used for comparison with the results shown in Fig 12(a) and 12(b). Although this land-use map is limited, it is very informative. Through comparative analysis, this study finds that POI data and DMR models can be used to correctly identify some commercial zones. However, many areas are misidentified as commercial areas, especially in remote areas, such as the mountainous area (A) and the mountainous plains mixed area (B) in the east of Huangpu District, the agricultural area (C) around the Beijing-Shenzhen line, and the southwest corner of the ring expressway (D). These areas include some commercial facilities, and the POI data for these areas are very sparse. Therefore, POI-sparse areas with few commercial POI facilities can be misidentified as commercial areas. With the development of the city, although many commercial facilities are preserved, some commercial areas gradually transform into residential areas, such as Wushan Campus (E) and Liuhuahu Park (F). Furthermore, the areas of Shangxiuju (G), Yuexiu District (H) and Chigang (I) belong to traditional commercial areas but are not identified by the DMR model, which may be caused by the inability of POI data to capture the long-term development of these areas.

Wu used POI data to identify multiple types of commercial centers [60]. This author’s results showed that the commercial area structure presents a dual-core distribution, with the urban services center and public service center located in Yuexiu District and the business center and financial center located in Tianhe District. These results are consistent with our findings shown in Fig 12(b). However, this research also revealed that POI data have the characteristics of large data volume but low information, as they ignore detailed information on the development characteristics and building age of geographical entities. Therefore, when identifying the boundary of the central area through POI data, the development characteristics and level information of the central area are easily overlooked, which will affect the accuracy of the boundary. Simultaneously, the model results cannot be used to analyze the development status of the central area.

Overall, the extraction of commercial areas using POI data can reflect the distribution of commercial facilities to some extent. However, the classification results cannot reflect the distribution of actual commercial zones. It is difficult to differentiate commercial areas based only on POI data; on the one hand, there are always more commercial POIs than residential POIs in an area at the same time. Therefore, for areas in which mining distribution features, most regions generally have commercial properties. On the other hand, due to the limitation of POI data, the model tends to ignore some key information such that the frequency of POIs cannot represent the potential structure of each region well.
Fig 12. (a) Commercial zone results based on POI data, (b) commercial zone results based on GPS and POI data, and (c) land-use map. These figures show three commercial districts in the research area. (a) presents the commercial zone results based on the POI data.

https://doi.org/10.1371/journal.pone.0202162.g012
This study also compares the results of Fig 12(b) with the land-use map in Fig 12(c). Although there are some misclassifications in the results, such as around the Tuhua overpass (J) and Shabei overpass (K) (in the northwestern and southeastern corners of the city ring), some emerging communities have increased in population but lack the support of commercial facilities, which are insufficient to form a new commercial space. Furthermore, in the traditional business district of Sports Center (L), the large number of people in transit reflects the high degree of functional mix, which makes the DMR model unable to identify the region as a commercial space. Overall, the results of the DMR model are closer to the actual land-use conditions than are the results of the DMR model, indicating that the use of GPS and POI data to discover urban functional zones provides more reliable results than the DMR model.

Conclusion

From the traditional perspective, qualitative analyses of urban commercial spatial structures and the associated changes have certain limitations, and the temporal resolution is low. With big data growing in popularity, obtaining information regarding the changes in a commercial spatial structure in a timely and effective manner has become possible. Based on the concepts of text categorization and traditional urban functional zoning, this paper uses big data associated with the daily movements of people and a semantic theme model (DMR model) to study the urban commercial districts in Guangzhou. Based on traditional methods, the study explores the research concepts and provides a new method that uses residents’ behaviors to study the urban commercial spatial structure; the results are important for guiding future studies and management strategies.

First, this paper discusses a new method of identifying urban commercial space, and the results are in accordance with the urban planning strategy in Guangzhou. This result suggests that taxi and POI data and the DMR model can be used to determine the spatial distribution and changes in an urban commercial space. In addition, after obtaining the distribution of commercial districts in Guangzhou in 2009 and 2013, the paper analyzes the trends and factors related to the evolution of the commercial space. The study also shows that the urban commercial structure in Guangzhou gradually changed from a single-center model to a multi-center model with dispersed clusters from 2009 to 2013. Moreover, the distribution of the other functional districts in the city and the overall spatial structure changed. Although the distribution remains heterogeneous among regions, these regional differences will likely diminish.

