

Article

Discovering Urban Functional Polycentricity: A Traffic Flow-Embedded and Topic Modeling-Based Methodology Framework

Shaodong Wang ^{1,2,*}, Yanbin Liu ^{3,*}, Wei Zhi ³, Xihua Wen ³ and Weihua Zhou ¹¹ School of Management, Zhejiang University, Hangzhou 310058, China; larryzhou@zju.edu.cn² Department of Public Policy, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong 999077, China³ CETHIK Group Corporation Research Institute, Hangzhou, Zhejiang 311121, China; zhiwei@cetiti.com (W.Z.); wenxh@cetiti.com (X.W.)

* Correspondence: wangshaodong@zju.edu.cn (S.W.); liuyanbin@cetiti.com (Y.L.)

Received: 23 January 2020; Accepted: 29 February 2020; Published: 2 March 2020



Abstract: With the rapid development of communication and transportation technologies, the urban area is increasingly becoming an ever more dynamic, comprehensive, and complex system. Meanwhile, functional polycentricity as a distinctive feature has been characterizing urban areas around the world. However, the spatial structure of the urban area has yet to be fully comprehended from a dynamic perspective, and understanding the spatial organization of polycentric urban regions (PUR) is crucial for issues related to urban planning, traffic control, and urban risk management. The analysis of polycentricity strongly depends on the spatial scale. In order to identify functional polycentricity at the intra-urban scale, this paper presents a traffic flow-embedded and topic modeling-based methodology framework. This framework was evaluated on real-world datasets from the Wujiang district, Suzhou, China, which contains 151,419 records of taxi trajectory data and 86,036 records of points of interest (POI) data. This paper provides a novel approach to examining urban functional polycentricity via combining urban function distribution and spatial interactions. This proposed methodology can help urban authorities better understand urban dynamics in terms of function distribution and internal connectedness and facilitate urban development in terms of urban planning and traffic control.

Keywords: urban region; polycentricity; functional clusters; topic modeling

1. Introduction

In the industrial era of urban development, urban areas are predominantly considered to be monocentric systems [1]. As traffic networks and communication facilities are largely stretching out in urban areas, cities are becoming dilated and contain more sub-regions with diverse functions, which impose more challenges on decision-making in the urban context. More formally, it is widely acknowledged that urban spatial structure is becoming increasingly polycentric [2,3]. The concept of polycentric urban regions (PUR) has been studied intensively in terms of definition, identification, measurement, and longitudinal and transverse comparisons [4]. Technically, there are two dimensions to measure urban polycentricity: the morphological dimension and the functional dimension [5]. The former corresponds to the nodality of urban regions [6], while the latter relates to their centrality [7,8]. Functional polycentricity extends its morphological counterpart by incorporating the functional interactions among different regions [5], or network density [9].

Many empirical studies have analyzed and assessed the polycentricity of urban areas using such intrinsic attributes as the distribution of population and employment [10], and the movement

of residents or products between distinct regions of the city, specifically, the commuting patterns of people [11,12], or the flows of products and services [13]. Although the population and employment patterns to some extent represent the functional linkages and interactions over the entire urban area, there are some considerable deficiencies, such as lack of flexibility and delayed updating. Therefore, more dynamic, accurate, and comprehensive data sources are urgently needed to evaluate the spatial structure of metropolises.

Nowadays, with the increasing pervasion of positioning sensors and mobile devices, real-time digital footprints of urban residents are being collected by transportation authorities and social media platforms, which allow the polycentricity of urban area to be evaluated in a more quantitative and dynamic manner. Residents' digital footprints not only indicate their mobility patterns, but also reveal the morphological and functional relations between origins and destinations due to the strong temporal and spatial regularity of human mobility patterns [14]. For example, the regular movement of people on weekdays occurs between residential and working areas within a city; the destinations are more likely to be recreation and entertainment regions during weekends. Therefore, both sub-regional connections and functional relations can be induced from urban human mobility patterns. Furthermore, the function of a region can be implied by the points of interest (POI) information of entities within the region since POI data contains the category of function information. However, POI data may provide static information and fail to distinguish the quality of the corresponding entity. For example, a POI of shopping services in a relatively remote region may only provide groceries, while another shopping construct in central area may act as the recreational function. Taking advantage of the dynamic nature of human mobility patterns and functional attribute of POI, their combination can well depict the sub-regional connections and urban function distributions.

The large-scale data generated by the modern sensing and monitoring technologies in the urban context can be processed and analyzed by the application of statistical learning and machine learning techniques [15]. Since the function inference of urban regions is analogous to topic extraction in the context of text mining with the urban region and function corresponding to the document and topic, respectively, topic models can be employed to identify urban function distribution [16]. The viable topic modeling approaches include Latent Semantic Indexing (LSI) [17], Latent Dirichlet Allocation (LDA) [18,19], and Dirichlet Multinomial Regression (DMR) [20]. Drawing on these sophisticated and effective sets of methods, this study synthesizes a solid methodology framework, in which three steps are implemented consecutively, that is, segmenting basic urban units, discovering functional zones, and aggregating polycentric urban regions. In specific, the urban area is segmented to basic urban units using road networks. Then, DMR is used to extract the function of basic urban units based on traffic patterns and POI data, and TF-IDF, LSI, and LDA are also employed by comparison. The graph clustering is utilized to obtain urban functional polycentric regions. A case study is conducted using the human mobility data and POI data of the Wujiang district, Suzhou, China based on the proposed methodological framework. As it has been validated in the literature that taxi trajectory data is effective in analyzing urban spatial structure and function distribution [21–23], one-week taxi trajectory data of Wujiang between September 4 and 10, 2017, is employed, which includes 151,419 records, and the POI data used in the study contains 86,036 records.

The major contributions of this work to academics and practitioners are of three aspects. Primarily, this paper touches and bridges a wide range of disciplines and research fields, such as urban spatial structure, graph theory, machine learning, and big data, which suggests the potential for applying advanced techniques to urban system research. The proposed methodology provides a practical and effective approach, which takes advantage of state-of-the-art topic models, and applies the traffic data and POI data into DMR as well as other topic models for comparison. Secondly, the phenomenon of urban polycentricity is revisited from a novel perspective, providing a non-territorial, dynamic, and generative approach to understanding polycentric urban structure. Last but not least, the urban functional structure of Wujiang is demonstrated with the proposed methodology, and the DMR model

shows more desirable results compared to other topic models. The novel approach to evaluating polycentric urban structure may facilitate urban planning and management.

The remainder of the paper is as follows. The second section presents the related work in terms of the identification of urban function and measurement of urban polycentricity. The third section describes the datasets and underlying topic modeling approach. In the following part, the methodology framework is elaborated upon in-depth in a step-by-step fashion. Further, based on the taxi trajectory data and POI data of Wujiang, the methodology is evaluated and assessed. Specifically, the results of urban functional clusters based on distinct topic models are checked and compared, with the final section presenting a detailed discussion of the results and prospects for future research.

