Front. Eng. Manag. 2024, 11(2): 231–246
https://doi.org/10.1007/s42524-024-0296-2

RESEARCH ARTICLE

Longzhu XIAO, Wangtu XU

Urban spatial cluster structure in metro travel networks: An explorative study of Wuhan using big and open data

Abstract Rail transit plays a crucial role in improving urban sustainability and livability. In many Chinese cities, the planning of rail transit routes and stations is focused on facilitating new developments rather than revitalizing existing built-up areas. This approach reflects the local governments' expectations of substantial growth to reshape the urban structure. However, existing research on transit-oriented development (TOD) rarely explores the spatial interactions between individual transit stations and investigates how they can be integrated to achieve synergistic effects and balanced development. This study proposes that rail transit systems impact urban structure through two "forces": the provision of additional and reliable carrying capacity and the reduction of travel time between locations. Metro passenger flow is used as a proxy for these forces, and community detection techniques are employed to identify the actual and optimal spatial clusters in Wuhan, China. The results reveal that the planned subcenters align reasonably well with the optimal spatial clusters in terms of spatial configuration. However, the actual spatial clusters tend to have longer internal travel times compared to the optimal clusters. Further exploration suggests the need for equalizing land use density within planned spatial clusters served by the metro system. Additionally, promoting concentrated, differentiated, and mixed functional arrangements in metro station areas with low passenger flows within the planned clusters could be beneficial. This paper presents a new framework for investigating urban spatial clusters influenced by a metro system.

Keywords urban spatial clusters, metro travel flows, land use, metro smartcard data, Wuhan

Received Jan. 17, 2023; revised Nov. 6, 2023; accepted Dec. 21, 2023

Longzhu XIAO, Wangtu XU (✉)
Department of Urban Planning, Xiamen University, Xiamen 361005, China
E-mail: ato1981@xmu.edu.cn

1 Introduction

The planning, design, and expansion of rail transit systems have been implemented in numerous cities worldwide with the aim of reducing the negative effects of private car usage and enhancing urban sustainability and livability. The initial underground rail transit system established in London in the 1860s marked the beginning of this trend (Ibraeva et al., 2020). However, the motivations behind these efforts are context-specific. While density-oriented transit planning is prioritized by federal agencies in the United States due to its ability to ensure ridership and fare revenue, many Chinese municipal governments have adopted a development-oriented approach (Yang et al., 2016). In China, rail transit stations are typically located in undeveloped or less developed land parcels, with the expectation of promoting compact and mixed-use development in these areas. The aim is to stimulate substantial and synergistic development through the implementation of rail transit systems. Thus, rail transit systems and their associated sphere of influence play a pivotal role in urban spatial restructuring.

The seminal concept of transit-oriented development (TOD) by Calthorpe (1993) has inspired many municipal governments in the United States to create self-contained neighborhoods within pedestrian catchment areas (PCAs) surrounding transit stations. The belief is that most residents' daily needs can be met within these PCAs (Curtis and Olaru, 2010). PCAs, however, are not isolated. TOD should be promoted in closely connected PCAs, requiring a deep understanding of the spatial network structure based on PCAs and the travel flows between them. Unfortunately, researchers and practitioners focusing on single PCAs often neglect to investigate the spatial interactions between them and how they can be combined to achieve synergistic effects and balanced development.

There are two related streams of analysis on urban spatial structure: morphological and functional. The morphological approach characterizes the urban form as monocentric or polycentric based on the clustering of population, employment, or human activities (Kloosterman and Musterd, 2001). Building on Castells' (1996) concept of "spaces of flows," an increasing number of studies emphasize the importance of considering functional linkages and spatial interactions between urban areas (Vasanen, 2012; Xiao et al., 2021). This functional approach moves beyond morphology and reveals urban spatial structure through various flows, such as commuting trips, shopping trips, and telephone calls. Several studies have used passenger flows from different modes of transportation (e.g., rail transit, taxi, and shared bike) to explore existing urban spatial structure (Roth et al., 2011; Liu et al., 2015; Wang et al., 2020; Chen et al., 2022). However, these studies rarely explore scenarios that could assist in joint TOD planning and design. Flows are influenced by urban form, which can be measured using indicators related to density, mixture, and function (Stead and Marshall, 2001). Understanding the relationship between land use in PCAs and passenger flows between them can guide development in PCAs and shape urban form more effectively.

Treating metro passenger flow as a proxy variable, this study aims to present a novel approach for identifying spatial clusters that are partially influenced by a metro system, consisting of stations and PCAs. In the existing literature, sub-centers refer to locations where population and various activities are concentrated, while sub-regions refer to partitioned urban areas that reveal a polycentric urban structure. In contrast to sub-centers and sub-regions, this study defines spatial clusters as sets of metro station areas (MSAs), where an MSA comprises a metro station and its corresponding PCA. Compared to an MSA, a spatial cluster is more self-contained due to the functional complementarity among MSAs within a cluster. Consequently, a greater volume of passenger flows can be observed between MSAs within a cluster when compared to other areas. These flows largely reflect the spatial interaction and functional linkage between different MSAs within a spatial cluster. On the one hand, these flows are influenced by land use attributes (e.g., function, mixture, and density) of PCAs in the cluster; on the other hand, they gradually reshape the existing land use of the city or region in question due to phenomena such as agglomeration, spillover, and economies of scale.

Recent literature and relevant public policies have underscored the significance and implications of polycentricism across varying geographic scales. Research indicates a positive relationship between polycentricism and spatial integration (Vasanen, 2013), commuting behaviors (Lin et al., 2015; Sun et al., 2016a), economic productivity (Li and Liu, 2018), and air quality (Li and Zhou, 2019) at the intra-urban level. However, these studies on polycentricism often narrowly focus on the identification and exogenous effects of sub-centers, overlooking the internal land use or functional composition of sub-regions and spatial clusters, as well as the flows within these areas. Therefore, this study proposes that analyzing the composition and flows would yield a better understanding of urban spatial configuration. By efficiently serving these spatial clusters, rail transit systems can minimize residents' overall travel time and enhance their overall quality of life, ultimately achieving the "city for people."

In this study, we utilize the community detection approach to identify spatial clusters formed by the metro system in Wuhan, China, based on smartcard data. We consider two types of flows: observed and potential passenger flows, which can be used to identify actual and optimal spatial clusters, respectively. By comparing actual spatial clusters, optimal spatial clusters, and planned sub-centers, as well as exploring the effects of land use on spatial clusters, we propose policies for polycentric spatial development. This study makes two key contributions: (1) It expands on the exploration of metro flow networks using a geographical approach, revealing the spatial cluster structure in both reality and the ideal scenario. This is in contrast to traditional urban designs and plans, which often provide subjective knowledge of urban spatial structure for the future. (2) Based on metro passenger flows, the spatial clusters identified in this research, along with further exploration of their relationships with land use, can provide insights to urban policymakers and practitioners on TOD planning and design strategies, promoting synergistic effects and balanced development.

2 Relevant literature

2.1 Urban spatial structure: toward a "flow" era

A significant body of literature focuses on urban spatial structure, particularly on identifying sub-centers and measuring polycentricity. The identification of subcenters has primarily relied on urban morphological attributes, with various methods proposed and employed. One of the simplest approaches involves using cut-off values, such as total employment numbers and minimum employment density within contiguous sub-areas (Giuliano and Small, 1991). More advanced methods include spatial statistical techniques (e.g., geographically weighted regression, as seen in McMillen, 2001), spatial clustering analyses (e.g., local Moran's I, as seen in Vasanen, 2012), and kernel density analysis (Leslie, 2010). The functional arrangement of urban land resources generates demands for interaction and gives rise to various flows (Stead and Marshall, 2001). These functional linkages, represented by different flows, connect discrete urban resources into an integrated system and reflect the spatial interactions between locations within the city (Liu et al., 2015). Recently, there has been increased attention on revealing urban spatial structure from a functional perspective, through the examination of the associations between land use and flows. Scholars often use functional linkages to detect and analyze urban polycentricity. For example, Vasanen (2012) proposed a connectivity field method as a measure of functional polycentricity and applied it to commuting flows in Finland. Instead of considering interconnections between sub-centers, Roth et al. (2011) explored the interactions between the London subway and arranged sub-centers to reveal polycentric urban spatial structure and organization. In recent studies, network science methods have been used to identify urban sub-regions and sub-centers and investigate functional linkages between sub-centers and within sub-regions based on various flows, particularly travel flows (Zhong et al., 2014; Liu et al., 2015; Munoz-Mendez et al., 2018; Wang et al., 2020; Zhang et al., 2021; Chen et al., 2022).

2.2 Network science methods: useful tools for urban spatial structure

While network science has a long history, its application in urban spatial studies is relatively new, particularly in the context of rail transit networks (Ducruet and Beauguitte, 2014). Network science methods have been utilized to reveal both inter-city and intra-city structures. Among these methods, network centrality metrics are commonly used to identify sub-centers, while community detection techniques are often employed to identify sub-regions.

A place's centrality in the spatial network indicates the number of people and activities it can attract. Therefore, identifying centrality is crucial for understanding its relationship with other places in terms of land use. For example, Burger and Meijers (2012), from an inter-city spatial network perspective, applied network centrality to determine the importance of a city within a polycentric metropolitan region, which reflects its connection strength with other cities. Building on this work, Liu et al. (2016) extended the definition by using total centrality and internal centrality to establish morphological polycentricity and functional polycentricity, respectively.