Furthermore, through the analysis of different regions, we find that the evolution of urban commercial districts reflects not only temporal changes but also specific geographical features. Early or historical commercial circles and community-oriented commercial districts for residents were largely retained. The commercial development of historic commercial circles was accompanied by a revival in the tourism industry. Additionally, community-oriented commercial districts are less likely to change because of the stability of the service groups in those areas. Moreover, under certain conditions, the commercial structure will begin to “recede” or “improve” based on a multi-center model, although the reasons for the two trends are different. Recession occurs because other functions in the region (such as leisure and residential functions) gradually become dominant, as they are affected by market and consumer behaviors. Conversely, improvement is associated with the macro-control and guidance provided by city planning, and these advantages are used to overcome the limitations of market-based economic development.

Although the study identifies and analyzes the commercial space in Guangzhou from the aspect of a floating car movement mode, the distributions of commercial space in different cities are heterogeneous, and the evolution processes are different. The factors that affect changes
in the distribution of commercial space in different regions are complicated and include the economy, society, culture, and other factors. If we rely only on the floating car movement mode, we cannot obtain a comprehensive summary of the movement in a city; however, the innovative research methods applied and the resulting trends are still noteworthy. Therefore, we use the multi-source traffic model based on economic, social, cultural, and other factors to study the urban commercial space structure and changes in China, and this model will be used and improved in future studies.

Supporting information

S1 Table. Data used in Fig 6. This comma separated value (CSV) file contains the departure taxi flow presented in Fig 6(a) and the arrival taxi flow data presented in Fig 6(b). The first column contains functional zone data, and the other column contains the daily taxi flow of each functional zone.

(SVG)

S2 Table. Data used in Fig 7. This comma separated value (CSV) file has 9 columns; the first column contains time data, and the remaining columns provide flow data for each functional zone. The data used in Fig 7(a) are the departure flows on weekdays, and those used in Fig 7(b) are the departure flows on weekends.

(SVG)

S3 Table. Data used in Fig 9. This comma separated value (CSV) file has 8 columns; the first column contains time data, and the remaining columns provide daily departure flows in the commercial zone.

(SVG)

S4 Table. Data used in Fig 10. This comma separated value (CSV) file has 8 columns; the first column contains time data, and the remaining columns provide daily arrival flows in the commercial zone.

(SVG)

Acknowledgments

The authors would like to express their sincere thanks to the journal editor, the guest editors of this special issue, and the four anonymous reviewers for their valuable suggestions and considerate support, as they have made the publication of this paper possible. This work was supported by the National Natural Science Foundation of China under grant 41371499 and Natural Science Foundation Team Project of Guangdong under grant 2014A030312010.

Author Contributions

Conceptualization: Shili Chen.

Data curation: Shili Chen, Haiyan Tao.

Formal analysis: Shili Chen.

Funding acquisition: Haiyan Tao, Li Zhuo.

Investigation: Shili Chen.

Methodology: Shili Chen, Haiyan Tao, Li Zhuo.

Project administration: Haiyan Tao.
Resources: Shili Chen, Li Zhuo.
Software: Shili Chen, Xuliang Li.
Supervision: Shili Chen.
Validation: Shili Chen.
Visualization: Shili Chen, Xuliang Li.
Writing – original draft: Shili Chen.
Writing – review & editing: Shili Chen, Haiyan Tao.