2. Related Work

2.1. Discovering Urban Function

Discovering urban functioning zones typically involves the use of remote sensing data generated by GIS [24], or the data obtained through extensive field interviews and questionnaires [25]. These approaches are often subject to either delayed updating or considerable consumption of time and financial resources, which impose significant constraints on their usefulness for urban authorities and researchers. Since human mobility in the urban area is closely associated with urban land use, or urban function, an increasing number of studies are adopting data generated by location-based services (LBS) techniques to evaluate human activities and functional distribution across the urban system.

Of these data sources, GPS data, cellular network data, and data generated by public transit including the metro and taxi are widely used in the current literature. These data sources feature large volumes, high accuracy, and high velocity, so they can be used to depict the human activities at the population level. For instance, Ratti and Frenchman [26] use location data obtained by mobile phones to investigate the intensity of human activities and city dynamics in their case-study of Milan; Roth et al. [27] utilize individual movement patterns based on Oyster card usage in subways to reveal the structure of London; and Guo et al. [21] extract mobility patterns from the origin-destination (OD) pairs of taxi trajectories.

With respect to the relation between human mobility patterns and urban functions, based on the GPS data of the taxis in Hangzhou, Qi et al. [22] demonstrate that the get-on/off behavior of taxi passengers depicts urban social dynamics, and further reveals the social function of the urban region. Pei et al. [28] take advantage of mobile phone data, which implies the activities of urban residents, to derive land functions. Zhong et al. [29] use the traffic data from surveys and smart card systems to extract the functions of the buildings within the urban area. With the proposed two-step method, the building functions can be inferred with a relatively high accuracy validated by substitute data sources.

Meanwhile, POI data is widely adopted by urban researchers due to the concrete functional information it contains. The integration of human mobility patterns and POI data can be more effective and comprehensive in terms of identifying urban functionality. For example, Zhang et al. [30] employ POI data as well as the public bicycle-sharing data to identify urban functions. Yuan et al. [23] use both POIs and taxi OD pairs as a proxy for human mobility patterns to probe the functional distribution over Beijing, and the results indicate that the combination of two data sources delivers superior performance relative to other situations involving only a single data source.

2.2. Identifying Urban Polycentricity

Identifying clusters or centers in the polycentric urban system requires data sources related to spatial interactions among the components in the system. Although such interactions can take various forms, the underlying mechanism characterizing these interactions is the labor market, which results in interrelated traffic patterns [3]. Admittedly, the polycentric urban region as a prevalent urban structure

has a long history, but the systematic illustration and formal analysis of the underlying interaction within polycentric urban region became a research focus during last two decades [31]. In urban and regional studies, urban polycentricity is a versatile concept without a normative definition [5,32]. In the vast amount of empirical work aimed at providing conceptual clarification and economic, social, and policy implications [4,33,34], researchers have come up with diverse measures to gauge urban polycentricity, such as network density [11], the gravity model [35], and rank-size distribution [5], based on commuting patterns [11,36] and business connections [37].

In the POLYNET project sponsored by the EU, Hall and Pain [38] utilize business network and information flow to investigate the functional polycentricity in the context of eight European metropolises. Notably, the concept of functional polycentricity was formulated and tested during this project. Taylor et al. [37] take advantage of firms' websites to pinpoint their business network and then construct their spatial structure. Green [9] explicitly articulates that functional polycentricity can be applied to multiple scales ranging from a building to a mega-city. Based on a hypothetical spatial place named Tuonela, he compares the polycentric properties of this hypothetical region using e-mail exchanges and commuting patterns among its settlements in line with predefined polycentric metrics.

More recently, in an ESPON report, three criteria including urban settlement structure, accessibility patterns, and territorial cooperation are utilized to identify polycentric development of Europe [39]. Liu et al. [40] use the intercity transportation network between major cities in China to measure the urban polycentricity. In the study of exploring the effect of polycentricity on urban economic performance, Kwon and Seo [41] employ traffic flow data to detect urban polycentricity in Korea. Therefore, the traffic patterns in the urban regions are effective in identifying polycentric urban regions, especially with the advantages of high data availability and prompt updating.

In summary, the two streams of literature, that is, urban function discovery and urban polycentricity analysis, both evaluate urban systems based on the intra or inter urban connections. However, they focus on functional and structural aspects, respectively. Combining these two related but conceptually distinct aspects will help our understanding of urban polycentricity in terms of function distribution and thus facilitate urban development. Although a considerable number of studies have investigated the two urban research fields, further studies are needed to integrate the two related aspects of urban research.

3. Data and Methods

3.1. Data

Two sources of data, taxi GPS data and POI data, are used in this study. As is shown in the related work, taxi trajectory data during certain time spans are effective to represent the human mobility in urban areas. Moreover, POI provides the contextual information concerning the whereabouts of the people being studied, thus allowing the contextual and functional awareness of the analysis of the urban region. As a regular public transportation mode in the urban area, the taxi possesses the advantages of high flexibility and efficiency, which enables citizens to reach the expected destination from their original location. Thus, it not only can satisfy the diverse travelling demands of residents, but also accurately and sufficiently represent the complex mobility patterns in the urban context. Moreover, the locational data is automatically generated by GPS devices, which ensures its accuracy and immediacy.

Due to the great demand for location-based-service (LBS), digital map technologies and applications are becoming increasingly widespread, which contributes enormously to the growth in the amount of POI data. POI provides useful information relating to categories of entities within a particular geographical area. When combined with POI information, such concrete social meanings of each trip as commuting, shopping, and entertainment, can be revealed. Therefore, the inclusion of POI helps construct mobility patterns in urban areas, which suggests its potential for extensive exploitation in

the research on urban spatial structure. There are 20 functional categories for POI data according to the standardized classification, as is shown in Table 1.

Table 1. The POI code and corresponding category.

Code	Category	Code	Category
1	Automobile service	11	Scenic spot
2	Automobile sales	12	Business and residential area
3	Automobile maintenance	13	Government agency and social organization
4	Motorcycle service	14	Science, educational and cultural service
5	Catering service	15	Transport facility
6	Shopping service	16	Finance and insurance service
7	Life service	17	Corporation
8	Sport and leisure service	18	Road affiliated facility
9	Health care service	19	Address information
10	Accommodation service	20	Public facility

3.2. Topic Modeling

Generally, there are two approaches to represent the content of a document: term weighting and indexing. The former aims to assign a weight to each term, a function often undertaken by TF-IDF [42], while the latter is used to assign indexing terms to each document. In particular, topic modeling—such as LSI [43], LDA [44], and DMR [20]—is extensively adopted to index documents. Unlike TF-IDF, which is closely related to the occurrence of terms in the corpus [45], topic models go beyond the lexical level of counting the frequency of terms and are able to discover latent thematic topics embedded in the text corpus. Specifically, LSI applied singular value decomposition (SVD) to the term-document matrix which is built based on the frequency of words in each document, and the latent topics of each document can be obtained according to the reduced matrix. LDA is essentially a three-level hierarchical Bayesian model. When modeling a text corpus, each document can be represented as a random mixture of underlying topics, and each topic can be represented as a distribution over a finite set of words.