The fundamental concept of community detection, also known as graph partitioning or graph clustering, is to identify sub-regions within a spatial network by dividing it into several sub-networks, or communities. These communities are characterized by stronger internal functional connections compared to external ones (Girvan and Newman, 2002). Effective community detection can provide insights into the configuration of land use among urban areas. For instance, if two geographically distant areas exhibit strong functional linkages, they are likely to be grouped into the same community.

With the advent of ubiquitous sensors and location-based services, community detection has been utilized in various studies to analyze sub-regional structures from an intra-city spatial network perspective with emerging geospatial movement data. The data includes smart card data (Zhong et al., 2014; Zhang et al., 2021), taxi trip data (Liu et al., 2015; Wang et al., 2020), bike sharing data (Munoz-Mendez et al., 2018; Chen et al., 2022), mobile phone positioning data (Tanahashi et al., 2012; Gao et al., 2013; Jin et al., 2021), location-based social networking data (Sun et al., 2016b), and among others.

For example, Zhong et al. (2014) and Zhang et al. (2021) studied the evolution of hubs, centers, and borders in Singapore and London, respectively. Their studies revealed a trend toward a polycentric and compact urban form. Zhong et al. (2014) also noted that the emergence of these hubs, centers, and borders aligns with the cities' master plan, while Zhang et al. (2021) investigated the effect of employment density and residential density on travel flow structures.

Tanahashi et al. (2012) applied community detection to identify sub-regions within New York City, based on phone records, and analyzed mobility patterns between these sub-regions. However, their study focused solely on flows between sub-regions and overlooked flows within them. Thus, they were unable to comprehensively understand the differences in internal spatial interaction patterns compared to inter-subregion patterns.

Liu et al. (2015), by categorizing taxi trips as either short or long-distance, discovered a two-level hierarchical polycentric structure in Shanghai. Additionally, they examined the internal and external characteristics of sub-regions using two network centrality metrics - betweenness centrality and closeness centrality. They also integrated land use information to elucidate the formation of urban sub-centers and sub-regions. In a similar vein, Wang et al. (2020) used similar data and methodologies to define the detected sub-regions as functional urban regions. They found that these regions did not align with administrative borders, leading them to conclude that policy interventions should focus on enhancing functional linkages between sub-centers in the future. However, the results failed to offer insights into the sub-regions' potential for promoting polycentricity.

These studies serve as examples of constructing spatial networks based on emerging geospatial movement data and utilizing community detection to identify spatial clusters. Furthermore, they shed light on the influence of land use on flow structures and functional linkages.

3 Methodology

3.1 Identifying spatial clusters among metro station areas

Rail transit systems have a significant effect on urban spatial structure through two "forces": providing additional and reliable carrying capacity and reducing travel time between different locations. These forces are reflected in the flow of passengers among MSAs. By treating the spatial network as a social network that comprised of nodes (vertices) and links (edges), we can consider MSAs as nodes and passenger flows between MSAs as weighted links. This allows us to construct a spatial network based on the configuration of the metro system and the flows between MSAs. Community detection techniques are then utilized to identify spatial clusters defined in the introduction, referred to as MSA clusters (MSACs). Therefore, an MSAC represents a cluster of MSAs that share stronger relationships with each other compared to others in a particular city or region.

Defining spatial network connections

The identification of MSACs through community detection raises an important question: how should we define the connections between MSAs? While observed passenger flows reflect the actual strength of spatial interactions, they may not necessarily capture the potential of metro systems or reasonable land uses. For instance, a small passenger flow between two MSAs with a short travel time might be considered an external connection across communities. In this case, the community detection algorithm would assign the two MSAs to different MSACs, even though they may have the potential to generate more travel demands and trips. Similarly, a large passenger flow with a long travel time may be viewed as an internal connection within a community. Essentially, while passenger flows provide a measure of connection in terms of actual spatial structure, there may be instances where unfavorable situations arise, such as large passenger flows with long travel times within MSACs or small passenger flows with short travel times between MSACs.

One of the primary advantages and explicit goals of transportation development is to reduce travel time. Thus, the inverse of travel time can serve as a proxy for the expected passenger flows that represent optimal spatial interaction patterns, maximizing benefits and approximating goals. By utilizing the inverse of travel time as a measurement for spatial network connections, we can establish an optimal spatial structure based on the metro system.

Creating spatial networks

A weighted spatial network, with MSAs as vertices and trip number between the MSAs as the weights of edges, is denoted by $G_1(V,E)$, where V is the number of vertices and E is the number of edges. Here, the weights are calculated by $W_{ij} = TN_{ij}$, where TN_{ij} denotes trip number between MSA i and MSA j.

Similarly, with MSAs as vertices and the inverse number of travel time between the MSAs as the weights of edges, the other type of weighted spatial network is denoted by $G_2(V,E)$. Here, the weights are calculated by $W_{ij} = -TT_{ij}$, where TT_{ij} denotes travel time between MSA i and MSA j.

Identifying spatial clusters

There is no universally accepted protocol for identifying community structures (Fortunato and Hric, 2016). To ensure scientific rigor, communities detected by different algorithms should be evaluated based on the knowledge specific to the research fields and objectives. In this study, the community detection algorithm known as Spinglass is employed to identify spatial clusters. Spinglass is an optimization algorithm that utilizes spin models and simulated annealing algorithm (Reichardt and Bornholdt, 2006). It solves the problem of community detection by identifying the ground state of an infinite range spin glass. To determine the ground state configuration and define the cluster structure of the spatial network, a quality function called Hamiltonian is utilized. The goal of the Hamiltonian is to minimize the energy of the spin glass and maximize the modularity of the spatial network system. The Hamiltonian is defined as follows:

$$H(\{s\}) = A_{ij} W p_{ij} \delta(s_i, s_j), \tag{1}$$

where $H(\{s\})$ represents the Hamiltonian of partition s, A_{ij} is the adjacent matrix of spatial network consisting of vertices i and j, as well as the edges between them, W is the weights of the edges, p_{ij} is the probability of the edges, s_i and s_j are the spin states (i.e., spatial cluster indices), $\delta(s_i, s_j)$ is a dummy variable that is equal to 1 if i and j are within the same spatial cluster, and it is equal to 0 otherwise. During the optimization process, an initial solution is randomly assigned and subsequently updated based on the quality function. In each iteration, candidate solutions are generated by moving vertices to other clusters or combining/dividing previous clusters.

The two types of spatial networks can be classified as positively and negatively weighted. The advantage of using the Spinglass algorithm over other algorithms is its ability to handle edges with negative weights (Traag and Bruggeman, 2009). Typically, a community detection procedure would identify communities with strong internal edges and weak external edges. However, in the case of networks with negative weights, this situation is reversed. In a negatively weighted network, the community structure is characterized by weaker negative edges within communities and stronger negative edges between communities. This research applies the Spinglass algorithm to detect two types of spatial clusters: actual MSACs and optimal MSACs. Actual MSACs refer to a set of MSAs that are connected to each other through larger passenger flows, while optimal MSACs refer to a set of MSAs that are connected to each other through shorter travel time. In other words, actual MSACs represent real urban spatial clusters with strong internal spatial interactions and weak external spatial interactions, while optimal MSACs represent an ideal situation where internal trips are generally more efficient than external trips.

3.2 Implementation of the empirical research

The overall procedures of the analysis are illustrated in Fig. 1. The first step involves constructing spatially

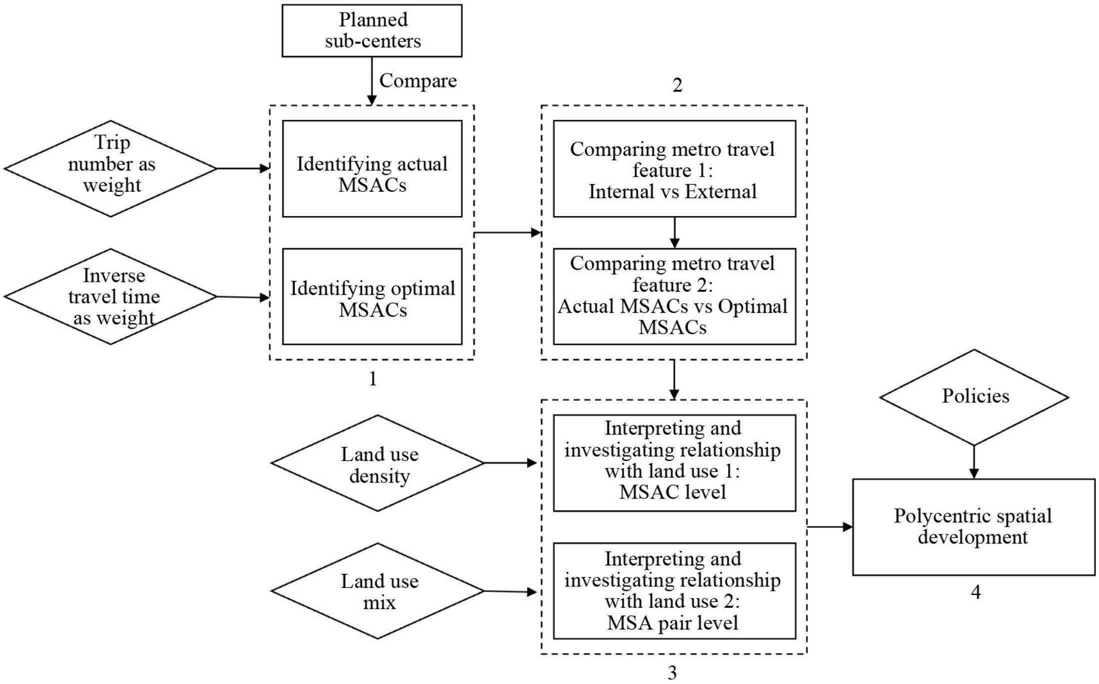


Fig. 1 Overall procedures of analyses.

