Computers, Environment and Urban Systems 112 (2024) 102143

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Exploring spatial complexity: Overlapping communities in South China's megaregion with big geospatial data

Chenyu Fang [a], Xinyue Gu [b], Lin Zhou [c,d,*], Wei Zhang [e,*], Xing Liu [f], Shuhua Liu [g,h], Martin Werner [a]

[a] Department of Aerospace and Geodesy, Professorship for Big Geospatial Data Management, Technical University of Munich, Munich 85521, Germany
[b] Department of Land Surveying and Geo- Informatics, The Hong Kong Polytechnic University, 999077, China
[c] Chair of Land Management, School of Engineering and Design, Technical University of Munich, Munich 80333, Germany
[d] Centre for Environment, Energy and Natural Resource Governance, University of Cambridge, Cambridge, United Kingdom
[e] College of Urban and Environmental Sciences, Peking University, Beijing 10087, China
[f] China Academy of Urban Planning & Design Shenzhen, Guangdong 518000, China
[g] College of Geo-Exploration Science and Technology, Jilin University, Jilin Province 130061, China
[h] Jilin institution of Geoinformation, Jilin Province 130062, China

ARTICLE INFO

Keywords:
Megaregional analysis
Overlapping community structures
Big geospatial data
Pearl River Delta urban dynamics
OCDDP methodology

ABSTRACT

Overlapping structures, often overlooked, are crucial in shaping comprehensive urban development and broader megaregional strategies. To address the gap, this study conducts the overlapping communities analysis in the Pearl River Delta (PRD), a megaregion in South China, using big geospatial data from 2018. A novel Overlapping Community Detection based on Density Peaks (OCDDP) is employed to generate multiple communities with diverse functions for different nodes in the commuting network of 60 sub-city divisions. We identify eight overlapping communities in PRD characterized by two categories of communities predominantly centered around Shenzhen and Guangzhou, revealing a bicentric spatial structure. Notably, central sub-cities are characterized by a low-overlap attribute, while peripheral sub-cities manifest a high-overlap tendency. Furthermore, the study investigates the driving forces behind these communities through ridge regression to analyze the impacts of various spatial flows, including policies, investment amount and times, branch funding and number, travel cost, and travel distance, co-patenting, and search index. This part found that four Shenzhen-centric communities are primarily driven by travel cost, co-patenting, branch funding, and number, while the four Guangzhou-centric communities are influenced by co-patenting, investment amount, and times. This study emphasizes differentiated functional linkages and the need for precise policy positioning and resource allocation, paving the way for a coordinated and holistic approach to megaregional development.

1. Introduction

Cities, which encompass various scales, are the outcome of multiple interactions among diverse components. This understanding has been effectively elucidated through the lens of complexity science, emphasizing the interwoven dynamics among system components, rather than focusing solely on individual properties (Palla et al., 2005). Complex systems are characterized by the nonlinear interactions and adaptive behavior of constituent elements (Bassolas et al., 2019; Caldarelli et al., 2023), demonstrating the nested and overlapping structure of the communities (Palla et al., 2005). Thus, it is crucial to consider the intricate dynamics of cities shaped by the multifaceted interactions among their diverse components.

Complex urban systems achieve a shift from geographically multiscale to functionally multilevel characteristics as demographic, social, and economic linkages become more nested in cities (Bai et al., 2017; Tang et al., 2021). This perspective implies that each community does not serve a single function in isolation but shares multiple functions with others (Ahn, Bagrow, & Lehmann, 2010). Although cities often fulfill multiple global roles, such as cultural centers, economic hubs, or transportation nodes, urban studies frequently approach these functions in isolation, with limited exploration of their multilevel overlapping dynamics.

The shift towards recognizing overlapping communities in urban

* Corresponding authors.
E-mail addresses: lin.zhou@tum.de (L. Zhou), zhangwei@stu.pku.edu.cn (W. Zhang).

https://doi.org/10.1016/j.compenvurbsys.2024.102143
Received 15 December 2023; Received in revised form 11 June 2024; Accepted 19 June 2024
Available online 2 July 2024

studies emerges from the growing complexity of city systems, which are no longer characterized by singular, homogeneous characteristics but rather by a rich mosaic of overlapping identities and functions (Zhang et al., 2020). This nuanced perspective offers a fresh lens to examine urban systems, moving beyond traditional compartmentalized analyses to reveal the intricate, multilayered interdependencies that underpin urban evolution and functionality (Peixoto, 2021). By pinpointing intersections where diverse social, economic, and infrastructural functions overlap, this approach uncovers subtle nuances of urban connectivity and cohesion, often overlooked by conventional non-overlapping models. Such insights not only enhance our understanding of the spatial and functional organization within individual cities but also shed light on broader megaregional structures. It highlights how cities within a megaregion are interconnected through layers of shared community functions, thus offering a more comprehensive view of urban dynamics (Palla et al., 2005). Through this lens, we gain insights into the fluidity and dynamism of urban spaces, providing a deeper understanding of how cities evolve and adapt in response to changing demographic and economic landscapes, and allowing for a fuller appreciation of the complex spectrum of interactions that shape megaregional structures.

Recent studies have regarded overlapping communities as a way to better comprehend the layered structure's multilayer nature in the fields of economics, sociology, and biology (de la Torre, Kalda, Kitt, & Engelbrecht, 2018; Pavesi et al., 2018; Zhou & de Vries, 2022). Finding groups of nodes in a network that perform similar tasks improves the representation of urban systems as a whole (Palla et al., 2005; L. Zhou, Zhang, Fang, Sun, & Lin, 2020; Zhou, Zhou, De Vries, Liu, & Sun, 2024). However, there is still a lack of research on the exploration of overlapped functions and the failure to uncover multilevel nested structures in urban studies. One way to better grasp the complex urban spatial structure and the hierarchical structures of mixed-use planning is to examine how various functions interact and how different functions might be combined. For this reason, it is essential to take a complex systems viewpoint and include overlapping community detection in studies of urban spatial organization (Peel, Larremore, & Clauset, 2017; Peixoto, 2021).

Another issue is in understanding the causes of overlapped structures. Many studies still remain on the exploratory analysis of urban structure, which spans various historical and contemporary perspectives. This comprehensive overview of spatial structure analysis in urban studies traces its evolution from clustering socioeconomic attributes to exploring flow- or network-based urban functions, examining historical approaches, spatial dependence perspectives, functional area delineation, and network-based methodologies (Li et al., 2022; Zheng et al., 2022; Cainelli et al., 2022). Despite these approaches, there remains a gap in understanding complex, interwoven urban structures and the driving forces behind their formation due to limited data availability and methodological challenges.

To fill these research gaps, we explored the multilevel spatial structure and its driving forces in the Pearl River Delta (PRD) of southern China, by adopting a novel overlapping community detection based on density peaks (OCDDP) and Ridge Regression, with the aim of providing valuable insights for effective megaregional integration strategies. We first tracked 1,718,068 flows from or to other sub-cities of PRD from a large anonymous mobile phone signaling (MPS) data set from 16 April to 1 May 2018. These data sets were distributed in 3600 pairs to construct the sub-city network and detect distinct types of overlapping communities. We utilized a series of novel multi-source spatial big data, including flows of branch flows, investment patterns, innovation networks, information exchange, policy interventions, and traffic patterns, to explore the causal effects of the formation of each overlapping community. We attempted to explore the following three research questions: (1) What are the underlying patterns and characteristics of overlapping communities? (2) What are the key driving forces behind the formation of overlapping communities, and how do these communities impact megaregional integration strategies? (3) How and to what extent do various factors contribute to the formation and stability of overlapping communities?

The rest of the paper is structured as follows. Section 2 reviews a new analytical perspective on overlapping structures as well as the driving forces of varying spatial structures. Section 3 describes the research area, dataset, and methods. Section 4 presents the empirical findings. Section 5 discusses and section 6 concludes.

2. Literature review

2.1. Redefining spatial structures in urban planning: Embracing the overlapping community paradigm

In urban planning discourse, complex network theory has emerged as an instrumental tool in interpreting spatial structures, with the availability and accessibility of abundant flow data. Numerous studies have harnessed multi-source spatial big datasets encompassing transport, human mobility, information trajectories, and capital movements to explore the spatial structure. For example, Bationo et al. (2023) applied private capital flow data to delineate connections between various production sectors, culminating in the formation of a national community structure. Similarly, Yu, Li, Yang, and Zhang (2020) utilized mobile data to unearth the commuting structure, thereby highlighting pronounced employment-housing linkages across urban districts.