Text documents are usually accompanied by ancillary information, such as annotated bibliographies. This information, formally defined as metadata, is in high demand to refine underlying topics and discover document patterns and configuration. Extensive research has therefore been conducted based on LDA to include metadata. This includes, for example, correspondence latent Dirichlet allocation (CorrLDA) [46], the Topics over Time (TOT) model [47], supervised latent Dirichlet allocation (sLDA) [48], and Dirichlet multinomial regression (DMR) [20]. DMR is the most capable and efficient among these LDA-based topic models, as it allows the inclusion of many kinds of metadata, including continuous and categorical data, and maintains relatively easy inference and flexibility. Therefore, we employ the DMR to extract the function of the urban region given traffic patterns and POI. Meanwhile, TF-IDF, LSI, and LDA are also employed for comparison.

3.3. Methodological Framework

As shown in Figure 1, the proposed methodology framework consists of three steps. In the first step, the whole city area is segmented into contiguous patches (i.e., basic units) using appropriate levels of traffic networks. Secondly, POI data and traffic patterns of each basic unit are fed into the DMR model, which allows the combination of both data sources. Related models such as TF-IDF, LSI, and LDA are also utilized by comparison. Then, the urban units with certain functional meaning are clustered with K-means clustering, resulting in functional zones with similar functions and roles in the urban system. Lastly, given the functional zones created in the third step, graph clustering is applied to the traffic patterns of these zones to generate urban regions, encompassing a subset of distinct functions, where intra-region travelling trips should be as large as possible while inter-regional trips should be as small as possible. Each of the regions can be considered as a cluster, featuring a set

of closely complementary functions and strongly interconnected traffic relations. The distinguishing features of this framework are the embeddedness of traffic flows and integration of topic models, in which the utilize traffic flows are utilized twice, and the performance of prevailing topic models is compared.

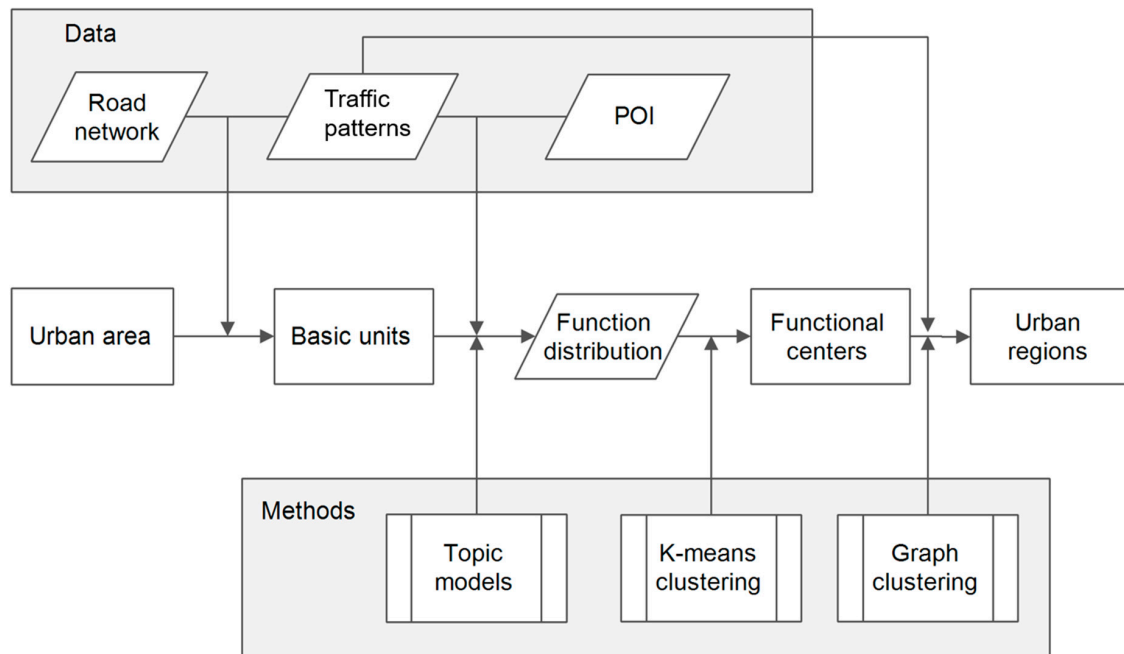


Figure 1. The methodological framework for discovering urban functional regions.

3.3.1. Division of Basic Urban Units

In urban studies, urban areas are often divided into a number of sub-regions to deal with practical problems related to traffic management and urban planning. The grid-based analysis is commonly used for its simplified implementation and ease of comprehension [49]. This approach partitions the whole area into adjoining equal-sized squares. Nevertheless, these squares only represent an abstract spatial territory and lack practical interpretation. In some cases, a large complex may be partitioned into several units. Another traditional approach exploits the spatio-temporal characteristics of sub-regions according to administrative boundaries [50], which brings about considerable deficiencies in practice, as the administrative boundaries blur in the increasingly integrated and complex urban system. The road networks naturally separate the city into sub-areas. A certain area encircled by roads usually encompasses a set of buildings with geographical and functional relations, and road-separated regions are often the natural locality of POIs and the trip origins and destinations. Therefore, this study uses road networks to divide urban areas into obtain basic urban units.

3.3.2. Discovery of Functional Zones

To satisfy the input data requirements of topic models, both the taxi OD data and POI data are processed to create the requisite data format. The taxi OD dataset contains the origin point, destination point, and their corresponding time. For a particular basic unit u_i , two kinds of mobility patterns (arriving and leaving) can be derived. The arriving pattern corresponds to trips that originate from other units and terminate at u_i , while its leaving counterpart corresponds to trips that originate from u_i and terminate at any other units. The traffic flow patterns of each unit are formulated on an hourly basis, and considering the distinct nature of mobility patterns between weekdays and weekends, they are constructed by averaging the number of trips on weekdays and weekends separately. Therefore, the mobility patterns of u_i can be built as a matrix with rows denoting basic units, columns representing

time bins, and each cell indicating the average number of trips originating from or terminating at u_i during the corresponding time bin. Further, the leaving and arriving cuboids are yielded by concatenating the corresponding mobility matrix of each unit.

Each record in the POI dataset refers to a single entity, which contains such properties as name, position, and category. According to the 20 categories listed in Table 1, the POI within each basic unit can be formulated into a vector, representing the categorical distribution. For basic unit u_i , the POI vector can be expressed as $V_i = (p_1, p_2, p_3, \dots, p_i, \dots, p_{20})$, in which p_i is the percentage of POI of the i^{th} category in total POI, i.e., $p_i = \text{number of POI of } i^{\text{th}} \text{ category} / \text{total number of POI}$. Similarly, the POI vector can be concatenated into the POI matrix.