References
1. United Nations. World urbanization prospects: the 2009 revision. New York: United Nations, Department of Economic and Social Affairs, Population Division; 2010.
2. Burgess EW. The growth of the city: an introduction to a research project. City. 2007; 18: 47–62.
3. Hoyt H. The structure and growth of residential neighborhoods in American cities. Atlanta: Federal Highway Administration; 1939.
4. Harris CD, Ullman EL. The nature of cities. Ann Am Acad Pol Soc Sci. 1945; 242: 7–17. https://doi.org/10.1177/000271624524200103
5. Scott AJ, Soja EW. The city: Los Angeles and urban theory at the end of the twentieth century. Urban Geogr. 1996; 18: 740–743.
6. Gordon P, Richardson HW, Wong HL. The distribution of population and employment in a polycentric city: the case of Los Angeles. Environ Plann A. 1986; 18: 161–173. https://doi.org/10.1068/a18016
7. Jones K, Simmons J. Location, location, location: analyzing the retail environment. Scarborough, Canada: Nelson Canada; 1993.
8. McLafferty SL, Ghosh A. Multipurpose shopping and the location of retail firms. Geogr Anal. 1986; 18: 215–226. https://doi.org/10.1538/4632.1986.tb00094.x
9. Berry BJL, Garrison WL. A note on central place theory and the range of a good. Econ Geogr. 1958; 34: 304–311. https://doi.org/10.2307/142348
10. Sang Y, Xiao L. A review of the theory and research method in commercial geography. Hum Geogr. 2003; 18: 15–17.
11. Christaller W. Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmaßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen. University New York: University Microfilms; 1933.
12. Cadwallader M. A behavioral model of consumer spatial decision making. Econ Geogr. 1975; 51: 339–349. https://doi.org/10.1037/142918
13. Berry BJ, Tennant RJ. Chicago commercial reference handbook. Chicago: University of Chicago Press; 1963.
14. Golledge RG, Stimson RJ. Spatial behavior: a geographic perspective. Econ Geogr. 1997; 74: 83–85.
15. Huff DL. A probabilistic analysis of shopping center trade areas. Land Econ. 1963; 39: 81–90. https://doi.org/10.2307/3144521
16. Rushton G. Analysis of spatial behavior by revealed space preference. Ann Assoc Am Geogr. 1969; 59: 391–400. https://doi.org/10.1111/j.1467-8306.1969.tb00678.x
17. Parr JB. Frequency distributions of central places in southern Germany: a further analysis. Econ Geogr. 1980; 56: 219–223. https://doi.org/10.2307/142932
18. Berry BJL, Parr JB, Epstein BJ, Ghosh A, Smith RHT. Market centers and retail location: theory and applications. Englewood Cliffs, NJ: Prentice Hall; 1988.
19. Berry BJL, Garrison WL. Recent developments of central place theory. Pap Reg Sci. 1958; 4: 107–120. https://doi.org/10.1011/j.1435-5597.1958.tb01625.x
20. Townsend AR, Bromley RDF, Thomas CJ. Retail change: contemporary issues. London: UCL Press; 1993.
21. Dawson JA. Retail geography. London: Croom Helm; 1980.
22. Golledge RG, Stimson RJ. Spatial behaviour: a geographic perspective. London: Guilford Press; 1997.
23. Davies RL. Structural models of retail distribution: analogies with settlement and urban land-use theories. Trans Inst Br Geogr. 1972; 63: 59–82. https://doi.org/10.2307/621554
24. Garrison WL, Berry BJ, Marble DF, Nystuen JD, Morrill RL. Studies of highway development and geographic change. Seattle: University of Washington Press; 1959.
25. Parr JB. Temporal change in a central-place system. Environ Plann A. 1981; 13: 97–118. https://doi.org/10.1068/a130097
26. Goodchild MF. Citizens as sensors: the world of volunteered geography. GeoJournal. 2007; 69: 211–221. https://doi.org/10.1007/s10708-007-9111-y
27. Goodchild MF. Geographic information systems and science: today and tomorrow. Ann GIS. 2009; 1: 1037–1043.
28. Li L, Goodchild MF, Xu B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartogr Geogr Inform Sci. 2013; 40: 61–77. https://doi.org/10.1080/15230406.2013.777139
29. Reades J, Calabrese F, Ratti C. Eigenplaces: analysing cities using the space—time structure of the mobile phone network. Environ Plann B. 2009; 36: 824–836. https://doi.org/10.1068/b34133t
30. Herrera JC, Work DB, Herring R, Ban XJ, Jacobson Q, Bayen AM. Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment. Transp Res C Emerg Technol. 2010; 18: 568–583. https://doi.org/10.1016/j.trc.2009.10.006
31. Joh CH, Hwang CA. A time-geographic analysis of trip trajectories and land use characteristics in Seoul metropolitan area by using multidimensional sequence alignment and spatial analysis. Washington, DC: AAG Annual Meeting; 2010.
32. Zhong C, Huang X, Müller Arisona S, Schmitt G, Batty M. Inferring building functions from a probabilistic model using public transportation data. Comput Environ Urban Syst. 2014; 48: 124–137. https://doi.org/10.1016/j.compenvurbsys.2014.07.004
33. Sun L, Lee DH, Erath A, Huang X. Using smart card data to extract passenger’s spatio-temporal density and train’s trajectory of MRT system. In: Proceedings of the ACM SIGKDD international workshop on urban computing. New York: ACM; 2012, pp. 142–148.
34. Chen S, Tao H, Li X, Zhuo L. Discovering urban functional regions using latent semantic information: spatiotemporal data mining of floating cars GPS data of Guangzhou. J Geogr Sci. 2016; 71: 471–483.
35. Li Q, Zhang T, Wang H, Zeng Z. Dynamic accessibility mapping using floating car data: a network-constrained density estimation approach. J Transp Geogr. 2011; 19: 379–393. https://doi.org/10.1016/j.jtrangeo.2010.07.003
36. Qi G, Li X, Li S, Pan G. Measuring social functions of city regions from large-scale taxi behaviors. In: 2011 IEEE international conference on Pervasive Computing and Communications Workshops (PERCOM workshops). Piscataway, NJ: IEEE; 2011, pp. 384–388.
37. Yue S, Chai Y. Study on commuting flexibility of residents based on GPS data: a case study of suburban mega-communities in Beijing. Acta Geogr Sin. 2012; 67: 733–744.
38. Gao S, Wang Y, Gao Y, Liu Y. Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. Environ Plann B. 2013; 40: 135–153. https://doi.org/10.1068/b38141
39. Gaubatz P. Changing Beijing. Geogr Rev. 1995; 85: 79–96. https://doi.org/10.2307/215557
40. Yin Z, Cao L, Han J, Zhai C, Huang T. Geographical topic discovery and comparison. In: Proceedings of the 20th international conference on World Wide Web. New York: ACM; 2011, pp. 247–256.
41. Pozdnoukhov A, Kaiser C. Space-time dynamics of topics in streaming text. In: Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks. New York: ACM; 2011, pp. 1–8.
42. Cranshaw J, Schwartz R, Hong JI, Sadeh N. The Livhoodes Project: utilizing social media to understand the dynamics of a city. In: International AAAI conference on weblogs and social media. Palo Alto, CA: AAAI; 2012. p. 58.
43. Kling F, Pozdnoukhov A. When a city tells a story: urban topic analysis. In: Proceedings of the 20th international conference on advances in geographic information systems. New York: ACM; 2012. pp. 482–485.
44. Yuan J, Zheng Y, Xie X. 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. New York: ACM; 2012. pp. 186–194.
45. Pulliam HR. Sources, Sinks, and Population Regulation. Am Nat. 1988; 132: 652–661. https://doi.org/10.1086/284880
46. Liu Y, Wang F, Xiao Y, Gao S. Urban land uses and traffic ‘source-sink areas’: evidence from GPS-enabled taxi data in Shanghai. Landsc Urban Plann. 2012; 106: 73–87. https://doi.org/10.1016/j.landurbplan.2012.02.012
47. Pan G, Qi G, Wu Z, Zhang D, Li S. Land-use classification using taxi GPS traces. IEEE T Intell Transp. 2013; 14: 113–123. https://doi.org/10.1109/TITS.2012.2209201