Yet, a notable cognition among these investigations is their concentration on single-layer structures in diverse functional urban spaces (Yin, Soliman, Yin, & Wang, 2017). Their objective often revolves around demystifying the city's networked system in cyberspace via precise spatial mapping. Such approaches inherently presuppose that each node in the network is a unique member of a single community. Such a reductionist perspective may inadvertently truncate the analysis, potentially omitting the multifaceted interrelations within the network and neglecting embedded layers (Yu et al., 2020; Palla et al., 2005; Peixoto, 2021).

Contrary to this prevailing methodological approach, empirical observations suggest that many networks comprise highly overlapping cohesive clusters or communities, with nodes that are intricately intertwined (Palla et al., 2005). Such nested structures resonate with various socio-cultural dimensions (Peixoto, 2021). This refreshed viewpoint on the spatial structure has garnered significant attention in various fields, including economics (de la Torre et al., 2018; Onnela, Chakraborti, Kaski, Kertész, & Kanto, 2003), biology (Pavesi et al., 2018; Spirin & Mirny, 2003), and sociology (Elyasi, Meybodi, Rezvanian, & Haeri, 2016; Lei, Zhou, & Shi, 2019).

The relevance and implications of community overlaps in network theory are profound. A network's complexity across different disciplines can be vividly portrayed by nodes spanning multiple communities. For example, an extremely intricate network built through characteristics of family, schoolmates, and hobbies can be illustrated. Protein can be associated with complexes or functions, and in this way, protein belonging to divergent communities establishes large and complicated networks. It is evident that these overlapping communities exhibit a functionally multi-level network structure than their non-overlapping counterparts, which facilitates the investigation of complex network's nested patterns (Ahn et al., 2010). In light of this, the exploration of overlapping communities emerges as a pivotal refinement to conventional methodologies, aiming to capture the nuanced interplay of multiple identities and roles within urban networks. This perspective underscores the critical need for an overlapped framework in understanding the intricacies of urban and megaregional structures, as it enables a more nuanced depiction of inter-community connections and the multifunctional nature of urban spaces (Lei et al., 2019; Palla et al., 2005).

While the mentioned studies have acknowledged the value and significance of overlapping structures in discerning interwoven communities, there remains a notable research gap in the field of geography and urban planning. This lacuna in our understanding of diverse communities' characteristics and needs limits our ability to delve into overlapping communities, which are pivotal in shedding light on urban boundaries, functions, and resource distribution (Bai et al., 2017; Bationo et al., 2023; Palla et al., 2005). Moreover, advancements in networked communication and transportation technologies have paved the way for optimizing urban facets such as population, land use, capital deployment, technological integration, and information dissemination. These innovations have fostered the emergence of sophisticated synergistic networks between cities and introduced more efficient resource allocation methods (Fang & Zhao, 2018; Priemus & Hall, 2004). The integration of an overlapping communities' perspective in urban planning and geography thus addresses a critical gap, offering a sophisticated lens through which to examine the dynamic interconnections and layered functionalities within and across urban regions. This approach is particularly relevant in the context of megaregional analysis, where understanding the depth and breadth of inter-city relationships is crucial for devising nuanced strategies for spatial organization and regional cohesion. Given these developments, there's an emergent and compelling demand to pivot towards network-centric analyses of overlapping structures, combined with finer-grained big data, to enhance our grasp of complex urban phenomena and inform effective planning initiatives.

2.2. Varying analytical perspectives on the driving force of spatial structure

The detection of spatial structure has progressed from identifying clusters of socio-economic attributes to probing the functional interactions within urban networks (Derudder & Taylor, 2018; Ross, 2012). To effectively understand spatial structures, it is crucial to examine a variety of structures across different flow data sets and analyze the key drivers behind them. This paper thoroughly reviews the factors and methodologies used to identify the determinants of spatial structures, leveraging empirical studies that offer a range of valuable insights.

Early studies on spatial structure often portrayed it as a mosaic, employing factor analysis and cluster analysis to delineate the spatial distribution of various attributes. These studies aimed to classify social areas by similarities in attributes, adopting two primary approaches for explaining the formation of these areas (Shevky, 1972; Feng Jian, 2018; Zhang et al., 2020). One approach involves estimating simple linear correlations among variables using methods like density functions, logit models, or higher-order factor analysis (Anas, 1990; Davies & Musson, 1978; Parr, 1985). The second approach is chronological, involving the overlay of maps from multiple periods to study the continuity and evolution of urban spatial patterns (Fulong & Yeh, 1999; Parr, 1979; Riitters et al., 1995). These studies typically incorporated census population, residential density, and business investment as the variables. While they focused primarily on regional homogeneity and heterogeneity, only traditional maps were used for the research.

A second perspective focuses on spatial dependence, examining the relationships between a focal unit and its neighbors to detect spatial structure through the aggregation and dispersion of spatial attributes (Getis & Ord, 2010; McMillen, 2001). Unlike the first perspective, this view prioritizes distance as a key variable, investigating spatial factors in relation to geographic distance using models like distance decay exponential models, gravity models, and distance proximity matrices (Griffith & Jones, 1980; Sohn, 2005). In order to understand how connections are formed within a specific distance, researchers also analyze the density gradients of populations and employments within each spatial unit (Getis & Ord, 2010; McMillen, 2001). This perspective goes beyond mere statistical analysis and incorporates spatial mathematical models that are related to geography and urban planning.

Furthermore, when considering spatial structures based on flows or networks, a third perspective aims to identify and define functional areas characterized by interdependent socio-economic links (Farmer & Fotheringham, 2011; Zhang et al., 2020). These functional areas can be regarded as spatial structures with strong commuting connections, linked to fundamental concepts such as central places, urban hierarchy, and commodity flows (Balland & Boschma, 2021; Casado-Díaz, 2000). This comprehensive approach seeks to elucidate the driving forces behind spatial structures, typically using data on commuting, employment-housing balance, regional demographics, land use, and goods flows (Hu, Yang, Yang, Tu, & Zhu, 2020; Lee, 2007; Xu et al., 2020).

In contrast to these approaches, recent studies have increasingly employed network-based methods to examine urban networks or spatial structures, utilizing advanced techniques in network visualization and mapping (Zhang et al., 2020; Zhang & Thill, 2019). These analyses vary in scale, ranging from small-scale urban studies to large-scale national or international. Small-scale analysis mainly studies within a city, which often employs centrality indices and binary metrics to reveal the formation of urban structures (Zhang & Thill, 2019; Zhao, Derudder, & Huang, 2017). Medium-scale analysis delves into spatio-temporal networks within megaregions, uncovering spatial structures shaped by converging events and activities (Zhang & Thill, 2017; Zhou, Yue, Li, & Wang, 2016). Large-scale analysis, spanning national or international scope, constructs city-pair or industry-pair matrices to analyze the city connectivity and decipher the driving forces behind these networks (Derudder et al., 2013; Z. Wang, Fu, Liu, & Liao, 2023).

Despite progress in comprehending network relationships in urban contexts, insights into the drivers of network structure are still scarce, primarily due to the scarcity of multi-source data and established research methods. Our study addresses this gap by employing ridge regression, a robust statistical tool known for its effectiveness in handling multicollinearity issues. This method, combined with the use of rare and valuable multi-source flow data, enables a more precise understanding of the dynamics within urban network systems. This comprehensive approach directly informs the creation of targeted, data-driven strategies for megaregional development.

3. Methodology

3.1. Research area

This research is centered on the Pearl River Delta (PRD), a prominent mega-city region in southern China, adjacent to the special administrative regions of Hong Kong and Macau (Fig. 1). As outlined in the *2019 Outline of the Development Plan of Guangdong, Hong Kong, and Macao Greater Bay Area,* this area is envisaged to transform into an integrated Greater Bay Area. The PRD encompasses a total of 60 sub-cities, with nine of these - Guangzhou, Shenzhen, Zhuhai, Foshan, Dongguan, Zhongshan, Zhaoqing, Huizhou, and Jiangmen - functioning as the primary prefecture-level administrative units.[1]

The cities within the PRD have significantly contributed to industrial development and talent attraction through enhanced connectivity and substantial improvements in transportation and infrastructure networks. This dynamic development has established the PRD as a globally recognized urban megaregion, distinguished by its complex and interwoven network of cities and sub-cities (Zhang et al., 2020). This intricate urban fabric exemplifies a successful model of regional integration, setting a precedent for urban development and planning.