Similar to the logic of topic inference for documents, topic models can be applied to discovering the function distribution of a region. As is shown in Table 2, a basic region unit, its functions, and traffic flow patterns in discovering urban function are analogous to a document, its latent topics, and words, respectively, in the situation of text mining. Specifically, POI data serves as metadata in the DMR model, which is equivalent to the part played by the bibliography information (e.g., author, date, or institution) in the topic extraction situation. As is shown in Figure 2, the process of DMR is:

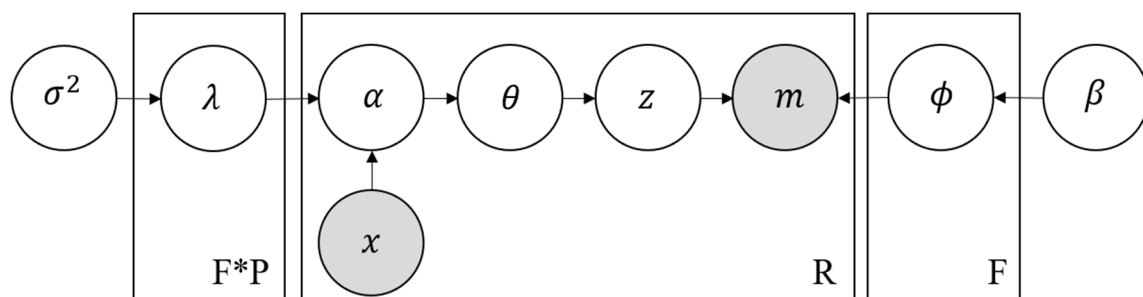


Figure 2. The illustrative scheme of DMR.

First, for each region function f ,

draw feature vector: $\lambda_f \sim N(0, \sigma^2)$

draw mobility pattern distribution: $\phi_f \sim \text{Dir}(\beta)$

Second, for each region r ,

for each function f , calculate the prior parameter: $a_{rf} = \exp(x_r^F \lambda_f)$

draw function distribution: $\theta_r \sim \text{Dir}(a)$

for each mobility pattern m_{rn} in region r ,

draw function assignment: $Z_{rn} \sim \text{Mult}(\theta_r)$

draw mobility pattern from the specific function: $m_{rn} \sim \text{Mult}(\beta_{Z_{rn}})$

where R represents regions, F denotes functions, P is the feature of POI, and β is the Dirichlet prior. Following the procedure of DMR, the function assignment for each basic unit can be discovered based on the traffic flow patterns and POI data.

For TF-IDF model, the number of POI pertaining to a certain category in a basic region unit resembles the term frequency in a document. Therefore, a POI feature vector can be built for each basic unit, and a POI matrix can be derived with each cell representing the POI percentage of a certain category. Similar to the case of discovering the high frequent terms of the regular document, this algorithm allocates a weight to each term and identifies those that are most notable and significant. These selected terms represent the elements underpinning the functions of each unit.

As a topic model, LSI enables the latent topics of each basic unit to be extracted from traffic flow patterns or POI data, which is the most distinguishing feature compared with TF-IDF. Along with the

standard process, the POI vector, or traffic flow patterns of each basic unit can be treated analogously as the words of a document. The SVD is then carried out on the constructed POI or traffic flow patterns, and the two approaches are designated as LSI_POI and LSI_MP. The dimensions or features representing the functions of basic unit are acquired by selecting a particular number of entries in the diagonal matrix.

Analogous to discovering the latent topics of a document, LDA can be employed to find the functions of the basic units in the urban area. In this scheme, the unit is equivalent to the document in a corpus, mobility patterns or POI to words, and functions to topics. The traffic flow patterns and POI data are loaded into the LDA model accordingly in LDA_MP and LDA_POI, and two types of function distribution can be obtained based on corresponding data sources.

As we have obtained six kinds of function representations using corresponding algorithms (i.e., TF-IDF, LSI_MP, LSI_POI, LDA_MP, LDA_POI, and DMR), K-means clustering is then applied to these function representations. As a consequence, the functional zones with certain functional topics are obtained.

Table 2. Models and corresponding data.

Model	Document	Representation	Word	Metadata
DMR	Basic unit	Latent index	Mobility pattern (MP)	POI
TF-IDF	Basic unit	Most weighted terms	POI	None
LSI	Basic unit	Latent index	POI or Mobility pattern (MP)	None
LDA	Basic unit	Latent index	POI or Mobility pattern (MP)	None

3.3.3. Aggregation of Urban Functional Clusters

The adjacent functional zones across the urban area are aggregated to form urban functional region, which is equivalent to the concept of polycentric urban region in the community of urban studies. PUR can be defined as a collection of settlements that are closely interconnected in terms of both functions and geographical locations [31,51]. In our case, the aim is to build up such PUR that the traffic volume within each centric area is as large as possible and that between centric areas as small as possible, suggesting strong connections among zones in the same centric area. As zones are characterized with functions, the PUR is inherently underpinned by functional attributes.

As a widely used graph analysis method, graph clustering provides an approach to grouping the vertices of a graph into clusters, also called communities [52], with the objective of having more edges within each cluster and relatively fewer between clusters [53]. Along with the logic, graph clustering is performed in the analysis to yield geographically contiguous regions, which possess various functions embodied in the functional zones. In this framework, each functional zone is further separated into several nonadjacent sub-zones, which are considered the nodes. The presence of a trip between sub-zones is regarded as the edge, and the number of trips is directly associated with edge weight. Therefore, using inter-zone travels, we can establish a weighted graph that is capable of depicting the morphological correlations among functional zones.

Graph clustering is implemented on the constructed traffic graph based on the derived zones in the previous step. Given six distinct categories of clustered urban functional zones, urban polycentricity can be evaluated based on the clustered results accordingly obtained from graph clustering.

4. Case Study and Results

4.1. Case Settings

The proposed methodology is applied to the case study on the Wujiang district, Suzhou, China. Wujiang used to be an independent city and incorporated into Suzhou in 2012. It is a well-developed region covering a relatively large geographical area of 1176.7 square kilometers. There are four major pillar industries for this region's economy: silk textiles, electronic information, optical cables, and

equipment manufacturing. In addition, the service sectors perform quite well, accounting for 46.4% of its economy. According to the 2017 national census data, its population is nearly 1 million, realizing a gross domestic product (GDP) of CNY 178.9 billion and 137,418 per capita, compared with the national level of CNY 59,660 Yuan per capita. Thus, Wujiang can be reasonably considered as a typical urban area with well-established industries and modern economies. Residents of Wujiang also enjoy a relatively high standard of living, with a disposable income of CNY 48,517 in 2017, compared with its national level of CNY 25,974. In terms of traffic models, residents have a wide range of choices and typically feel no extra burden in taking such premium public transit modes as the taxi and DiDi.

Two geo-related datasets regarding Wujiang are utilized, namely the taxi trajectory and POI data. It has long been authenticated in the literature and in practice that urban commuting patterns possess high weekly periodicity, and therefore one week's taxi data was collected from 4 to 10 September 2017, comprising 151,419 individual trip records, which accounts for 16.04% of the public traffic flow. It is worthy to note that the taxi positioning data is produced every 10 minutes by the onboard navigation system, which can be used to depict the taxi's traveling routine. This dataset contains abundant information concerning each trip, such as the real-time locations, OD points, boarding and alighting time, travel distance, time duration, and fees. POI data is acquired via API provided by Gaode, a map server provider in China, and POI data obtained for Wujiang contains 86,036 records.

In the first step, the urban area is divided using the road networks of Wujiang, which are accessed via API provided by Gaode. There are 117 roads in Wujiang at Level 2 and above, including freeways, ring roads, and secondary roads. However, not all of the segmented areas are significant entities in terms of traffic and POI. For example, several closely located roads may form roundabouts with trivial intra-traffic significance. Therefore, the district is segmented with minor manual adjustment, from which 187 basic units are derived, as is shown in Figure 3.