48. Yang Z, Franz ML, Zhu S, Mahmoudi J, Nasri A, Zhang L. Analysis of Washington, DC taxi demand using GPS and land-use data. J Transp Geogr. 2018; 66: 35–44. https://doi.org/10.1016/j.jtrangeo.2017.10.021

49. Nong Y, Zhou S, Liu L, Li Q, Peng Y, Hao X. Structural cities: delimiting retailing center boundaries and their hierarchical characteristics in urban China based on GPS-enabled taxi data. J Plan Educ Res. 2017.

50. Zhou T, Zhai C, Gao Z. Approaching bus OD matrices based on data reduced from bus IC cards. Urban Transp China. 2007; 5: 48–52.

51. Song Z, Qin Z, Xu J, Wu D. 2016. Research on large-scale OD matrix estimation method based on bus IC card data. Appl Res Comput. 2016; 33: 2007–2013.

52. Johnson A. Bus transit and land use: illuminating the interaction. J Public Transp. 2003; 6: 21–39. https://doi.org/10.5038/2375-0901.6.4.2

53. Chen W, Liu L, Liang Y. Characterizing the spatio-temporal evolution of retail business at transfer hubs of Guangzhou Metro. Acta Geogr Sin. 2015; 70: 879–892.

54. Chen M, Chien S. Determining the number of probe vehicles for freeway travel time estimation by microscopic simulation. Transp Res Rec J Transp Res Board. 2000; 1719: 61–68. https://doi.org/10.3141/1719-08

55. Zhang CB, Yang XG, Yan XP. Probe vehicles sample size for mobile traffic detection system. China J Highway Transp. 2007; 20: 96–101.

56. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res. 2003; 3: 993–1022.

57. Wei X, Croft WB. DMR-based document models for ad-hoc retrieval. In: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM; 2006. pp. 178–185.

58. Liu Y, Liu X, Gao S, Gong L, Kang C, Zhi Y, et al. Social sensing: a new approach to understanding our socioeconomic environments. Ann Assoc Am Geogr. 2015; 105: 512–530. https://doi.org/10.1080/00045608.2015.1018773

59. Mimno D, McCallum A. Topic models conditioned on arbitrary features with dirichlet-multinomial regression. Amherst: University of Massachusetts; 2012. pp. 411–418.

60. Wu K, Zhang H, Wang Y, Wu Q, Ye Y. Identify of the multiple types of commercial center in Guangzhou and its spatial pattern. Progress in Geography. 2016; 35: 963–974. https://doi.org/10.18306/dlkxjz.2016.08.005