[1] Significantly, Dongguan and Zhongshan stand out as two unique cities that do not utilize the traditional district and county divisions. For conducting analyses on the sub-city level, we reference the subdivisions outlined in each city's Master Plan. Specifically, Dongguan is divided into six sub-city clusters: Urban Core, East, Songshan Lake, Southeast, Binhai, and New Town. In contrast, Zhongshan is organized into five sub-city clusters: Northeast, East, South, Northwest, and Urban Core.

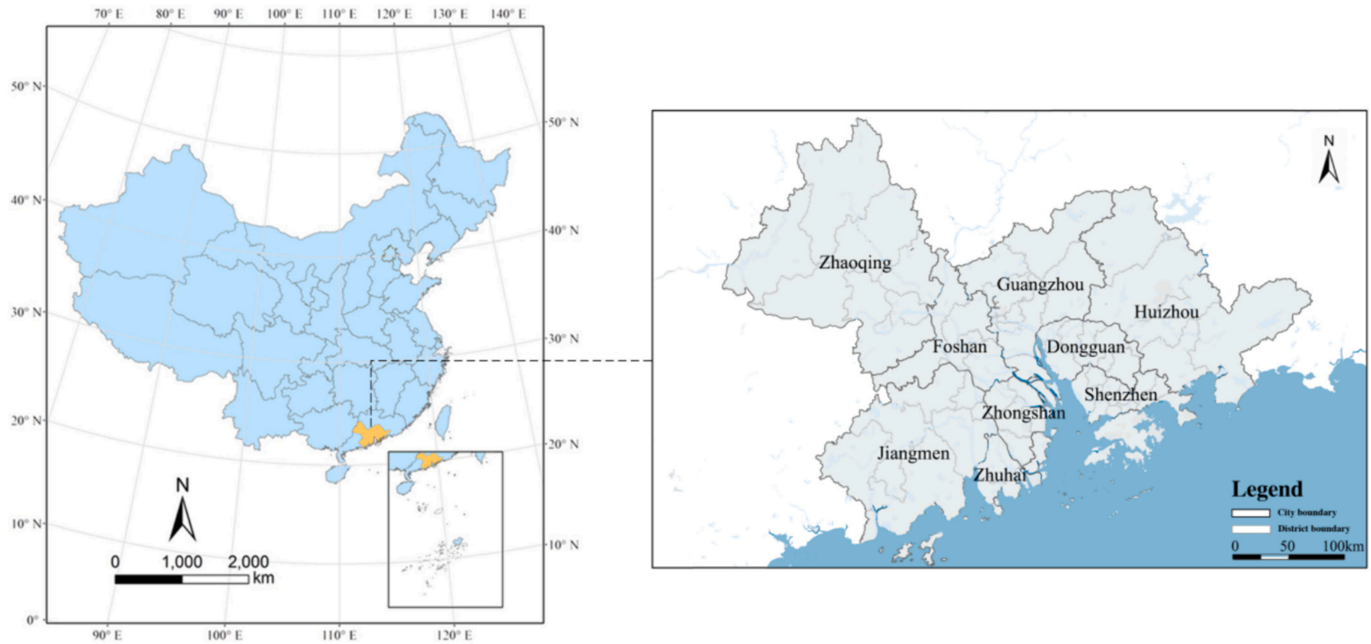


Fig. 1. Geographical Overview of 9 Cities and 60 Sub-Cities within the PRD Megaregion, China.

3.2. Data collection

This research employs mobile phone signaling (MPS) data to detect the overlapped structures within urban areas. The dataset includes information on cell phone users' estimated residential and workplace locations in the Pearl River Delta (PRD) from April 16th to May 1st, 2018, derived from mobile phone trajectories. Residential locations were inferred from the longest stops between 9 PM and 8 AM, while workplaces were identified from stops between 7 AM and 7 PM. The study focuses on users exhibiting regular movement between these points to map out a commuting network, operating under the assumption that such consistent patterns are indicative of commuting behavior. The analysis encompassed over 12 million commuters, primarily within major PRD cities. While 85.31% of commuting takes place within the same districts, this research underscores the significance of 1,718,068 inter-city flows distributed across 3600 city pairs, which are instrumental in deciphering the complexities of regional connectivity. Fig. 2 displays these flows on a 500*500 m grid, constructing a detailed commuting network that reveals interconnections between sub-cities. Unlike traditional methods, such as field surveys or geo-tagged social media data, using big data on mobile phone trajectories allows for a more detailed network analysis that transcends the boundaries of cities or counties. Each record in this dataset includes anonymized user details, residence and workplace locations, and residence time, with its richness and breadth being pivotal for a comprehensive examination of spatial interrelationships and community dynamics within the PRD.

The study delineates a comprehensive network of the Pearl River Delta's (PRD) urban landscape by employing a multi-dimensional dataset to analyze the driving forces behind the formation of overlapping communities. This dataset encapsulates various element flows that signify the economic, innovative, informational, policy, and traffic interconnections among the sub-cities within the PRD, as illustrated in Table 1 and Fig. 3. Table 1 encapsulates the key data variables and metrics utilized to analyze the inter-city connectivity and the multifaceted interactions that underlie the formation of overlapping communities within the region. Fig. 3 depicts the spatial distribution of various flows within the Pearl River Delta.

Branch number flows (*BRnum*) represent the count of enterprise branches between two city pairs, signifying the magnitude of business expansion and economic interdependency. Branch funding flows (*BRfund*) indicate the transfer of funds due to the movement of goods between headquarters, reflecting the financial transactions that facilitate economic activities across the region. These variables elucidate the economic linkages between city pairs, underscoring the pivotal role such connections play in facilitating individual mobility. As individuals navigate between various firm branches, they engage in a quest for information, market opportunities, collaborative ventures, and intra-organizational communication (Argote & Ingram, 2000). In light of this, *BRnum* and *BRfund* are incorporated within our analysis to delineate the influence exerted by enterprise branches on the movement patterns observed.

Investment times flows (*IVnum*) and investment amount flows (*IVcapital*) are indicative of the frequency and volume of investments made between headquarters, serving as a barometer for the region's capital investment climate. Innovation flows (*patent*) quantify instances of innovation cooperation, as evidenced by co-patenting activity (Wang et al., 2020), which serves as a direct indicator of the cities' collaborative research and technological development. Financial activities and knowledge cooperations rely heavily on private and face-to-face communication, which are instrumental in the exchange of tacit knowledge and complex information (Bhimani & Langfield-Smith, 2007; Gehrig, 1998; Tchamyou, 2019). Thus, these two variables come with high intensity flows of individuals, which would impact the identification of overlapping communities.

Information flows (*search*), captured through the pairwise frequency of sub-cities searched on web pages (Zhong et al., 2019), reflect the digital connectivity and the prominence of the cities in the information space (Zhen et al., 2019). Policy flows (*policy*), deduced from the frequency of sub-cities' mentions in policy documents, offer insights into the extent of policy influence on regional integration and development. These variables not only capture the digital and legislative connections between sub-cities but also hint at their potential impact on the movement patterns of individuals. The interplay between digital visibility, policy engagement, and physical mobility underscores the complex dynamics at play in shaping urban and regional landscapes in the information age (Lim, 2014; Milbourne et al., 2014; van der Waerden et al., 2019).

Traffic flow data for the Pearl River Delta was calculated using the

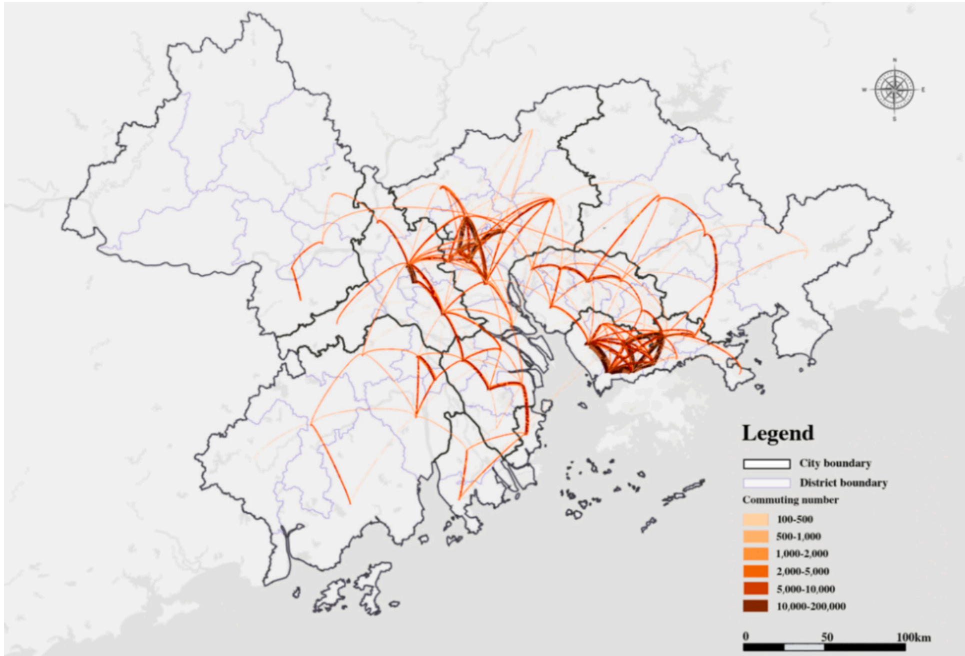


Fig. 2. The network of mobile signaling data in the 60 sub-cities in the PRD. (Note: Only the links with >100 commuters are shown).