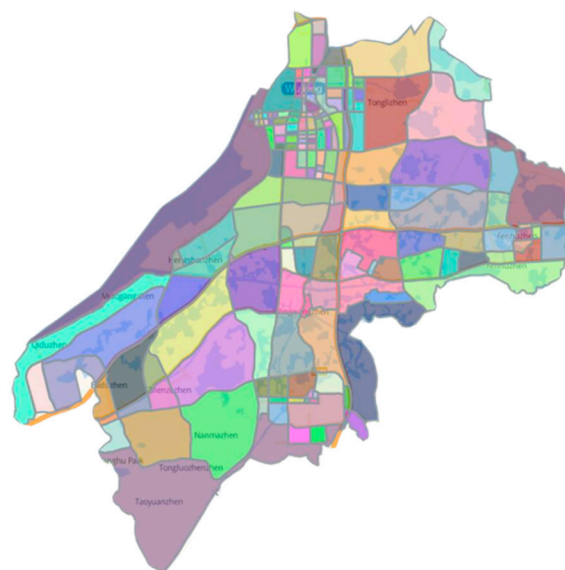


Figure 3. Basic units segregated by the road network.

4.2. Functional Zones Analysis

Based on the taxi trajectory and POI data of Wujiang, the function distributions can be extracted using DMR as well as other models. Subsequently, K-means clustering is carried out to obtain the functional zones. Since the number of clusters must be decided prior to the clustering procedure, an elaborate scheme for evaluating clustering performance is applied to determine the optimal number. This evaluation scheme encompasses three quantitative indices, CH, DB, and Silhouette. For the DMR model, the number of clusters is 7 according to the assessment results. Table 3 shows the index values and their corresponding cluster number K for all models.

Table 3. Clustering assessment by three indices.

Model	CH (K)	DB (K)	Silhouette (K)
DMR	5.917 (7)	2.628 (7)	0.04997 (6)
TF-IDF(POI)	47.08 (6)	0.80006 (7)	0.3592 (6)
LSI(MP)	3.867 (7)	0.3486 (8)	0.6188 (7)
LSI(POI)	49.09 (8)	0.7059 (7)	0.7426 (8)
LDA(MP)	16.83 (6)	1.483 (6)	0.2632 (8)
LDA(POI)	12.08 (7)	1.672 (6)	0.1278 (6)

As Figure 4 shows, the 187 basic units are clustered into 7, 6, 7, 8, 6, 6, and 7 functional zones for DMR, TF-IDF, LSI_MP, LSI_POI, LDA_MP, and LDA_POI, respectively. It is worth noting that the same color in different sub-figures may not necessarily represent the same functional zone. To evaluate the results of functional zone clustering, we consult local people and check the urban development plan of Wujiang. Taking TF-IDF as an example, as is shown in Figure 4b, region 1 is a cluster of scientific parks specializing in electronics and communications, which is evidently different from its left adjacent parks specializing in residences. Region 2 contains the Tongli Ancient Town (同里古镇), and region 3 mainly consists of Zhenze Ancient Town (震泽古镇). Although these two areas are quite apart, they are supposed to be sorted into the same functional zone. Region 4 and 5 are typical commercial area and industrial area, and thus should be distinguished from each other. Part or all of these problems also exist in LSI and LDA models. Technically, since TF-IDF, LSI, and LDA models utilize either POI data or mobility pattern data, they deliver distinct yet undesirable outcomes. In contrast, DMR model identifies the aforementioned functional zones properly, due to including information on both POI and mobility patterns. Therefore, we come to a conclusion that DMR presents the most satisfactory results compared with other models.

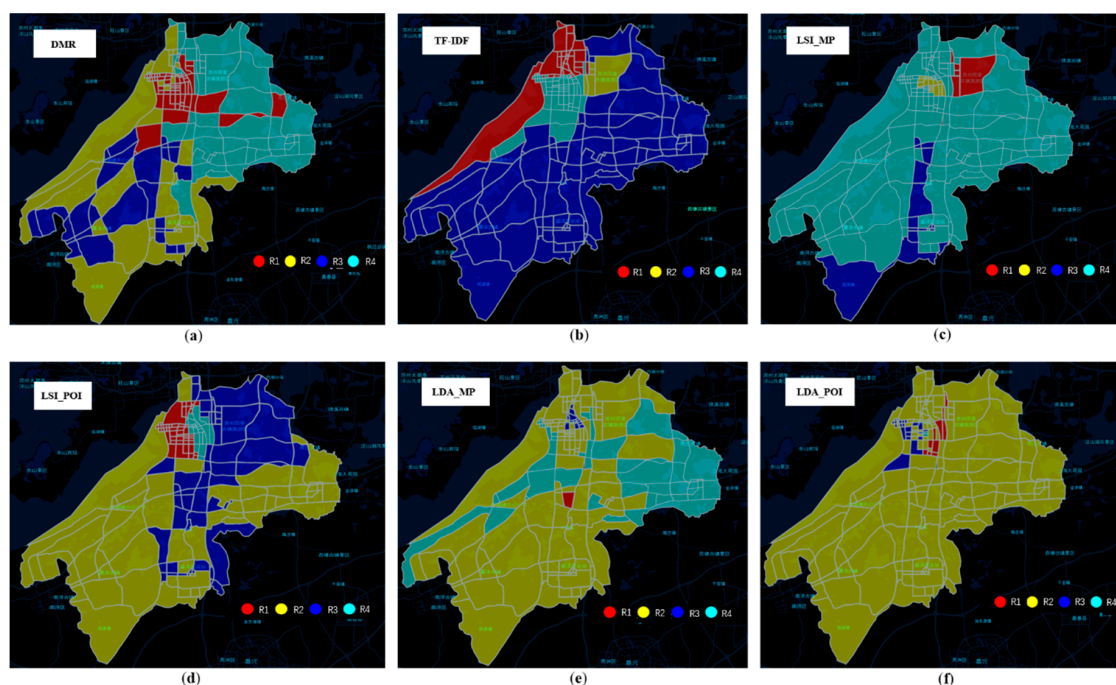


Figure 4. (a)–(f): Clusters of functional zones derived according to DMR, TF-IDF, LSI_MP, LSI_POI, LDA_MP, and LDA_POI, respectively.

Table 4 shows the POI frequency with respect to each category for aggregated functional regions using DMR. They are labeled with respective functional tags in consideration of the POI distribution and actual state of each cluster.

Table 4. The POI frequency of functional areas by DMR.

POI	A1	A2	A3	A4	A5	A6	A7
1	55	4	497	154	138	36	77
2	18	0	93	21	36	6	30
3	29	3	248	84	77	12	36
4	2	2	56	9	16	0	3
5	396	76	3068	1491	571	33	606
6	661	255	10,381	4181	2494	82	1349
7	404	77	4286	1909	783	64	829
8	119	7	666	260	145	4	121
9	60	19	780	266	123	7	97
10	37	13	586	172	109	6	117
11	15	9	210	26	46	2	11
12	78	19	845	243	157	18	119
13	107	21	1091	279	225	23	142
14	70	9	800	349	177	8	75
15	299	47	1966	707	490	87	250
16	77	5	557	220	73	11	71
17	736	179	8102	2092	1684	542	933
18	2	2	14	6	9	6	1
19	743	371	12,811	2430	3311	495	1145
20	62	15	461	192	128	8	70

A1: *Recreational and commercial area*. This area clearly contains many commercial and recreational facilities. Based on the mobility patterns, people tend to go to these places after work or on weekends.