embedded metro travel networks and identifying the actual MSACs and optimal MSACs using the community detection method. The identified spatial clusters will then be compared to the planned urban sub-centers established by the local government. Next, to evaluate the performance of the community detection method and gain a better understanding of the structures of actual and optimal spatial clusters, metro travel features such as trip number and travel time between stations will be compared, specifically focusing on internal versus external and actual MSACs versus optimal MSACs. Subsequently, we will explore the relationships between the identified spatial cluster structure and land use features, specifically land use density and land use mix. This analysis will be conducted at both the MSAC level and the internal MSA pair level. Finally, based on the findings, we will propose policy recommendations for local development.

4 Empirical study

4.1 The study site

Wuhan, located in east-central Hubei in central China, was chosen as the empirical study site for two reasons. First, Wuhan is a densely populated city with a growing number of rail transit lines. It is highly likely that these rail transit lines have significantly influenced the polycentric structure of the city. As of 2019, the population of Wuhan was approximately 8070000, with a population density of 2475 persons/km^2 within the Core Urban Development Zone (Wikipedia, 2019). In 2019, Wuhan had a metro system consisting of 9 lines and 189 stations, serving about 3.05 million trips per day. With a long-term goal of reaching a total length of 1100 km by 2049, ambitious expansion projects are currently underway. Secondly, Wuhan has long been recognized for its polycentricism due to the presence of rivers and lakes. The city is divided into three parts by the Yangtze River and Han River, namely Wuchang, Hankou, and Hanyang. Before the introduction of cross-river rail transit lines, this division resulted in fragmented landscapes and polycentricity (Liu and Wang, 2016). These geographic factors provide favorable conditions for promoting polycentricity and spatial cluster development in Wuhan. The concepts of polycentric urban spatial development were incorporated into the Wuhan Master Plan (2017–2035).

TOD is also emphasized in the Wuhan Master Plan (2017–2035). Like many cities in China, Wuhan is striving for development through the integration of rail transit and land use. Therefore, studies on MSACs in Wuhan can offer valuable insights for other cities. Additionally, these studies can address the research gap in the existing literature, which has focused less on MSCAs using emerging geospatial movement data and network science methods.

The study area includes MSAs (metro stations and their PCAs) in Wuhan in 2019. A PCA refers to a buffer area around a transit station that is easily accessible by foot. Typically, it is defined within a radius of 400–800 m from a transit station. Jun et al. (2015) and Chakour and Eluru (2016) compared different radii in their empirical studies and suggested a radius of 600 m as the most effective analysis scale for TOD. Therefore, this study adopts a 600-meter radius.

4.2 Data

Smartcard data

The smartcard data used in this study covers a typical week, specifically March 11–17, 2019. The original data set contains records of passengers tapping in and out of metro stations in Wuhan, China. As our focus is on passenger flows between stations, we extract the relevant trips from the original data set. In total, there are more than 16.7 million trips, with each trip including information such as the boarding station, boarding timestamp, alighting station, alighting timestamp, and duration.

Figure 2 displays the hourly trip numbers and average travel times across the metro system over a period of seven consecutive days. By comparing these figures, we can identify distinct travel patterns between weekdays and weekends. Consequently, our analyses are divided accordingly for weekdays and weekends. To facilitate the construction of the spatial networks, we aggregate the weekday and weekend trips into station-based matrices based on trip numbers and travel times. Please note that this study does not consider the directions of trips, but rather focuses on the strength of potential spatial interaction.

Points of interest (POI) data

A growing number of studies utilize open data, particularly points of interest (POI) data obtained through data mining techniques, to portray land use characteristics (Yue et al., 2017). Compared to traditional land use data, POI data offer higher precision and more comprehensive details. They describe land use characteristics through a broader range of facilities and destinations, rather than being limited to specific land use types and sizes.

For this study, we obtained POI data of Wuhan from Amap.com in 2019. The data set comprises two levels of classification. At the first level, there are 13 categories, and at the second level, more than 60 categories exist. We subjectively omit or condense certain information at the second level. For instance, we assume that agriculture facilities, moving companies, and laundry services have negligible relevance to metro travel behavior, and therefore exclude them from our analysis. Additionally, similar categories such as museums, art galleries, exhibition halls, science and technology museums, and libraries are merged into a single category. Following these procedures, we obtain 13 categories at the first level and 50-five categories at the second level, as depicted in Fig. 3. We adopt the second level of classification to achieve enhanced detail and accuracy in representing land use. Subsequently, we extract a subset of POIs located within the study area.

Based on these POIs, land use density and land use mix by PCA are calculated. The land use density for POIs of category i in an PCA j is measured by $Density_{ij} = N_i/Area_j$, where N_i denotes the number of POIs of category i, and $Area_j$ denotes the area of PCA j. Entropy is a widely used measurement of land use mix, which assumes that land use with equal percentage for each type creates the best mix. However, this assumption is theoretically problematic (Song et al., 2013). In this study, we employ an adapted entropy proposed by Song et al. (2013), which overcomes the problem by incorporating a reference geography. Based on POIs, it is calculated as follows: Suppose that R is a reference geography (it refers to a well-balanced area, here, we consider the city as a reference geography), for land use $i = 1, 2, \ldots, 55$, the percentage of type i within R

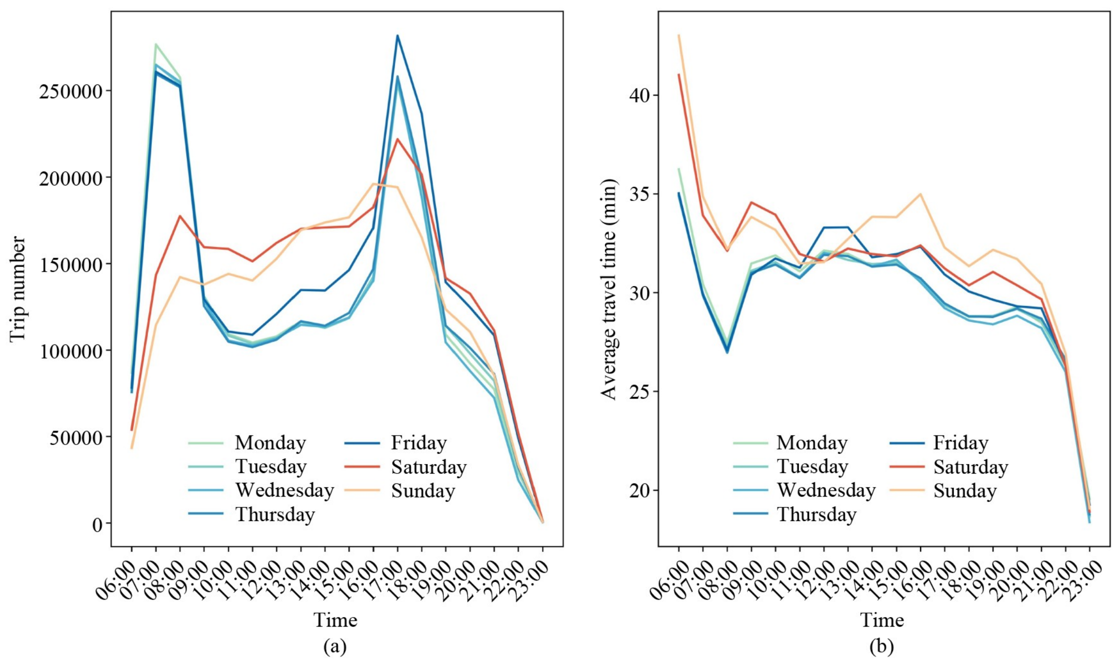


Fig. 2 Hourly trip number (a) and hourly average travel time (b) across the metro system within a typical week.