Table 1
Overview of Key Data Variables and Metrics for PRD.

Data Type	Variable Name	Acquisition Time	Data Source	Total Pairwise links/ funding amount
Branch number	*BRnum*	Jan 1 to Dec 31, 2019	National Enterprise Credit Info. Publicity System in China	80,519
Branch funding	*BRfund*	Jan 1 to Dec 31, 2019	National Enterprise Credit Info. Publicity System in China	43.9 billion yuan
Investment times	*IVnum*	Jan 1 to Dec 31, 2019	National Enterprise Credit Info. Publicity System in China	144,520
Investment amount	*IVcapital*	Jan 1 to Dec 31, 2019	National Enterprise Credit Info. Publicity System in China	2724.7 billion yuan
Innovation	*patent*	March 1 to 31, 2021	Co-patenting data (Wang et al., 2020)	120,000
Information	*search*	March 1 to March 31, 2021	Google Web pages	54,000
Policy	*policy*	Jan 1, 2019 to Dec 31, 2021	Policy documents from government websites	16,953
Traffic cost	*TRcost*	March 1 to March 31, 2021	Baidu Maps API	2.69 million yuan
Traffic distance	*TRdis*	March 1 to March 31, 2021	Baidu Maps API	270,000

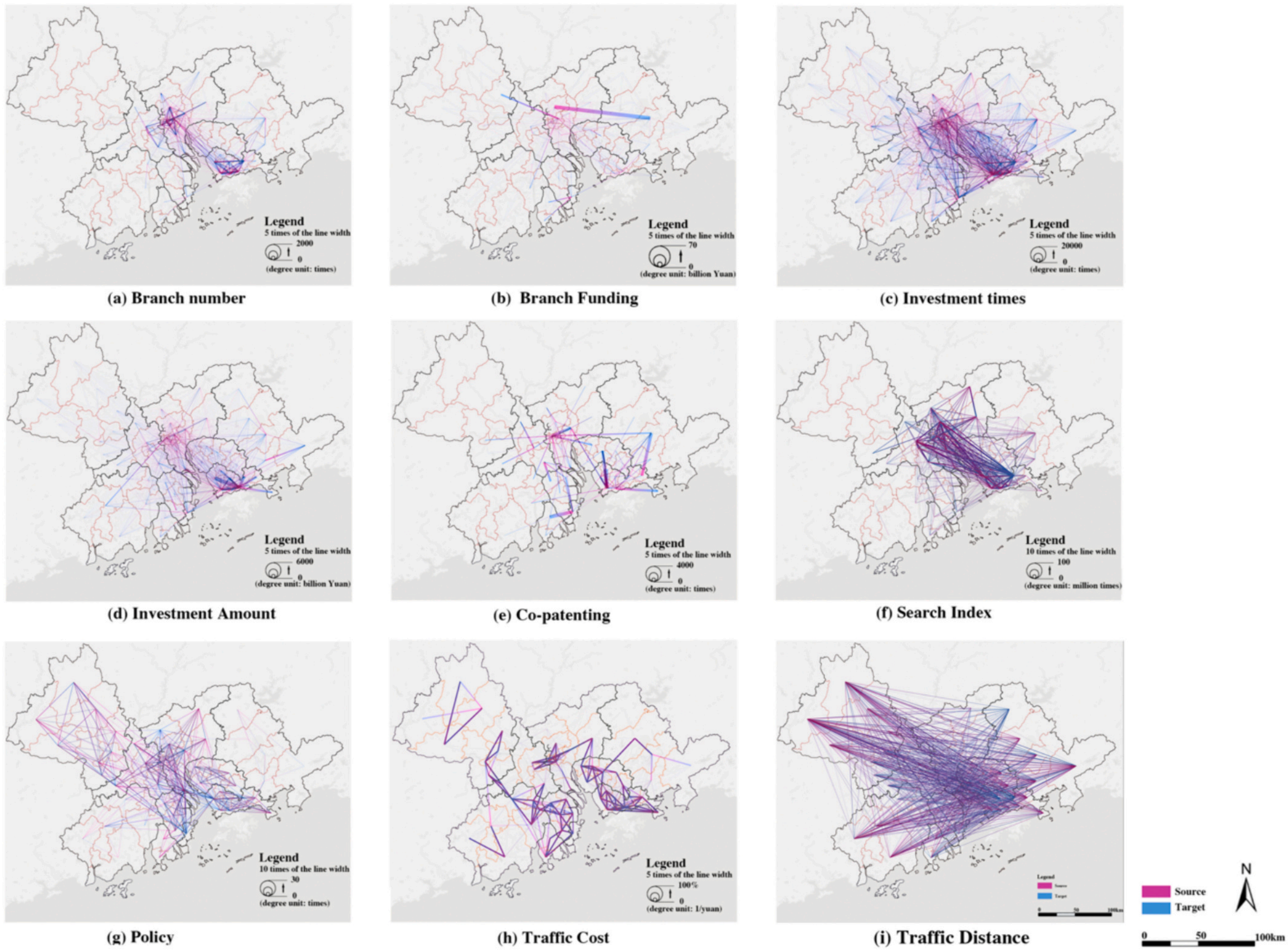
Baidu Maps API. This tool provided estimates for travel time and cost between sub-city pairs, reflecting the accessibility and economic aspects of transportation (Zhang et al., 2020). Data was collected for different times of the day to capture peak and off-peak variations, specifically at 9 am, 12 pm, 3 pm, 6 pm, and 10 pm, over the month of March 2021. To improve data visualization, travel costs were inverted, with zero costs set to one, thus standardizing the dataset for analysis. This process resulted in 270,000 data points representing the region's traffic flow. Traffic flow data, representing the intricacies of transportation linkages between sub-cities, plays a pivotal role in shaping human mobility. Accordingly, this crucial variable has been incorporated into our model.

Together, these data points represent the interconnectedness of the sub-cities, forming an immense and intricate network that facilitates the emergence of cohesive and overlapping communities. These data types serve as a foundation for the textual elaboration that follows, discussing how each flow type contributes to the functional synergy of the PRD and thereby influences the spatial structure and community dynamics of the region.

3.3. Methodology: overlapping community detection based on density peaks (OCDDP)

This study identifies the overlapping structure of the PRD using a directed commuting network, with 60 nodes indicating sub-city divisions (e.g., districts or counties) and 3600 links representing sub-city pairs. The commuting numbers in each sub-city are not specifically adjusted for the Modifiable Areal Unit Problem (MAUP), given our primary aim to understand intrinsic commuting flows rather than compare commuting volumes across different spatial scales. We also consider the context-specific nature of commuting patterns in the Pearl River Delta, where the scale of sub-cities is consistent with the functional and administrative realities of the region, aiming to reflect the region's specific urban overlapping structure. Besides, we adopt a research design that sidesteps hierarchical assumptions in favor of an equitable view of each community. This design is shaped by the complex nature of urban systems, where rigid hierarchical structures are often inadequate to capture the nuanced interplay of interconnections.