A2: *Undeveloped area*. The area is either encircled by highways or expressways, or mainly covered with lakes and rivers.

A3: *Commercial and residential area*. This area comprises the major part of the whole city, the prominent feature of which is the mixture of residence and commercial and business services.

A4: *Industrial and residential area*. This area features industrial and residential buildings, and the traffic is relatively heavier on weekdays.

A5: *Public service area*. This area is characterized by public services such as schools, hospitals, and public institutions located in the area.

A6: *Traditional industrial area*. This area mainly comprises industrial plants, such as the clothing firms, elevator manufacturers, metal smelting companies, and other traditional industrial firms.

A7: *Mature residential area*. This area has such developed facilities for residents as hotels, restaurants, shopping stores, and banking services.

4.3. Urban Polycentricity Analysis

There are two criteria in the framework to evaluate urban polycentricity, namely traffic volume and function distribution. The difference between intra-region traffic volume and inter-region traffic volume indicates the degree of interconnection among these regions, while the function distribution of each region is manifested in the proximity of underlying functions, thus accounting for the concept of functional polycentricity.

Geographic elements are not considered in the procedure of functional clustering. Accordingly, the zones within the same functional category can be sparsely scattered. These areas of similar function may or may not be intensely related in terms of transportation. As for the formation of PUR, the disconnected zones of the same function are regarded as distinct entities, and then aggregated into separate regions using graph clustering. Due to the unique property of graph clustering, each aggregated region is physically connected, resulting in a geographically continuous area. Meanwhile, since functional zones are clustered based on traffic OD pairs, the clustered region has a relatively larger internal traffic volume and smaller external ones. Following this logic, therefore, the integral urban area can be

divided into several tight-knit regions with specific functions, enabling functional polycentric regions to be derived.

Based on the functional configuration generated by DMR, the urban area is clustered into four distinct regions denoted by four colors, as shown in Figure 5a. The comparison with other results derived by the corresponding algorithms can be seen in the following sub-figures. Table 5 shows the traffic volume between each pair of regions and its corresponding proportion, and Table 6 shows the distribution of functions in these regions. Except for region 3, the traffic volume within regions dominates its counterparts between different regions. According to Figure 5a, a considerable part of region 3 is surrounded by, or borders on, region 2, so inter-traffic volume is likely to be quite large. Region 1 is the most diverse in terms of functional distribution, with many functions including residences, industries, businesses, and public services. This region is widely acknowledged as the central area with various functions and a relatively large traffic volume. In contrast, region 2 is merely composed of function 3 with significant intra-traffic, which shows that internal industrial and residential activities are highly connected. Regions 3 and 4 vary in terms of their predefined functions, and their traffic volume distributions are quite different.

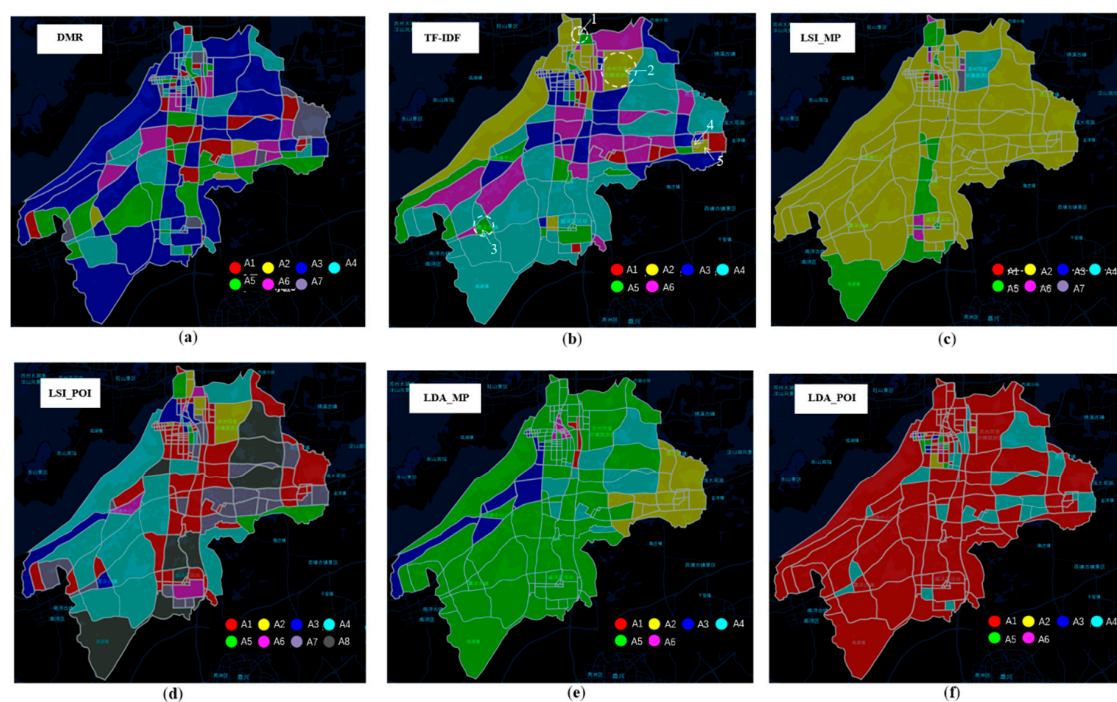


Figure 5. (a)–(f): Functional polycentric clusters for DMR, TF-IDF, LSI_MP, LSI_POI, LDA_MP, and LDA_POI, respectively.

Table 5. Traffic volume and its proportion between polycentric regions.

Region	Region 1		Region 2		Region 3		Region 4	
Region 1	15,799	0.666	3618	0.076	91	0.010	8617	0.230
Region 2	3372	0.142	33,955	0.716	5041	0.549	5288	0.141
Region 3	66	0.003	5054	0.107	3752	0.409	247	0.007
Region 4	4474	0.189	4796	0.101	290	0.032	23,364	0.623

Table 6. Distribution of function zones.

Region	Function 1	Function 2	Function 3	Function 4	Function 5	Function 6	Function 7
Region 1	8	0	2	7	4	2	3
Region 2	0	0	1	0	0	0	0
Region 3	4	1	0	6	5	0	3
Region 4	3	3	2	3	2	2	4

5. Discussion and Conclusions

This study develops a methodology framework for systematically partitioning the urban area into functional regions based on large volume of traffic data and associated POI information, which responds to the research issue concerning the polycentric urban region in the urban study landscape. As urban traffic and POI data is employed to detect the urban function and structure in the proposed methodology, the model needs to be validated in the context of mature urban areas, which are characterized by high traffic accessibility. The proposed methodology is illustrated with the case study examining the functional polycentricity of Wujiang, a highly urbanized area. In the case study, DMR delivers optimal results regarding the function distribution of each basic unit compared with other approaches. Based on function distributions, this area is separated into seven categories of functional zones, ranging from undeveloped periphery to commercial and residential center. These function zones are further clustered into four regions. The function distributions and traffic patterns of these regions verify the functional polycentricity property of Wujiang.