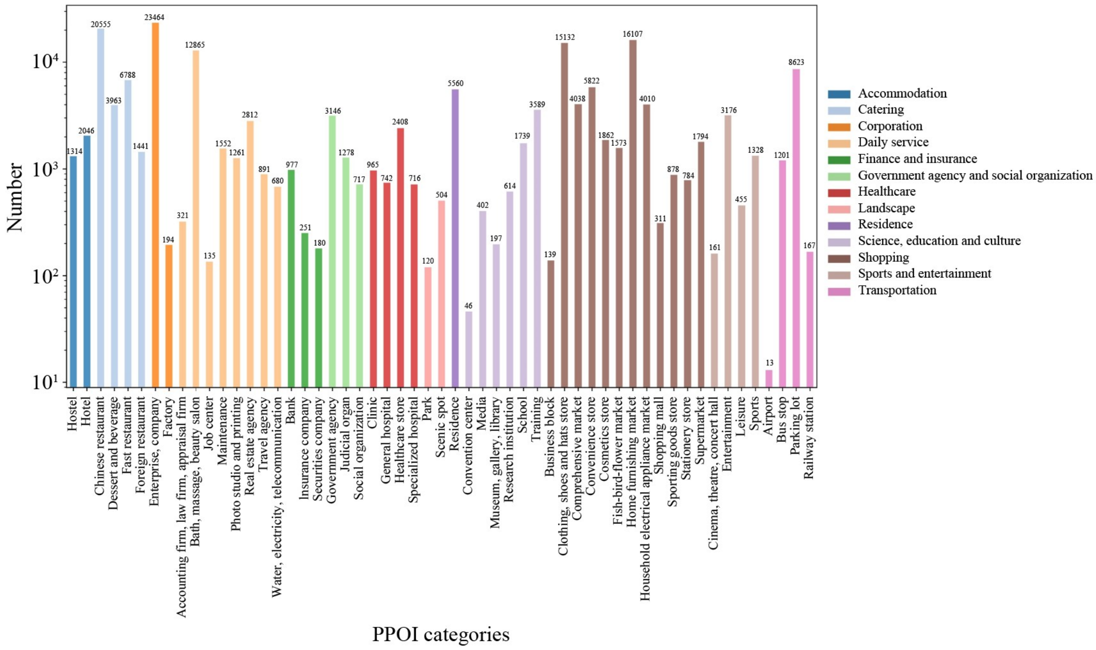


Fig. 3 A summary of POI data.

is r_i, $\sum r_i = 1$. For a PCA j, with the percentages of POIs of all categories $p_1, p_2, \ldots p_i$, $\sum p_i = 1$. Let $q_i = p_i / r_i$, recreate land use percentages for PCA j, $s_i = q_i / \sum q_i$, $\sum s_i = 1$. Finally, the adapted entropy is calculated by $Entropy_j = -\sum_i (s_i \ \ln s_i) / \ln 55$.

4.3 Analysis and discussion

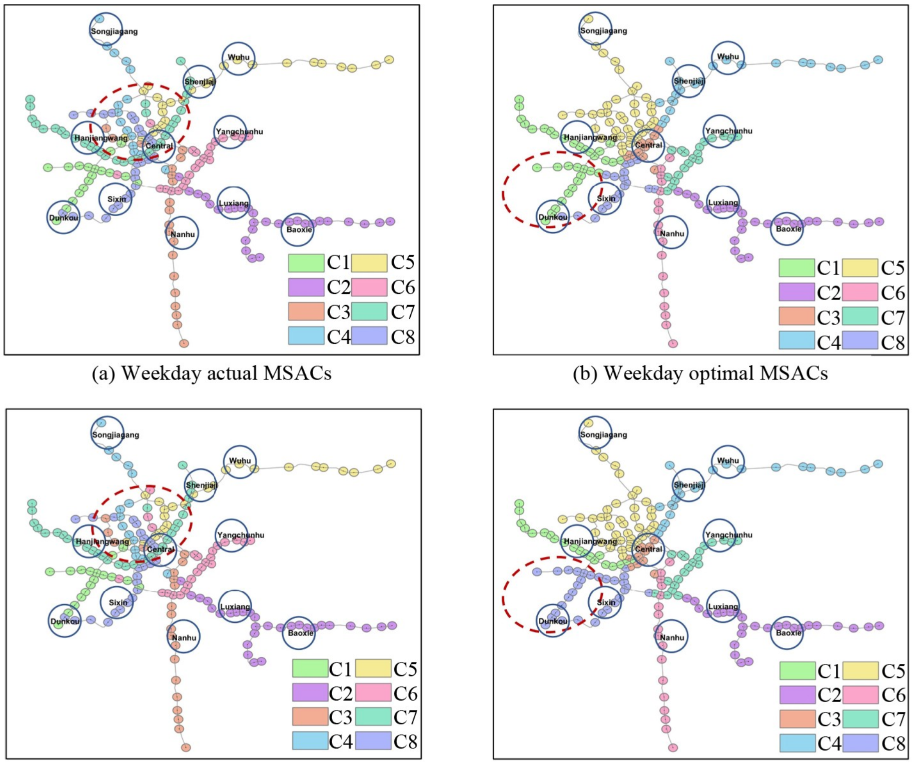
4.3.1 Actual and optimal spatial cluster structures

Identifying MSACs via community detection

The MSACs in Wuhan are investigated from two perspectives: the actual and the optimal. We create two types of spatial networks, i.e., $G_1(V = 189, E = 17112)$ and $G_2(V = 189, E = 17112)$ for weekdays, $G_1(V = 189, E = 16728)$ and $G_2(V = 189, E = 16728)$ for weekends, where the edge weights denote the actual and potential spatial interaction strengths, respectively. The community detection analysis is performed using the Spinglass algorithm, provided in the igraph package in R. This algorithm enables us to handle both positively weighted spatial networks and negatively weighted spatial networks. It is important to note that as a stochastic heuristic optimization algorithm, the operation of the algorithm may generate different results each time. To ensure the robustness of the results, we conduct pre-analyses one thousand times and set seeds for each analysis. Then, we determined the median spatial cluster number and performed formal analyses based on the corresponding seed. Four sets of MSACs, namely actual MSACs and optimal MSACs for weekdays and weekends, are identified and displayed in Fig. 4. Each MSA in the maps is color-coded based on its respective cluster.

As depicted in Figs. 4(a) and 4(c), there are slight disparities observed in the actual MSACs between weekdays and weekends, with the central area of the city being the main location for such differences. In contrast, noticeable discrepancies can be seen in the optimal MSACs between weekdays and weekends in the western part of the city, as illustrated in Figs. 4(b) and 4(d). The volume of metro travel flows on weekdays and weekends exhibit similarities with subtle variances in terms of both the number of trips and travel duration. Throughout the week, metro travel flows in the central area display a higher level of instability in terms of trip numbers compared to the peripheral area. This can primarily be attributed to the central area's more complex and multifunctional nature when compared to the peripheral area. Furthermore, there are significant changes in metro travel time to and from the western part of the city. This can be attributed to variations in congestion levels within the metro system and different metro departure frequencies between weekdays and weekends. On one hand, the consistency of metro travel flows between weekdays and weekends leads to similar optimal spatial clusters. These spatial clusters can facilitate TOD practices. On the other hand, the subtle distinctions give rise to partial differences in both actual and optimal spatial clusters, necessitating adaptive TOD policies that take into account both weekday and weekend scenarios.

Interestingly, the actual MSACs exhibit a spatial mix within the central area, while they remain spatially cohesive in the peripheral area. This reflects the intricacies of metro travel activities in the real world and aligns with the urban form and functional arrangement across different urban spaces. The central area is often characterized by



(a) Weekday actual MSACs (b) Weekday optimal MSACs

(c) Weekday actual MSACs (d) Weekday optimal MSACs

Fig. 4 The detected actual and optimal spatial clusters and the planned sub-centers.

dense buildings and mixed-use developments, whereas the peripheral area tends to have low-density and functionally homogeneous landscapes. As anticipated, the optimal MSACs exhibit greater spatial cohesion, with the MSAs therein being contiguous due to the metro travel time being roughly consistent with the corresponding Euclidean distance. Such cohesive spatial clusters can contribute to the promotion of urban polycentricity, which necessitates the rearrangement of land use.

The optimal spatial clusters are consistent with planned urban sub-centers

Typically, in a polycentric spatial structure, each sub-center would dominate a specific area. To assess whether the spatial cluster structure of a metro system contributes to polycentric spatial development, we compare the planned sub-centers with the identified spatial clusters. We examine whether each planned sub-center is located within a cluster (Fig. 4). According to the Wuhan Master Plan (2017–2035), there are 11 planned sub-centers. Eight sub-centers fall within one actual MSAC, while three exceptions (Central, Shenjiaji, and Zhuankou) span across multiple actual MSACs. In contrast, the optimal MSACs align more closely with the planned sub-centers. This can be attributed to the spatial cohesion of the optimal MSACs and the consideration of urban residents' travel time characteristics in the planning process. Therefore, optimizing the existing spatial cluster structures and realizing the optimal ones would support polycentric spatial development in Wuhan.

4.3.2 The travel features of actual and optimal spatial cluster structures

Comparing external and internal travel features

By analyzing passenger flows associated with actual and optimal spatial clusters identified through community detection, we can gain insights into the spatial cluster structures from the perspective of "flow." Figure 5 presents boxen plots comparing the distribution of trip numbers and travel times between internal and external edges for the four sets of MSACs. Figures 5(a) and 5(b) illustrate that the trip numbers of internal edges are generally greater than those of external edges for both actual and optimal spatial clusters on both weekdays and weekends. Notably, this trend is more pronounced for the actual MSACs. Conversely, as shown in Figs. 5(c) and 5(d), the distributions of travel times exhibit the opposite pattern. The travel times of internal edges are generally shorter than those of external edges. This observation is more prominent for the optimal MSACs. These findings are consistent with our expectations regarding actual and optimal spatial cluster structures, indicating the suitability of the adopted community detection method for this study.