The chosen method, Overlapping Community Detection Based on Density Peaks (OCDDP) from the complex networks field, effectively captures the region's nuanced interconnections, recognizing the unique and sometimes non-linear ties among sub-cities (Bai et al., 2017). This method provides an unbiased analysis of the PRD's urban network, revealing not only the number of communities to which each sub-city



(a) Branch number (b) Branch Funding (c) Investment times (d) Investment Amount (e) Co-patenting (f) Search Index (g) Policy (h) Traffic Cost (i) Traffic Distance

Fig. 3. Spatial distribution of flows based on multi-source dataset. (Note: 85% of the flows are shown).

belongs but also identifying their core nodes,[2] thus highlighting the distinct role of each without the imposition of artificial structural constraints. Overlapping nodes are typically located at the boundaries between two communities, connecting with nodes from both communities, which leads to relatively high local density in certain instances. Through this analysis, we gain valuable insights into the intricate patterns and interconnections present within overlapping structure of the PRD.

In this model, each sub-city represents a node $i = 1, 2, \ldots N$ where N denotes the number of divisions. The communities of each node are calculated as follow:

$$\begin{cases} sab_{ij}^{0} = a_{ij} \\ sab_{ij}^{a} = sab_{ij}^{(a-1)} + \frac{1}{2}\sum_{k \neq i,j} sab_{ik}^{(a-1)} (\alpha > 0) \end{cases} \tag{1}$$

$$str_{ij} = \frac{t \bullet sab_{ij}^{a}}{mean(sab^{a})} + \frac{(1-t) \bullet n \bullet sab_{ij}^{(a)}}{\sqrt{\sum_{n} sab_{im}^{(a)} \sum_{q} sab_{qj}^{(a)}}} \tag{2}$$

$$d_{ij} = \begin{cases} \frac{1}{str_{ij}}, & i \neq j \\ 0, & i = j \end{cases} \tag{3}$$

where, a_{ij} represents the linkage strength between node i and j, d_{ij} is the distance between point i and j, sab_{ij} indicates the absolute strength of linkage between them is denoted, str_{ij} denotes the strength of linkage between node i and j, t represents weights of absolute and relative strength of linkage.

To distinguish the cores of each community, the OCDDP uses regularization to separate the two communities when calculating the distance δ_i^*, the calculation is as follows:

$$\rho_i = \sum_{j} X(d_{ij} - d_c) \tag{4}$$

$$X(x) = \begin{cases} 1 \text{ if } x < 0 \\ 0 \text{ } otherwise \end{cases} \tag{5}$$

$$\delta_i = \begin{cases} \max_{j} d_{ij} \text{ } if \text{ } \rho_i = \max_{k} \rho_k \\ \min_{j:\rho_j > \rho_i} d_{ij} \text{ } otherwise \end{cases} \tag{6}$$

$$\delta_i^* = exp\left(-\left(\frac{d_a}{\delta_i}\right)^2\right) \tag{7}$$

[2] OCDDP proves to be a valuable algorithm for detecting overlapping communities in networks, particularly in the case of real social networks, where it surpasses all other methods in comparison with the top existing method SLPA and SpeakEasy.

where, ρ_i is the local density of each point i, and d_c is the cutoff distance. The threshold d_a is selected from the list of δ. In this algorithm d_a locates around 80% of the ascending δ list in which all δ s are smaller than $1/\varepsilon$.

3.4. Driving force analysis:Ridge regression

In our study, we employed ridge regression to analyze the determinants of spatial structure. Ridge regression, an augmented form of linear regression, is adept at addressing issues of multicollinearity and limited sample sizes, making it particularly suitable for our analytical context (Hoerl & Kennard, 2000). This methodological approach enhances the traditional linear regression by integrating a regularization term, also known as the L2 penalty term, into the residual sum of squares equation. This inclusion yields a refined ridge regression estimator. The L2 penalty term systematically adjusts the magnitude of all coefficients, effectively tempering the influence of each predictor and mitigating the risk of overfitting—a common concern when variables are interdependent and the dataset for model training is constrained (Dorugade, 2014; Hoerl & Kennard, 1970; McDonald, 2009). By ensuring a more balanced consideration of predictors and reducing the likelihood of overfitting, ridge regression emerges as the optimal analytical model for assessing the impact of various potential driving forces in our investigation of spatial dynamics.

In this model, community structure (Y) is a binary dependent variable. When sub city i belongs to community c, its value is set as $Y_{ci} = 1$; otherwise, $Y_{ci} = 0$. Then we constructed 9 independent variables to represent the potential driving forces underlying the formation of overlapping community structure, as we illustrated in Fig. 3. To match with the dependent variable in the community-sub-city level, we measure each independent variable with the strength of the link between the core city c of community c and each sub-city i. Subsequently, the following ridge regression model is estimated:

$$Y_{ci} = \alpha_0 + \beta_1 patent_{ri} + \beta_1 IVcapital_{ri} + \beta_1 IVnum_{ri} + \beta_1 BRfund_{ri} + \beta_1 BRnum_{ri} + \beta_1 search_{ri} + \beta_1 policy_{ri} + \beta_1 TRcost_{ri} + \beta_1 TRdis_{ri} + \varepsilon \tag{8}$$

Where c is a community, r is the core city of community c, and i represents sub cities. Y is a dummy variable about whether one sub-city i belongs to one specific community c. $patent_{ri}$ is the number of co-patenting between sub-cities i and the core city r in community c. $IVcapital_{ri}$ is the total amount of investment capital while $IVnum_{ri}$ is the total number of investment projects between sub-cities i and core city r in community c. $BRnum_{ri}$ is the total number branches and $BRfund_{ri}$ refers to the total amount of funds transfered through branches between sub-cities i and core city r in community c. $search_{ri}$ is the times that sub city i and core city r in community c are searched together on a web page. $policy_{ri}$ represents times that sub-city i and core city r in community c are mentioned to be collaborative pair in the outline of the national economic and social development plan for each sub-city i. $TRcost_{ri}$ indicates travel cost while $TRdis_{ri}$ is travel distance between sub-city i and core city r in community c. Table 2 provides descriptive statistics for various independent variables, displaying both original and normalized data for comparability, with normalized values ensuring consistent result interpretation across all variables

4. Results

4.1. Characteristics of overlapping communities in different megaregion cities

Table 3 articulates the overlapping characteristics within the regional commuting network of PRD, structured around 60 sub-cities across 9 cities and delineated into 8 distinct yet intersecting communities via the OCDDP algorithm. 8 detected overlapping communities elucidate the intricate commuting networks and the central sub-cities' crucial roles in shaping the PRD's complex spatial configuration. The tabulation underscores the multifaceted nature of sub-cities: 40 of them are identified as being part of more than one community, exhibiting their functional versatility within the region. Specifically, 13 sub-cities are shared between two communities, 15 among three, and 12 are interlinked across four communities, indicating a complex web of shared functions and responsibilities.

The cores of each community within the regional commuting network are 8 sub-cities of Nanshan, Baoan, Futian, and Longhua in Shenzhen, along with Tianhe, Haizhu, Baiyun, and Yuexiu in Guangzhou. These 8 sub-cities (marked with a red asterisk in Table 3) represent the cores of the 8 overlapping structures, exclusively associated with a single community each, reflecting their specialized urban roles within the broader PRD region. Additionally, there are 14 other sub-cities that are similarly identified with only a single community structure. This configuration suggests a concentrated development within these sub-cities, signifying a strong, specialized influence that integrates the diverse sub-cities of the PRD into a cohesive urban entity.

Conversely, sub-cities associated with higher overlap in community structures typically exhibit more distributed resources and functions. This is evident in the 12 sub-cities that concurrently belong to up to four community structures, including Doumen District in Zhuhai, Longmen District in Huizhou, Liwan, Nansha, and Conghua Districts in Guangzhou, Shunde District in Foshan, Deqing, Duanzhou, and Sihui Districts in Zhaoqing, and Enping, Kaiping, and Taishan Districts in Jiangmen. These high-overlap sub-cities, often situated on the periphery of the PRD, fulfill diverse roles within the megaregion's network, enhancing the interconnectivity of the non-central areas.

Fig. 4 depicts eight visual layouts illustrating the grouping features and overlapping structures. The concentration of colors within certain districts indicates the intensity and clustering of network connections, thereby revealing the central hubs (highlighted with darker colors) and the extent of their influence within each community. Structures a to d focus on sub-cities predominantly located in the southeastern coastal area of the PRD, while Structures e to h encompass sub-cities primarily in the northwestern inland part of the region. This bifurcation illustrates the existence of two primary urban clusters centered around the major cities of Guangzhou and Shenzhen.

Table 2
Descriptive Statistics for independent varaibles.