In the research literature concerning urban spatial structure, although researchers have developed various models to conceptualize and measure urban functional polycentricity [5,9,36], this concept is still encircled with certain fuzziness related to its analytical measurement [31]. This paper contributes to the stream of this literature in terms of integrating urban function distribution into urban polycentricity. Specifically, this methodology novelly applies topic modeling to the research on urban functional polycentricity using the traffic patterns and POI data, providing a more dynamic, holistic, and real-time urban partitioning approach than hitherto. Moreover, the case study verifies the applicability of the proposed methodology, and confirms the superiority of the DMR model which enables the combination of urban traffic patterns and POI data, corresponding to the urban dynamic and static features, respectively.

As for policy and managerial implication, it can be used to inform urban administration decision-making, thus facilitating urban planning, traffic management, urban disaster prevention, etc. It is acknowledged that urban networks and clusters are becoming more important for regional competitive advantage and economic growth. In this respect, this work provides an alternative solution for entangling urban polycentric networks, and helps policy makers formulate coordinated regional development policies that better reconcile regional cohesion and competition. In addition to benefiting urban management, the findings of this work are conducive to individuals and businesses. For urban residents, the urban functional zones and polycentric structure can provide general guidance in choosing residential places, recreational locales, and traffic paths. The study also has substantial implication for business activities. Based on the functional distribution in a PUR, business managers can be better informed in choosing store locations and primary places to launch marketing campaigns, which can increase their efficiency and effectiveness, and eventually lead to higher revenue.

This study also has several limitations, which should be addressed in future research. Firstly, the optimal K of clustering is not consistently derived by the evaluation methods, and therefore future studies are needed to improve the clustering procedure, and develop more elaborate clustering methods. Secondly, as the result evaluation regarding urban function identification mainly relies on real-life experience in the current work, more quantitative and comprehensive evaluation approaches are needed to assess the urban polycentric development. Thirdly, since the region of Wujiang studied in this work is a relatively small urban area, broader geographical area needs to be considered for evaluating the scalability of the proposed methodology. Lastly, the study used taxi OD data in topic

models, which may fail to fully capture the interconnection among urban regions. Thus, data generated by other traffic models such as bus, bicycle, as well as social media check-in data, can be employed for this end.

Author Contributions: Conceptualization, S.W. and Y.L.; methodology, S.W.; software, W.Z. (Wei Zhi); investigation, X.W.; resources, X.W.; data curation, W.Z. (Wei Zhi); writing—original draft preparation, S.W.; writing—review and editing, W.Z. (Weihua Zhou); visualization, W.Z. (Wei Zhi); funding acquisition, Y.L. and W.Z. (Weihua Zhou). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 71571160 and Information Technology: Industrial Transformation and Upgrading in 2016 (Made in China 2025), grant number 0714-EMTC02-5737/5.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hall, P. Modelling the Post-Industrial City. *Futures* **1997**, *29*, 311–322. [\[CrossRef\]](#)
- Kloosterman, R.; Musterd, S. The Polycentric Urban Region: Towards a Research Agenda. *Urban Stud.* **2001**, *38*, 623–633. [\[CrossRef\]](#)
- Parr, J. The polycentric urban region: A closer inspection. *Reg. Stud.* **2004**, *38*, 231–240. [\[CrossRef\]](#)
- Kloosterman, R.C.; Lambregts, B. Clustering of Economic Activities in Polycentric Urban Regions: The Case of the Randstad. *Urban Stud.* **2001**, *38*, 717–732. [\[CrossRef\]](#)
- Burger, M.; Meijers, E. Form Follows Function? Linking Morphological and Functional Polycentricity. *Urban Stud.* **2012**, *49*, 1127–1149. [\[CrossRef\]](#)
- Taylor, P.J.; Hoyler, M.; Verbruggen, R. External Urban Relational Process: Introducing Central Flow Theory to Complement Central Place Theory. *Urban Stud.* **2010**, *47*, 2803–2818. [\[CrossRef\]](#)
- Storbeck, J.E. The Spatial Structuring of Central Places. *Geogr. Anal.* **2010**, *20*, 93–110. [\[CrossRef\]](#)
- Preston, R.E. A Comparison of Five Measures of Central Place Importance and of Settlement Size. *Tijdschr. Voor Econ. Soc. Geogr.* **1975**, *66*, 178–187. [\[CrossRef\]](#)
- Green, N. Functional Polycentricity: A Formal Definition in Terms of Social Network Analysis. *Urban Stud.* **2007**, *44*, 2077–2103. [\[CrossRef\]](#)
- McDonald, J.F.; McMillen, D.P. Employment Subcenters and Land Values in a Polycentric Urban Area: The Case of Chicago. *Environ. Plan. A* **1990**, *22*, 1561–1574. [\[CrossRef\]](#)
- Burger, M.J.; de Goei, B.; van der Laan, L.; Huisman, F.J.M. Heterogeneous Development of Metropolitan Spatial Structure: Evidence from Commuting Patterns in English and Welsh City-Regions, 1981–2001. *Cities* **2011**, *28*, 160–170. [\[CrossRef\]](#)
- Parr, J.B. Spatial Definitions of the City: Four Perspectives. *Urban Stud.* **2007**, *44*, 381–392. [\[CrossRef\]](#)
- Ogden, K.W. The Distribution of Truck Trips and Commodity Flow in Urban Areas: A Gravity Model Analysis. *Transp. Res.* **1978**, *12*, 131–137. [\[CrossRef\]](#)
- Gonzalez, M.C.; Hidalgo, A.R.C.; Barabasi, A. Understanding Individual Human Mobility Patterns. *Nature* **2008**, *453*, 779–782. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zheng, Y.; Capra, L.; Wolfson, O.; Yang, H. Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans. Intell. Syst. Technol.* **2014**, *5*, 38. [\[CrossRef\]](#)
- Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Trans. Knowl. Data Eng.* **2014**, *27*, 712–725. [\[CrossRef\]](#)
- Kurashima, T.; Iwata, T.; Hoshide, T.; Takaya, N.; Fujimura, K. Geo Topic Model: Joint Modeling of User's Activity Area and Interests for Location Recommendation. In Proceedings of the WSDM 2013 - Proceedings of the 6th ACM International Conference on Web Search and Data Mining, Rome, Italy, 4–8 February 2013; pp. 375–384. [\[CrossRef\]](#)
- Coffey, C.; Pozdnoukhov, A. Temporal Decomposition and Semantic Enrichment of Mobility Flows. In Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, Orlando, FL, USA, 5 November 2013; pp. 34–43. [\[CrossRef\]](#)
- Kling, F.; Kildare, C.; Pozdnoukhov, A. When a City Tells a Story: Urban Topic Analysis. In Proceedings of the GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 7–9 November 2012; pp. 482–485. [\[CrossRef\]](#)