Comparing travel features between actual and optimal spatial cluster structures

To further investigate the differences between the actual MSACs and optimal MSACs, we compare their total trip numbers and average travel times internally and

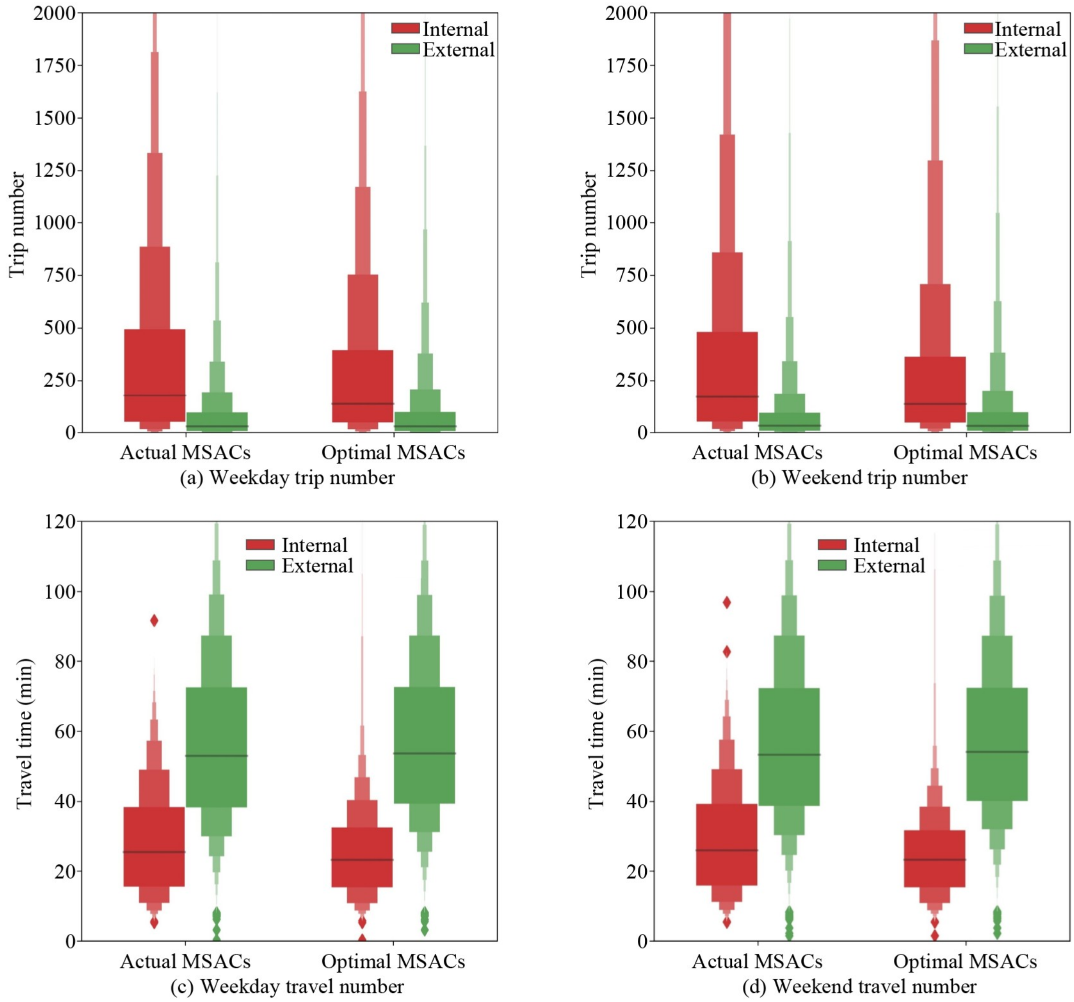


Fig. 5 Comparison of internal and external travel features.

externally. On one hand, the actual MSACs display higher internal total trip numbers and lower external total trip numbers on both weekdays and weekends, as depicted in Figs. 6(a) and 6(b). On the other hand, the optimal MSACs exhibit shorter internal average travel times and longer external average travel times on both weekdays and weekends, as illustrated in Figs. 6(d) and 6(e).

The ratios of internal total trip number to external total trip number and internal average travel time to external average travel time are utilized for further comparison in order to analyze the degree of concentration of trips in MSACs and the efficiency of travel within MSACs compared to outside. A higher ratio of the former signifies a higher level of trip “internalization” and a clearer spatial cluster structure, while a lower ratio of the latter indicates higher time-efficiency, a higher degree of potential trip “internalization,” and a clearer potential spatial cluster structure. Figures 6(c) and 6(f) demonstrate that actual MSACs have a higher ratio of internal total trip number to external total trip number, while optimal MSACs have a lower ratio of internal average travel time to external average travel time. These ratios suggest that the actual MSACs are not efficient enough in terms of saving travel time. To maximize the benefits of time saving provided by the metro system and achieve efficient spatial cluster development, current land use should be promoted to optimize people’s trip-making behaviors, resulting in MSACs characterized by larger internal total trip number and shorter internal average travel time.

4.3.3 The relationship between spatial cluster structure and land use

As previously argued, land use arrangements partially influence passenger flows between MSAs (Stead and Marshall, 2001). Examining the effect of land use on spatial clusters derived from metro trips can provide insights for guiding spatial development. Our exploration in Wuhan focuses on land use density and land use mix at two scales: the MSAC scale and the MSA pair scale.

The MSAC scale

The flows are closely related to land use, meaning that the actual spatial cluster structure reflects the actual land

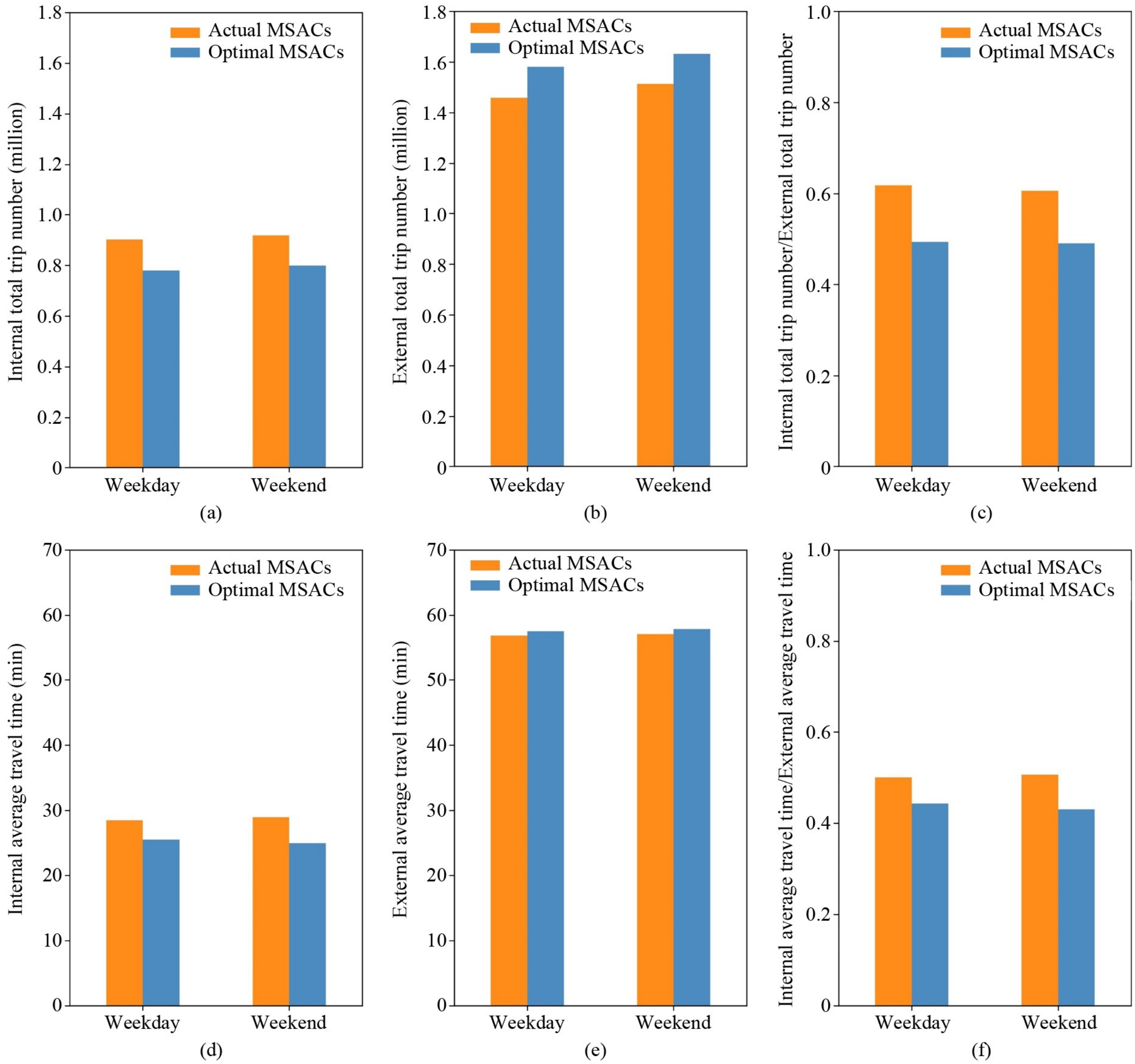


Fig. 6 Comparison of travel features between the actual MSACs and optimal MSACs.

use at the spatial cluster level. In contrast, other potential spatial cluster structures, such as the optimal one, may not reflect land use to the same extent. Therefore, differences in land use patterns between the actual and optimal spatial cluster structures may exist, offering insights into achieving the optimal spatial cluster structure. To test this assumption, we compare land use density and land use mix between actual MSACs and optimal MSACs.