Variables	N	Original Data				Normalized Data			
		mean	sd	min	max	mean	sd	min	max
$patent_{ri}$	432	43.3	279	0	4.39E+03	1.79e-09	0.992	−0.466	7.360
$IVcapital_{ri}$	432	9.38E+05	5.88E+06	0	9.92E+07	5.27e-10	0.992	−0.485	7.481
$IVnum_{ri}$	432	196	1.05E+03	0	1.97E+04	5.09e-10	0.992	−0.713	7.313
$BRfund_{ri}$	432	7.97E+03	1.41E+05	0	3.01E+06	−2.66e-09	0.992	−0.441	7.551
$BRnum_{ri}$	432	92.2	238	0	2.14E+03	−0	0.992	−0.626	5.897
$search_{ri}$	432	3.08E+07	2.18E+07	6.14E+05	1.00E+08	−1.51e-09	0.992	−1.381	5.023
$policy_{ri}$	432	1.86	1.16	0	6	−6.59e-10	0.992	−1.556	3.578
$TRcost_{ri}$	432	62.4	41.6	0	202	−7.54e-11	0.992	−1.961	3.074
$TRdis_{ri}$	432	1.20E+05	6.44E+04	9.61E+03	3.75E+05	−8.44e-10	0.992	−1.857	3.397

Table 3
Overlapping communities and core sub- cities in the PRD's commuting network.

Communities	a	b	c	d	e	f	g	h	Communities	a	b	c	d	e	f	g	h
Shenzhen (SZ)									Guangzhou (GZ)								
Nanshan		*							Tianhe							*	
Baoan			*						Huangpu							*	
Futian				*					Haizhu						*		
Longgang				*					Baiyun								*
Luohu				*					Huadu								*
Longhua	*								Yuexiu					*			
Dapeng		*	*						Zengcheng					*	*		*
Pingshan		*	*						Panyu						*	*	*
Yantian		*	*						Liwan					*	*	*	*
Guangming	*		*	*					Nansha					*	*	*	*
Zhuhai (ZH)									Conghua					*	*	*	*
Xiangzhou				*					Zhongshan (ZS)								
Jinwan		*	*	*					Northeast							*	*
Doumen	*	*	*	*					East							*	*
Dongguan (DG)									South							*	*
Urban Core				*					Northwest							*	*
East				*					Urban Core						*	*	*
Songshan Lake				*					Zhaoqing (ZQ)								
Southeast		*	*						Dinghu		*	*					
Binhai	*		*	*					Fengkai						*	*	*
New Town					*	*		*	Gaoyao						*	*	*
Huizhou (HZ)									Guangning						*	*	*
Boluo				*					Huaiji						*	*	*
Huicheng		*	*	*					Deqing					*	*	*	*
Huidong		*	*	*					Duanzhou					*	*	*	*
Huiyang		*	*	*					Sihui					*	*	*	*
Longmen				*		*	*	*	Jiangmen (JM)								
Foshan (FS)									Jianghai								*
Chancheng							*		Xinhui				*				
Gaoming							*	*	Heshan							*	*
Nanhai							*	*	Pengjiang						*	*	*
Sanshui							*	*	Enping					*	*	*	*
Shunde					*	*	*	*	Kaiping					*	*	*	*
									Taishan					*	*	*	*

Note: $i = a, b, \ldots h$ where i denotes the id of each community structure. The core sub-city of each community is highlighted in the red text.

Structures a to d encapsulate the southeastern coastal cluster of the PRD, where 25 sub-cities demonstrate intricate resource distribution and connectivity patterns. The figures from Structure a to Structure d reveal an increasing trend in the number of sub-cities, particularly within the sub-cities associated with Shenzhen(SZ), Dongguan(DG), and Huizhou(HZ). Their proximity fosters a dense network of interlocking industries and transportation routes, facilitating robust inter-city commuting links. Specifically, Structure a is depicted as the most compact community, with only four coastal sub-cities. Moving to Structure b, there is an expansion into the inland with the inclusion of Dinghu District from Zhaoqing(ZQ), suggesting a break from the traditional coastal-centric development. Structure c further broadens the coastal community, incorporating a wider range of surrounding regions and indicating a more expansive coastal influence. However, Structure d highlights a significant coastal agglomeration effect, with a stronger emphasis on the coastal city of Zhuhai(ZH) and a reduced connection to southeastern regional relations and inland communities.

Conversely, Structures e to h depict the northwest's inland cluster with 35 sub-cities, where the degree of overlap intensifies, suggesting a more uniform resource distribution. The increasing presence of sub-cities from Guangzhou(GZ), Foshan(FS), Zhaoqing(SQ), Jiangmen (JM), and Zhuhai(ZH) in these community structures speaks to the geographical contiguity and shared regional identity of these areas. Specifically, Structure e is focused on the western peripheral sub-cities contributing to cultural, industrial, and transportation networks, while Structure f extends these connections into non-peripheral regions of Zhongshan(ZS) and Huizhou(HZ). Following this, Structure g and h illustrate expansive urban networks, with Structure g capturing a collective of the aforementioned five cities, and Structure h encompassing nearly all sub-cities in the PRD's northwest.

Notedly, the Zhuhai sub-city is included in the community structures (a)-(d) centered around Shenzhen, despite the lack of geographical contiguity. This is primarily attributed to the Overlapping Community Detection by Disjoint Paths (OCDDP) algorithm, which prioritizes high-density commuting links over geographic proximity to define community structures within the Pearl River Delta. Although the external commuting volumes from Zhuhai to the core sub-cities of both Shenzhen and Guangzhou are generally low, the commuting data still reveals a relatively stronger link between Zhuhai and Shenzhen compared to other potential pairings. This connection is significantly influenced by developed shipping, railway, and road systems, which enhance the functional connectivity despite the geographical distance. Our analysis demonstrates that, especially for relatively isolated cities like Zhuhai, specific infrastructural developments can foster strong commuting flows and suggest that economic and social interactions are more indicative of community structure than mere geographic proximity.

4.2. Driving forces of overlapping communities

We conduct ridge regression to analyze the impacts of potential factors on the formation of each type of overlapping community. All variables are normalized in order to make the results of these nine divergent driving factors comparable. The results for eight communities are displayed in Table 4. We can see that the determinants behind each overlapping community are different, with only several key factors being significant. These key factors are the main forces underlying the formation of each community, which represent the external and internal associations and interactions between the sub-cities.

For Structure a in Fig. 4, the results for investment amount and search index are positively significant, indicating that the development