20. Mimno, D.M.; Mccallum, A. Topic Models Conditioned on Arbitrary Features with Dirichlet-Multinomial Regression. *Uncertain. Artif. Intell.* **2008**, 411–418.
21. Guo, D.; Andris, C. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Trans. GIS* **2012**, *16*, 411–429. [[CrossRef](#)]
22. Qi, G.; Li, X.; Li, S.; Pan, G.; Wang, Z.; Description, A.D. Measuring Social Functions of City Regions from Large-Scale Taxi Behaviors. In Proceedings of the 2011 International Conference on Pervasive Computing and Communications Work, Seattle, WA, USA, 21–25 March 2011; pp. 384–388. [[CrossRef](#)]
23. 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, Beijing, China, 12–16 August 2012; pp. 186–194. [[CrossRef](#)]
24. Xiao, J.; Shen, Y.; Ge, J.; Tateishi, R.; Tang, C. Evaluating Urban Expansion and Land Use Change in Shijiazhuang, China, by Using GIS and Remote Sensing. *Landsc. Urban Plan.* **2006**, *75*, 69–80. [[CrossRef](#)]
25. Evier, E.L.S.; Muller, D.; Zeller, M. Land Use Dynamics in the Central Highlands of Vietnam: A Spatial Model Combining Village Survey Data with Satellite Imagery Interpretation. *Agric. Econ.* **2002**, *27*, 333–354.
26. Ratti, C.; Frenchman, D. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environ. Plan. B Plan. Des.* **2006**, *33*, 727–749. [[CrossRef](#)]
27. Roth, C.; Kang, S.M.; Batty, M.; Barthélemy, M. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLoS ONE* **2011**, *6*, 2–9. [[CrossRef](#)]
28. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.; Li, T. A New Insight into Land Use Classification Based on Aggregated Mobile Phone Data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [[CrossRef](#)]
29. Zhong, C.; Arisona, S.M.; Huang, X.; Batty, M.; Schmitt, G. Detecting the Dynamics of Urban Structure through Spatial Network Analysis. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2178–2199. [[CrossRef](#)]
30. Zhang, X.; Li, W.; Zhang, F.; Liu, R.; Du, Z. Identifying Urban Functional Zones Using Public Bicycle Rental Records and Point-of-Interest Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 459. [[CrossRef](#)]
31. Vasanen, A. Functional Polycentricity: Examining Metropolitan Spatial Structure through the Connectivity of Urban Sub-Centres. *Urban Stud.* **2012**, *49*, 3627–3644. [[CrossRef](#)]
32. Markusen, A. Fuzzy Concepts, Scanty Evidence, Policy Distance: The Case for Rigour and Policy Relevance in Critical Regional Studies. *Reg. Stud.* **2003**, *37*, 701–717. [[CrossRef](#)]
33. Meijers, E. Polycentric Urban Regions and the Quest for Synergy: Is a Network of Cities More than the Sum of the Parts? *Urban Stud.* **2005**, *42*, 765–781. [[CrossRef](#)]
34. Meijers, E.J.; Burger, M.J. Spatial Structure and Productivity in US Metropolitan Areas. *Environ. Plan. A* **2010**, *42*, 1383–1403. [[CrossRef](#)]
35. Van Oort, F.; Burger, M.; Raspe, O. On the Economic Foundation of the Urban Network Paradigm: Spatial Integration, Functional Integration and Economic Complementarities within the Dutch Randstad. *Urban Stud.* **2010**, *47*, 725–748. [[CrossRef](#)]
36. De Goei, B.; Burger, M.J.; van Oort, F.G.; Kitson, M. Functional Polycentrism and Urban Network Development in the Greater South East, United Kingdom: Evidence from Commuting Patterns, 1981–2001. *Reg. Stud.* **2010**, *44*, 1149–1170. [[CrossRef](#)]
37. Taylor, P.; Evans, D.; Pain, K. Organization of the Polycentric Metropolis: Corporate Structures and Networks. In *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*; Routledge: London, UK, 2012; pp. 67–78. [[CrossRef](#)]
38. Hall, P.; Pain, K. From Metropolis to Polyopolis. In *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*; Routledge: London, UK, 2012; pp. 18–31. [[CrossRef](#)]
39. European Observation Network for Territorial Development and Cohesion (ESPON). Polycentric Territorial Structures and Territorial Cooperation. *ESPON Policy Br.* **2016**, *6*, 1–12.
40. Liu, X.; Derudder, B.; Wu, K. Measuring Polycentric Urban Development in China: An Intercity Transportation Network Perspective. *Reg. Stud.* **2016**, *50*, 1302–1315. [[CrossRef](#)]
41. Kwon, K.; Seo, M. Does the Polycentric Urban Region Contribute to Economic Performance? The Case of Korea. *Sustainability* **2018**, *10*, 4157. [[CrossRef](#)]
42. Ramos, J. Using Tf-Idf to Determine Word Relevance in Document Queries. In Proceedings of the First Instructional Conference on Machine Learning, Piscataway, NJ, USA, 3–8 December 2003; Volume 242, pp. 133–142.

43. Hofmann, T. Unsupervised Learning by Probabilistic Latent Semantic Analysis. *Mach. Learn.* **2001**, *42*, 177–196. [[CrossRef](#)]
44. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2012**, *3*, 993–1022. [[CrossRef](#)]
45. Zhang, L.; Luo, S.; Xia, H. An Investigation of Intra—Urban Mobility Pattern of Taxi Passengers. *arXiv* **2016**, arXiv:1612.08378.
46. Blei, D.M.; Jordan, M.I. Modeling Annotated Data. In Proceedings of the International Acm Sigir Conference on Research and Development in Information Retrieval, Toronto, ON, Canada, 28 July–1 August 2003; pp. 127–134. [[CrossRef](#)]
47. Wang, X.; Mccallum, A. Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends. In Proceedings of the Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; pp. 424–433. [[CrossRef](#)]
48. Blei, D.M.; Mcauliffe, J. Supervised Topic Models. In Proceedings of the 20th International Conference on Neural Information Processing, Kitakyushu, Japan, 13–16 November 2007; pp. 121–128.
49. Powell, J.W.; Huang, Y.; Bastani, F.; Ji, M. Towards Reducing Taxicab Cruising Time Using Spatio-Temporal Profitability Maps. In Proceedings of the Symposium on Large Spatial Databases, Minneapolis, MN, USA, 24–16 August 2011; pp. 242–260. [[CrossRef](#)]
50. Jiang, S.; Alves, A.; Rodrigues, F.; Ferreira, J.; Pereira, F.C. Computers, Environment and Urban Systems Mining Point-of-Interest Data from Social Networks for Urban Land Use Classification and Disaggregation. *Comput. Environ. Urban Syst.* **2015**, *53*, 36–46. [[CrossRef](#)]
51. Parr, J.B. Perspectives on the City-Region. *Reg. Stud.* **2005**, *39*, 555–566. [[CrossRef](#)]
52. Newman, M.E.J. Fast Algorithm for Detecting Community Structure in Networks. *Phys. Rev. E* **2004**, *69*, 66133. [[CrossRef](#)]
53. Fortunato, S. Community Detection in Graphs. *Phys. Rep.* **2010**, *486*, 75–174. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).