Heatmaps in Fig. 7 display the land use density for the four sets of MSACs. The density data has been normalized to eliminate the influence of POI categories. A more equal pattern is evident among the actual MSACs in comparison. C3, an extremely dominant cluster, is found in both weekdays and weekends for the optimal MSACs. The difference between the actual and optimal MSACs for each land use category can be quantified by $D_i = \sigma_i^{\text{actualMSAC}} - \sigma_i^{\text{optimalMSAC}}$, where $\sigma_i^{\text{actualMSAC}}$ denotes the standard deviation of normalized density for land use i in the actual MSACs, $\sigma_i^{\text{optimalMSAC}}$ denotes the standard deviation of normalized density for land use i in the optimal MSACs. As shown in Figs. 7(c) and 7(f), the difference values are predominantly negative. These results substantiate the observation that land use density is more equal among the actual MSACs than among the optimal MSACs. Despite the spatially unequal and discrete distribution of land resources and functional arrangements, metro travel flows connect these urban spaces into equal and integrated spatial clusters. Based on the relationship between land use and metro travel behaviors, it is reasonable to conclude that arranging land and functional development equally among spatial clusters is a prerequisite for generating a polycentric urban structure.

Entropy values for individual MSAs and the four sets of MSACs are presented in Fig. 8. Although the entropy values of the MSACs are significantly higher (generally

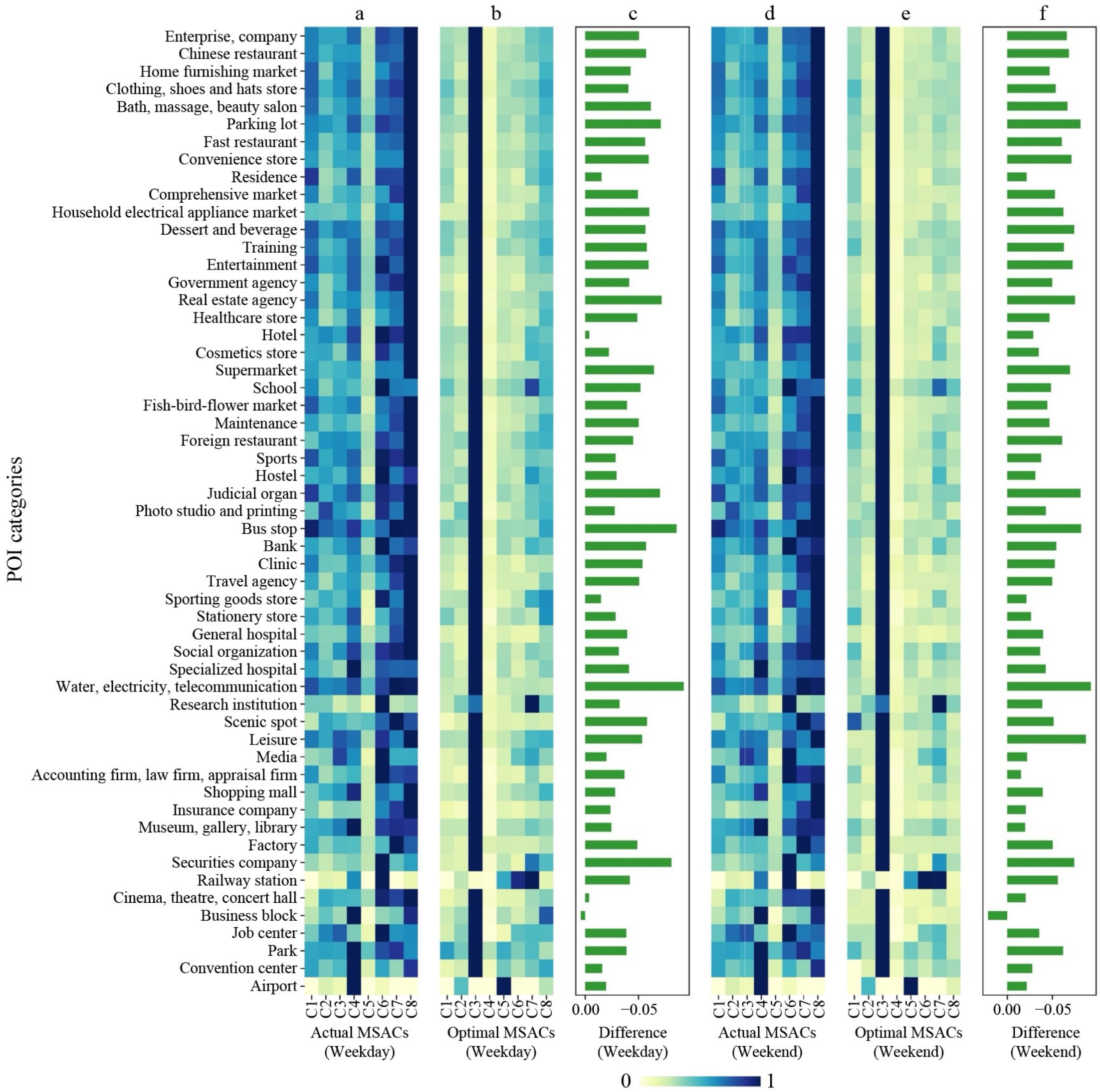


Fig. 7 Comparison of land use density between the actual MSACs and optimal MSACs.

above 0.9) than that of individual MSAs, through t-tests[1)], we find no significant differences between the actual MSACs and optimal MSACs (p-value is 0.566 for weekdays, p-value is 0.832 for weekends). Land use mix at the MSAC scale probably does not play a significant role in the formation of spatial cluster structure.

The MSA pair scale

Another mesoscopic approach involves investigating spatial interactions within the MSACs. As mentioned earlier, two types of undesirable passenger flows exist. The first type occurs within the actual MSACs, characterized by high trip numbers and long travel times. The second type occurs within the optimal MSACs, characterized by low trip numbers and short travel times. Ideally, the passenger volume of the former should be minimized, while that of the latter should be maximized. This approach aims to achieve optimal spatial cluster structures.

By grouping all internal passenger flows into quartiles, two types of undesirable passenger flows can be identified, as shown in Fig. 9. The upper quartile within the actual MSACs represents the former, while the lower quartile within the optimal MSACs represents the latter. These two types of flows represent opposite poles of spatial interactions. To understand the influence of land use factors on these flows, our focus is on the origin and destination MSAs. By examining the differences in land use characteristics between the upper quantile MSA pairs and lower quantile MSA pairs, we can gain insights into why these two types of undesirable spatial interactions exist. These insights can be valuable for urban planners in the development of TOD policies.

1) In this study, before performing t-test, Levene's test is first conducted to test the assumption of equality of variances. If p-value $\leqslant$ 0.05, t-test is run assuming equal variances, otherwise, t-test is run assuming unequal variances.

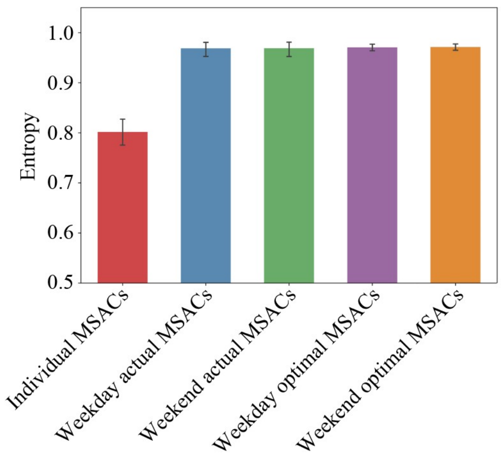


Fig. 8 Bar plots of land use mix among individual MSAs, actual MSACs, and optimal MSACs (error bars: 95% confidence interval).

The land use density of MSA pairs can influence passenger flows in two possible ways. The sum and difference of land use density values between origin and destination MSAs might influence travel demand and passenger volume. Our *t*-tests for all land use categories produce consistent results, indicating that the upper quartile MSA pairs are significantly different from the lower quartile MSA pairs in both the sum and difference of land use density values (*p*-value < 0.05). Figure 10 presents the sum and difference of land use density values for the upper and lower quantiles, with the values for the upper quantile MSA pairs being larger than those for the lower quantile ones. In other words, if lower quantile MSA pairs have larger sum and/or difference of POI density values, they are more likely to have more trips.

Furthermore, we compare the land use mix values between the upper and lower quantile MSA pairs, as shown in Fig. 11. The entropy values for the upper quantile MSA pairs are significantly larger than those for the lower quantile ones (*t*-tests, *p*-value < 0.05). This indicates that if each pair of lower quantile MSAs has more mixed functions, they are more likely to have heavier passenger flows.