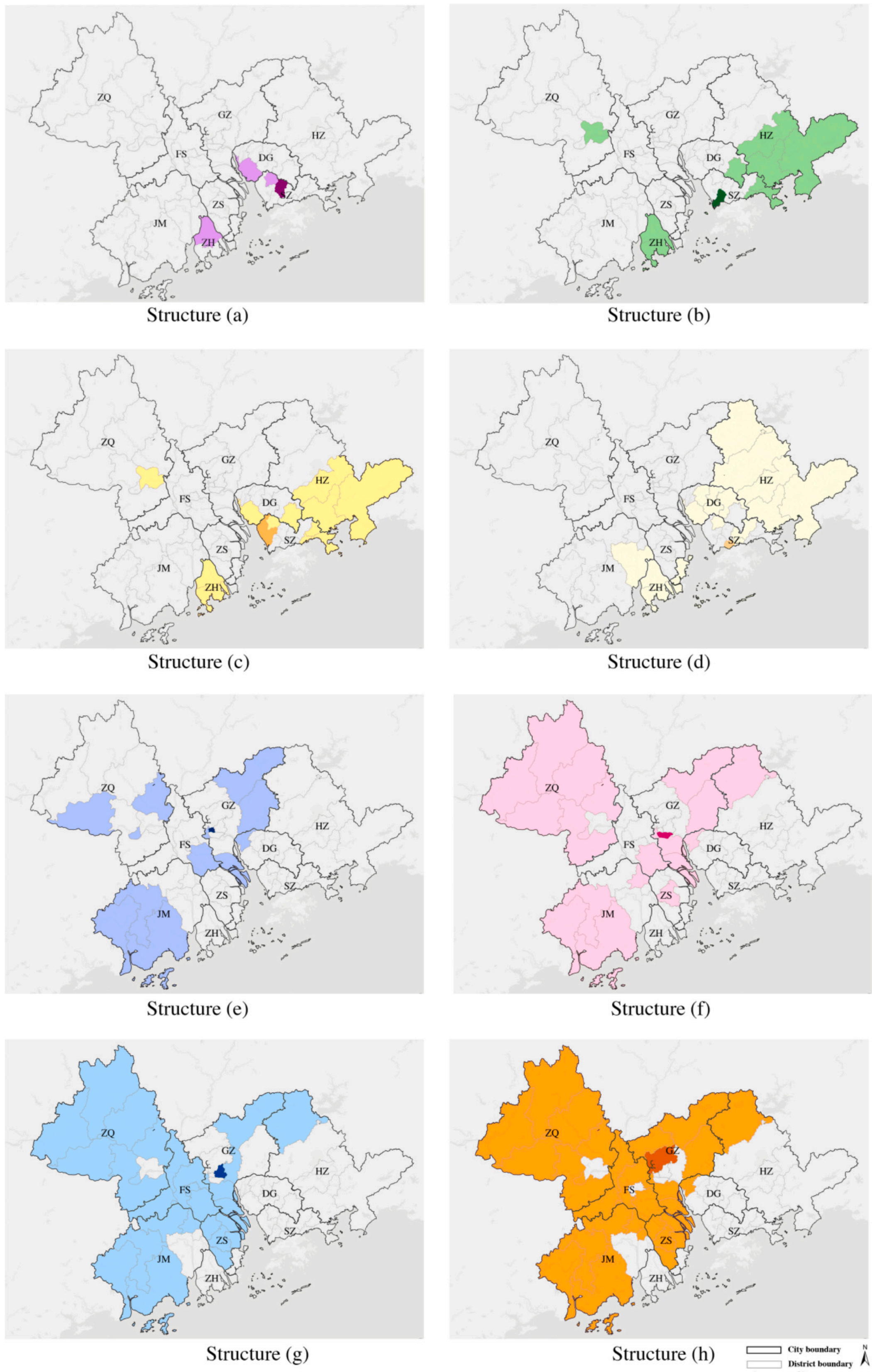


Fig. 4. Megaregional overlapping communities and the central sub-city (highlighted by darker colors) for each in the commuting network of the PRD.

Table 4
The correlation coefficient (estimate) result of bivariate logistic regression analysis to measure the driving forces of the eight community structures in the PRD.

Variables	a	b	c	d	e	f	g	h
$patent_{ri}$	0.049	−0.033	0.105*	−0.061	−0.035	0.047	0.055	0.044
	(1.332)	(−0.861)	(1.901)	(−1.267)	(−0.735)	(0.664)	(−0.508)	(0.816)
$IVcapital_{ri}$	0.066**	−0.024	−0.078	−0.487	0.098**	0.037	0. 157	−0.008
	(2.079)	(−0.513)	(−1.328)	(−0.413)	(2.03)	(0.327)	(0.936)	(−0.172)
$IVnum_{ri}$	0.041	−0.163	0.204***	0.668	−0.053	−0.147	−0. 170	−0.048
	(0.782)	(−1.14)	(4.286)	(0.472)	(−0.689)	(−0.661)	(−0.880)	(−0.746)
$BRfund_{ri}$	−0.034	0.023	−0.047	−0.121	−0.093	−0.053	−0. 007	−0.035
	(−1.370)	(0.473)	(−1.129)	(−1.376)	(−1.481)	(−0.963)	(−0.120)	(−0.890)
$BRnum_{ri}$	0.036	0. 104	−0.039	−0.130	−0.081	0.086	−0.009	0.019
	(0.726)	(0.743)	(−0.730)	(−0.417)	(−0.754)	(0.558)	(−0.072)	(0.279)
$search_{ri}$	0.093***	0. 239***	0.109**	0.180***	0.036	0.300***	0. 275***	0.011
	(3.222)	(5.841)	(2.179)	(3.263)	(0.485)	(5.647)	(4.590)	(0.238)
$policy_{ri}$	−0.017	0.120*	0.045	0.291***	0.241***	0.026	0.127	0.385***
	(−0.437)	(1.772)	(0.696)	(4.352)	(4.101)	(0.367)	(1.523)	(9.371)
$TRcost_{ri}$	0.010	−0.050	−0.024	−0.006	−0.036	−0.223	−0.105	0.109
	(0.197)	(−0.485)	(−0.265)	(−0.042)	(−0.233)	(−1.024)	(−0.584)	1.539
$TRdis_{ri}$	0.059	0.046	0.033	0.097	−0.032	0.327	0.069	−0.173**
	(1.113)	(0.529)	(0.386)	(0.781)	(−0.201)	(1.508)	(0.386)	(−2.466)
Constant	0.051***	0.169***	0.203***	0.271***	0.203***	0.339***	0.475***	0.508***
	(2.776)	(4.775)	(4.507)	(6.106)	(4.685)	(6.768)	(8.335)	(13.124)
Observations	59	59	59	59	59	59	59	59
R-Squared	0.630	0.5615	0.5274	0.5110	0.430	0.451	0.363	0.709

Note: $i = a, b, \ldots h$ where i denotes the id of each community structure.
Robust t-statistics in parentheses.
*** $p < 0.001$.
** $p < 0.01$.
* $p < 0.05$.

of this smallest community is mainly drived by capital flow and exchange of information between sub-cities and the core city of this community, Longhua. However, the practical significance of these two variables on the formation of community a is relatively smaller than other communities, with 0.066 for investment amount and 0.093 for the search index. As for Structure b, the main driving force is the search index, representing that information exchange impacts the development of this community. Besides, the coefficient of policy is significant at a 10% statistical level, showing that apart from information exchange, the policy also contributes to this process. For Structure c, the coefficients for investment number and search index are positively significant at a 5% statistical level, and the coefficient for patent cooperation is significant at a 10% statistical level. Thus, we can see that although with similar spatial distribution, Structure b, and Structure c are formed under the influence of different factors. Even though both are driven by information exchange, the development of Structure b is more policy-oriented while Structure c is more investment and innovation-oriented. As for Structure d, with Futian as the core city, this community is mainly forced by search index and policy, both positively significant at 1% statistical level.

In the above four communities, sub-cities like the Nanshan, Baoan, Futian, Longgang, Luohu, and Longhua Districts in Shenzhen(SZ) are mainly identified in one or two overlapping communities at the same time and mostly as the community centers, indicating that they each play a unity and prominent functional role in the PRD. On the contrary, Zhuhai(ZH), Dongguan(DG), and Huizhou(HZ) have sub-cities that are simultaneously identified in three or four overlapping communities, such as the Doumen District in Zhuhai(ZH), the Binhai, and New Town in Dongguan(DG), and the Huicheng, Huidong, Huiyang and Longmen Districts in Huizhou(HZ), indicating that these sub-cities take on multiple roles of policy, innovative cooperation, and industry at the same time.

The formation of Structure e are positively related to the investment amount and policy, indicating that capital flow among enterprises between sub-cities and the core city, Yuexiu, and the policy guide issued by government are the key forces underlying this community. Moreover, the coefficient of search index for Structure f and Structure g are both positively significant, and the results indicate that information exchange is the only factor that influence the development of these two communities. Finally, for Structure h, the development of the community is positively related with policy and negatively related to travel distance. Thus, the policy guide between Baiyun and sub-cities facilitate the formation of Structure h while the travel distance impede this process.

In the latter four communities, the major sub-cities are detected in more than two communities, especially the districts or counties in Zhongshan, Zhaoqing, and Jiangmen, showing that they perform multiple city functions such as information, industry, and investment in the PRD. However, there are also some sub-cities, like the Tianhe, Huangpu, Haizhu, Baiyun, Huadu, and Yuexiu Districts in Guangzhou, only detected in one community, which means that they assume a main function of crucial importance in the whole megaregion.

In sum, we can see from the results that the search index plays a significant role in the development of most communities, indicating the importance of information exchange, especially in the world of today. The significance of information exchange exceeds the limitation of space and explains the existence of a cross-regional community. Compared with the inland sub-cities in the latter four communities, the influence of information change is more significant for the eastern coastal sub-cities in the first four communities. Moreover, apart from the search index, investment between headquarters and government policy impacts the formation of several communities in this study, showing the power of the market and government. Market power is displayed more with coastal sub-cities in the first four communities, while government power is more essential for inland sub-cities in the latter four communities. The effect of transportation is only significant for community h, the largest community that includes nearly all the sub-cities in the northwest of the PRD. Thus, the transportation system still works as crucial channels connecting sub-cities, especially those in periphery regions, to the core cities. Besides, the effect of innovation is only significant for community c at a 10% statistical level, which is different from our expectation. This phenomenon indicates that innovation cooperation still hasn't played a full role in the development of communities.