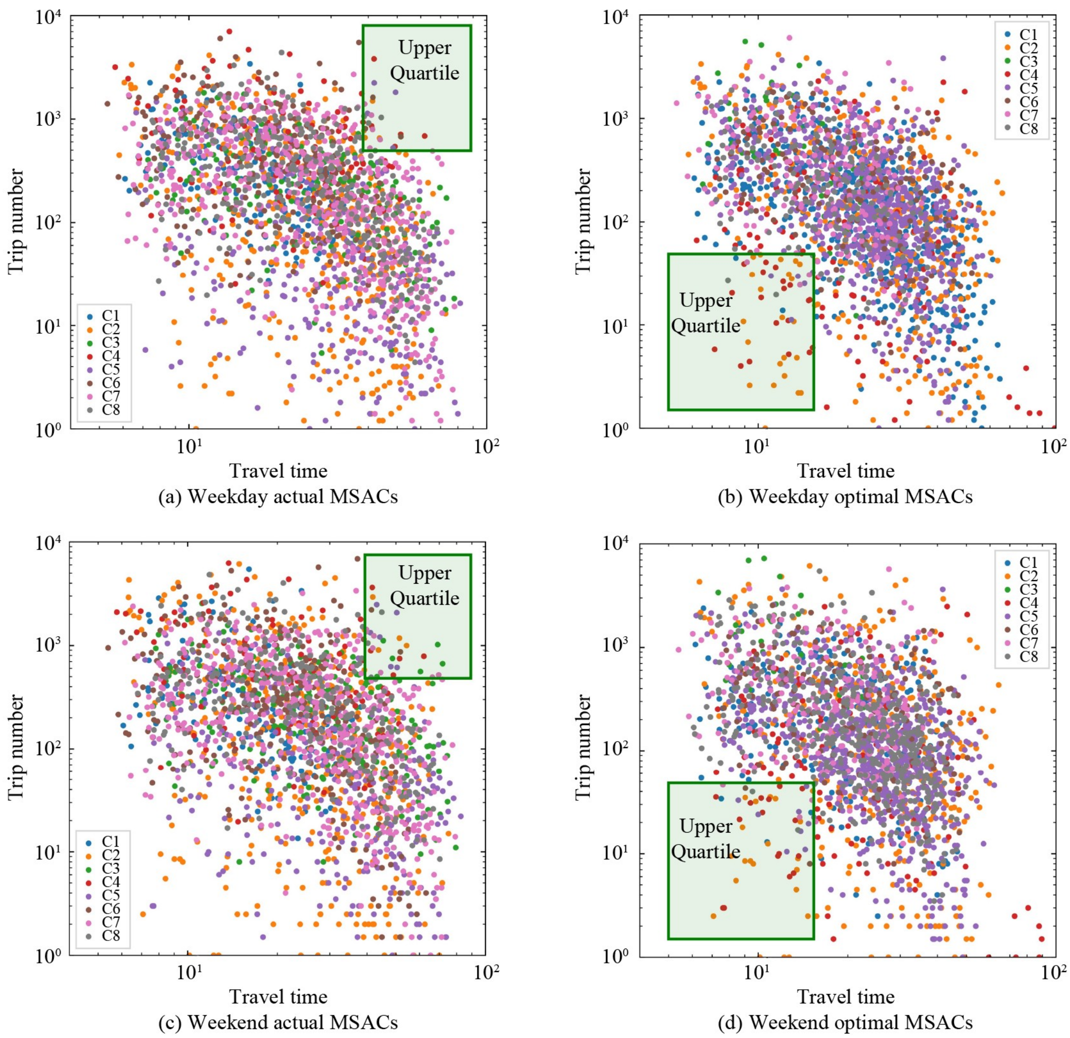


Fig. 9 Trip number and travel time of internal edges for the four sets of MSACs ((a) and (c) show the upper quartile for the actual MSACs; (b) and (d) show the lower quartile for the optimal MSACs).

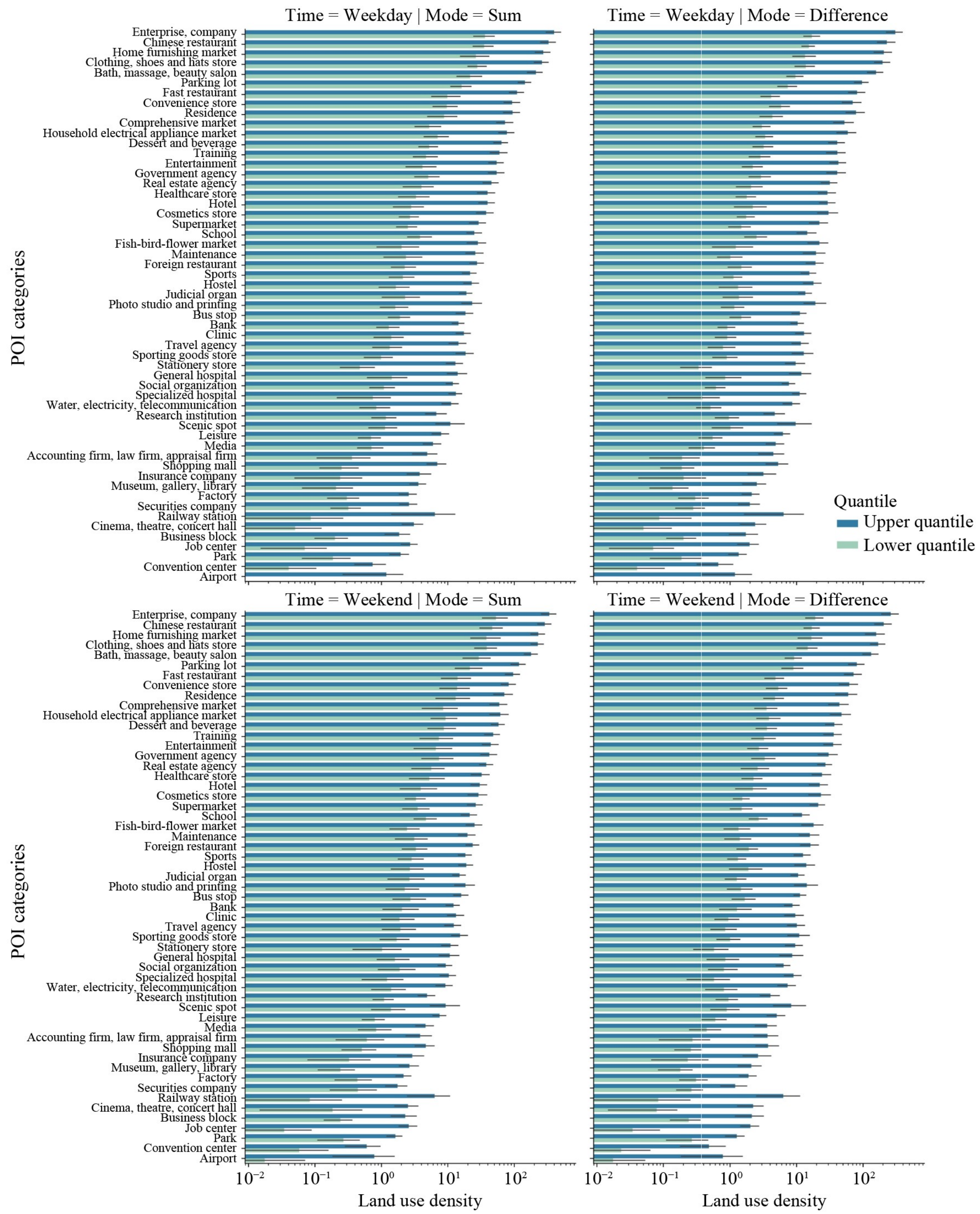


Fig. 10 Bar plots of land use density for the upper and lower quantile MSA pairs in terms of sum and difference (error bars: 95% confidence interval).

5 Conclusions

With the shift in focus toward "flow," this study argues for the importance of spatial organization in MSAs based on observed and potential metro passenger flows. Taking the case of Wuhan, this paper applies community detection to analyze two types of passenger flows and reveal the actual spatial clusters as well as the optimal spatial clusters. It is found that the identified optimal spatial clusters align with the planned sub-centers proposed by the local government. Additionally, these optimal spatial clusters exhibit shorter internal travel times compared to the actual spatial clusters, making them more desirable. Pursuing the optimal spatial cluster structure will ultimately decrease overall travel time for residents living in

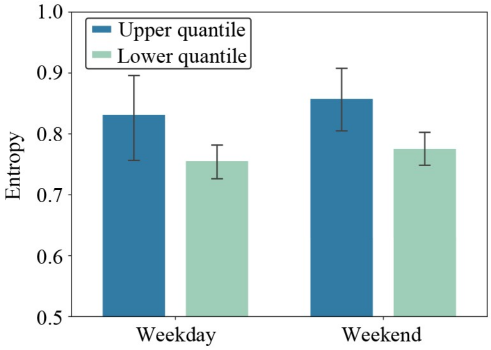


Fig. 11 Bar plots of land use mix for upper and lower quantile MSA pairs (error bars: 95% confidence interval).

MSAs, thereby improving their quality of life. Simultaneously, it will also enhance urban sustainability and livability.

We examine the effect of land use on urban structure and provide recommendations for land use policies that support optimal spatial cluster structure. The principle of promoting concentrated and mixed development within a single MSA is widely recognized as a key principle in TOD planning. It is now commonly accepted among urban planners (Jacobs, 1961; Cervero and Guerra, 2011; Song et al., 2013) that increasing land use density and promoting a mix of land uses around transit stations are essential. Building on this conventional wisdom, our study expands on the concepts of land use density and land use mix at two larger scales: the scale of MSACs and the scale of MSA pairs. The main findings and potential policy implications are as follows.

At the MSAC scale, the standard deviations of land use density values among the actual MSACs are higher than those among the optimal MSACs. These statistical results suggest that in order to promote TOD and encourage polycentric spatial development, urban planners and policymakers should aim to maintain relatively equal land use density among transit station area clusters that belong to different planned sub-regions.

At the MSA pair scale, we focus on undesirable internal travel flows within the actual MSACs, which are characterized by large trip numbers and long travel times, as well as flows within the optimal MSACs, which involve smaller trip numbers and shorter travel times. Our research reveals that the sum and difference of land use density values between the origin and destination MSAs in the former are significantly greater than those in the latter. This trend also applies to land use mix values. To promote TOD and polycentric spatial development, urban practitioners can target the elimination of the latter type of undesirable travel flows by promoting concentrated, differentiated, and mixed functional arrangements in both the origin and destination MSAs. This approach may redirect passengers from the former to the latter, resulting in a more efficient metro travel flow structure.

While many TOD researchers and planners primarily focus on the scale of a single MSA, our research investigates the spatial interactions between MSAs. We aim to provide urban policymakers and practitioners with planning strategies for functional arrangements among MSACs and MSA pairs in order to achieve synergistic effects and promote polycentric spatial development. To conclude, this research provides a comprehensive understanding of spatial cluster structures based on metro passenger flows. Additionally, it offers practical spatial planning strategies for urban planners and policymakers interested in promoting TOD and polycentric spatial development. However, due to the rapid urban growth and global expansion of rail transit, there is still much work to be done in fully understanding the relationships between urban spatial structure, travel behaviors, and land use characteristics.

It is important to approach the application of network science methods to spatial analysis with caution, particularly when considering the meaning of community detection results. In this study, the performance of the selected algorithm is assessed by examining the characteristics of the detected communities, as community detection presents an ill-defined problem (Fortunato and Hric, 2016).

Furthermore, in order to explore the evolution of spatial networks and their land use determinants, it is necessary to incorporate the factor of time. The utilization of emerging dynamic clustering techniques and the analysis of big and open data would undoubtedly facilitate this exploration. By employing the concept of flows, further investigation can be conducted to determine the extent to which nodes in a spatial network share socio-economic attributes with their neighborhoods.

Moreover, the findings of this study can be extrapolated to other cities that share similar socio-economic conditions, population sizes, and urban forms with Wuhan, and are also seeking TOD and polycentric development. Nonetheless, it is important to exercise caution when applying these conclusions to monocentric cities such as Beijing. Researchers and practitioners should proceed with careful consideration. Furthermore, additional comparative studies and empirical evidence across diverse urban contexts are necessary to establish more robust conclusions in the future.

Competing Interests The authors declare that they have no competing interests.

References

Burger M, Meijers E (2012). Form follows function? Linking morphological and functional polycentricity. Urban Studies, 49(5): 1127–1149

Calthorpe P (1993). The Next American Metropolis: Ecology, Community, and the American Dream. Princeton: Princeton Architectural Press

Castells M (1996). The Rise of the Network Society. Oxford: Blackwell
Cervero R, Guerra E (2011). Urban densities and transit: A multi-dimensional perspective
Chakour V, Eluru N (2016). Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal. Journal of Transport Geography, 51: 205–217
Chen W, Chen X, Cheng L, Liu X, Chen J (2022). Delineating borders of urban activity zones with free-floating bike sharing spatial interaction network. Journal of Transport Geography, 104: 103442
Curtis C, Olaru D (2010). The relevance of traditional town planning concepts for travel minimization. Planning Practice and Research, 25(1): 49–75
Ducruet C, Beauguitte L (2014). Spatial science and network science: Review and outcomes of a complex relationship. Networks and Spatial Economics, 14(3–4): 297–316
Fortunato S, Hric D (2016). Community detection in networks: A user guide. Physics Reports, 659: 1–44
Gao S, Liu Y, Wang Y, Ma X (2013). Discovering spatial interaction communities from mobile phone data. Transactions in GIS, 17(3): 463–481
Girvan M, Newman M E (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America, 99(12): 7821–7826
Giuliano G, Small K A (1991). Subcenters in the Los Angeles region. Regional Science and Urban Economics, 21(2): 163–182
Ibraeva A, De Almeida Correia G H, Silva C, Antunes A P (2020). Transit oriented development: A review of research achievements and challenges. Transportation Research Part A, Policy and Practice, 132: 110–130
Jacobs J (1961). The Death and Life of Great American Cities. London: Vintage
Jin M, Gong L, Cao Y, Zhang P, Gong Y, Liu Y (2021). Identifying borders of activity spaces and quantifying border effects on intra-urban travel through spatial interaction network. Computers, Environment and Urban Systems, 87: 101625
Jun M J, Choi K, Jeong J E, Kwon K H, Kim H J (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. Journal of Transport Geography, 48: 30–40
Kloosterman R C, Musterd S (2001). The polycentric urban region: Towards a research agenda. Urban Studies, 38(4): 623–633
Leslie T F (2010). Identification and differentiation of urban centers in Phoenix through a multi-criteria kernel-density approach. International Regional Science Review, 33(2): 205–235
Li F, Zhou T (2019). Effects of urban form on air quality in China: An analysis based on the spatial autoregressive model. Cities, 89: 130–140
Li Y, Liu X (2018). How did urban polycentricity and dispersion affect economic productivity? A case study of 306 Chinese cities. Landscape and Urban Planning, 173: 51–59
Lin D, Allan A, Cui J (2015). The impact of polycentric urban development on commuting behaviour in urban China: Evidence from four subcentres of Beijing. Habitat International, 50: 195–205
Liu X, Derudder B, Wu K (2016). Measuring polycentric urban development in China: An intercity transportation network perspective. Regional Studies, 50(8): 1302–1315
Liu X, Gong L, Gong Y, Liu Y (2015). Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography, 43: 78–90
Liu X, Wang M (2016). How polycentric is urban China and why? A case study of 318 cities. Landscape and Urban Planning, 151: 10–20
McMillen D P (2001). Nonparametric employment subcenter identification. Journal of Urban Economics, 50(3): 448–473
Munoz Mendez F, Klemmer K, Han K, Jarvis S (2018). Community structures, interactions and dynamics in London's bicycle sharing network. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 1015–1023
Reichardt J, Bornholdt S (2006). Statistical mechanics of community detection. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 74(1): 016110
Roth C, Kang S M, Batty M, Barthélemy M (2011). Structure of urban movements: Polycentric activity and entangled hierarchical flows. PLoS One, 6(1): e15923
Song Y, Merlin L, Rodriguez D (2013). Comparing measures of urban land use mix. Computers, Environment and Urban Systems, 42: 1–13
Stead D, Marshall S (2001). The relationships between urban form and travel patterns. An international review and evaluation. European Journal of Transport and Infrastructure Research, 1: 113–141
Sun B, He Z, Zhang T, Wang R (2016a). Urban spatial structure and commute duration: An empirical study of China. International Journal of Sustainable Transportation, 10(7): 638–644
Sun Y, Fan H, Li M, Zipf A (2016b). Identifying the city center using human travel flows generated from location-based social networking data. Environment and Planning. B, Planning & Design, 43(3): 480–498
Tanahashi Y, Rowland J R, North S, Ma K L (2012). Inferring human mobility patterns from anonymized mobile communication usage. In: Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia. ACM: 151–160
Traag V A, Bruggeman J (2009). Community detection in networks with positive and negative links. Physical Review E, 80: 036–115
Vasanen A (2012). Functional polycentricity: Examining metropolitan spatial structure through the connectivity of urban sub-centres. Urban Studies, 49(16): 3627–3644
Vasanen A (2013). Spatial integration and functional balance in polycentric urban systems: A multi-scalar approach. Tijdschrift voor Economische en Sociale Geografie, 104(4): 410–425
Wang T, Yue W, Ye X, Liu Y, Lu D (2020). Re-evaluating polycentric urban structure: A functional linkage perspective. Cities, 101: 102672
Wikipedia (2019). Wuhan
Xiao L, Lo S, Zhou J, Liu J, Yang L (2021). Predicting vibrancy of metro station areas considering spatial relationships through graph convolutional neural networks: The case of Shenzhen, China. Environment and Planning. B, Urban Analytics and City Science, 48(8): 2363–2384
Yang J, Chen J, Le X, Zhang Q (2016). Density-oriented versus development-oriented transit investment: Decoding metro station location

selection in Shenzhen. Transport Policy, 51: 93–102

Yue Y, Zhuang Y, Yeh A G, Xie J Y, Ma C L, Li Q Q (2017). Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. International Journal of Geographical Information Science, 31(4): 658–675

Zhang Y, Marshall S, Cao M, Manley E, Chen H (2021). Discovering the evolution of urban structure using smart card data: The case of London. Cities, 112: 103157

Zhong C, Arisona S M, Huang X, Batty M, Schmitt G (2014). Detecting the dynamics of urban structure through spatial network analysis. International Journal of Geographical Information Science, 28(11): 2178–2199