5. Policy implication and discussion

The study's findings carry significant implications for regional cooperation division and rational resource allocation in the PRD and the Greater Bay Area. The multi-level community structures in PRD highlights how various drivers present spatial characteristics through the growth sub-cities. It is evident that resource allocation in the PRD remains highly unequal, necessitating further adjustments in terms of industrial, transportation, and residential balance among communities. The regression results indicate that coastal cities exhibit stronger investment and information connections, whereas inland cities have closer policy and transportation linkages, which partly aligns with Zhao et al. (2017). Coastal cities are more adept at forming intangible capital and informational cluster effects, whereas inland cities rely more on government policy and traditional transportation. These observations highlight the importance of strengthening connections between inland and coastal cities in future policy planning. Such efforts would facilitate the flow and integration of various elements across the region, ultimately promoting a more balanced and equitable distribution of resources.

The development of multiple urban clusters has been an essential planning strategy for regional development. Based on geographic location and urban connections, the *Pearl River Delta Regional Spatial Plan (2016–2020)* divided the PRD into three major clusters: "Guangzhou-Foshan-Zhaoqing," "Shenzhen-Dongguan-Huizhou," and "Zhuhai-Zhongshan-Jiangmen." However, in this study, the community structure analysis reveals that the PRD is currently divided into only two major segments, leaving cities in the last cluster integrated into the first two. Zhuhai is more closely connected to the Shenzhen metropolitan area, while Zhongshan and Jiangmen are more strongly linked to the Guangzhou metropolitan area. These findings align with previous studies on the networks and spatial structures of the PRD (Zhang et al., 2020; Zhang, Fang, Zhou, & Zhu, 2020; Zhao et al., 2017). This indicates that a polycentric structure hasn't been achieved yet, and industrial linkages and resource allocation are still influenced by the dominant central cities.

Additionally, the identification of high-overlap sub-cities highlights the problem of resource allocation imbalance existing in peripheral regions, which are mainly situated in the non-central cities of the PRD, such as Zhaoqing, Jiangmen, and Huizhou (He, Cao, & Zhou, 2021). Among these cities, certain transportation infrastructure facilitates cross-regional connections. For instance, in Structure b, a distant sub-city, Dinghu District, is connected with coastal sub-cities through its high-speed railway station, "Zhaoqing East Station," which is linked to the Shenzhen intercity railway. Moreover, Dinghu District has been actively developing an innovative and entrepreneurial industrial base, and subsequently fostering connections with the southern sub-cities in the PRD. However, the overall accessibility in these peripheral cities remains inadequate compared to other cities. To address this issue, enhancing commuter networks to connect the periphery cities with the core cities is crucial. Furthermore, Dongguan and Huizhou cities, which are adjacent to Shenzhen, need to develop a regional-level transportation system that connects longer distances. This approach will help break the over-reliance on central cities and foster the growth of new first-tier cities in the PRD. Moreover, collaboration with head enterprises and innovative capital will play a vital role in this endeavor, accelerating the development of a more balanced and sustainable urban network in the region.

Furthermore, it is important to highlight that the majority of sub-cities overlap in similar communities. For example, all districts and counties in Zhongshan are identified in Structures g and h, and all districts and counties in Zhaoqing overlap in structures g and h, except for Dinghu District, which belongs to the Shenzhen-centric community. These cities that are primarily composed of high-overlap sub-cities require special attention from the government, as they currently lack a strong pillar industry and innovative capital. The emphasis should be on not only establishing industrial connection and layout structures but also achieving continuous turnover and development in technology-related industries. Moreover, policies in these sub-cities should focus on attracting technology and high-tech talents. By fostering a favorable environment for technology-driven industries, these cities can effectively position themselves for sustainable growth and development in the ever-evolving economic landscape.

Indeed, the multilevel community structure in the PRD necessitates the implementation of related policies for coordinated cooperation. This study also underscores the significance of flow networks in driving the formation of the spatial structure. Thus, it is crucial for government to focus on facilitating the establishment of external linkages between sub-cities in the megaregion (Zhang et al., 2020). For instance, the branching of headquarters enterprises fosters economic ties and industrial activities between different urban communities, which influences the planning of high-tech industrial parks and the layout of innovative industries. Patent cooperation, on the other hand, facilitates the flows of innovative knowledge between communities, and subsequently encourages the integration of innovative industries, research institutes, and universities to form larger innovative communities. As such, the spatial structure and planning policies of mega-regions should identify differentiated functional linkages to achieve more accurate policy positioning and resource allocation.

6. Conclusion

This study introduces an innovative understanding of analyzing community structures: an overlapped perspective of how sub-cities within megaregions interact and contribute to overall spatial patterns. This insight shifts the focus from traditional, singular community analyses to a more integrated approach, acknowledging the multifaceted nature of megaregion cities and their influence on regional development. First, it identified eight distinct community structures driven by various factors, underscoring the significance and effectiveness of the multi-source data and the novel OCCDP method for finer-grained, complex, and comprehensive network analysis. Second, it provides empirical evidence on the megaregional complex dynamics and resource distribution, where coastal cities exhibit stronger investment and informational networks, and inland cities rely more on policy and transportation linkages. This calls for strategic policy interventions to facilitate balanced growth and integration between inland and coastal areas. Third, the polycentric development strategies for existing regional planning are driven by the gravitational pull of central cities like Guangzhou(GZ) and Shenzhen(SZ) and their influence on surrounding sub-cities, highlighting the challenges in resource allocation and infrastructure development of peripheral sub-cities. It advocates for enhancing transportation infrastructure and fostering innovation in these regions to facilitate cross-regional connections and reduce dependency on central cities.

This study makes significant contributions to analyzing the megaregional spatial development from three key aspects. Firstly, we introduce the concept of overlapping communities as a novel perspective to comprehend the spatial structure of megaregions. By utilizing mobile phone signaling data, we unveil the complexity and multi-functionality of sub-cities and further explore the mechanisms behind the multi-functional nature of these cities. Next, unlike previous studies relying on field surveys or geotagged data points, we leverage multi-source big flow data, encompassing human mobility, capital flow, patent flow, information flow, and traffic flow, to conduct a comprehensive analysis of the structure and driving forces of overlapping communities. This approach enables a detailed network analysis that transcends administrative boundaries and resolutions, providing a more comprehensive understanding of spatial relationships. Lastly, this study pioneers the use of the OCDDP method to explore the overlapping community structure at the megaregion scale. This method performs excellently in effectively identifying community cores and peripheral areas. Consequently, the analytical perspective of overlapping communities and utilization of multi-source data warrants further exploration in urban structure studies.

While this study provides valuable insights into urban structure, there are still certain limitations, and further improvements can be made in the following aspects. Firstly, the study is constrained by the availability of flow data and does not focus on the connection between Hong Kong, Macau, and the PRD. As the Guangdong-Hong Kong-Macao Greater Bay Area continues to develop, understanding the roles of urban communities in Hong Kong and Macau within the entire Greater Bay Area is essential. Secondly, considering the precision of flow data in time and space, this study only examines external connections in sub-cities. Future research could target certain types of flow data with higher accuracy and explore smaller-scale units for a more detailed spatial analysis. Finally, conducting comparative studies on the structure of overlapping communities in global megaregions would be valuable. Comparing the differences and variations in network and spatial structures of megaregions across various World Class Bay Areas could provide broader insights and enrich our understanding of urban development on a global scale.

CRediT authorship contribution statement

Chenyu Fang: Conceptualization, Formal analysis, Funding acquisition, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Xinyue Gu:** Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Lin Zhou:** Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Wei Zhang:** Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Xing Liu:** Data curation, Investigation, Resources. **Shuhua Liu:** Data curation, Investigation. **Martin Werner:** Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

None.

Acknowledgements

This study was conducted as part of the Ph.D. research program at the Professorship of Big Geospatial Data Management, Technical University of Munich(TUM). We extend our gratitude to the China Scholarship Council (No.202108080306) for their financial support of these doctoral studies. Additionally, we express our appreciation to the experts, mentors, doctoral candidates and reviewers whose insightful discusssions and constructive feedback significantly enhanced this paper.

References

Ahn, Y.-Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature, 466*(7307), 761–764. https://doi.org/10.1038/nature09182

Anas, A. (1990). Taste heterogeneity and urban spatial structure: The logit model and monocentric theory reconciled. *Journal of Urban Economics, 28*(3), 318–335. https://doi.org/10.1016/0094-1190(90)90031-H

Argote, L., & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational behavior and human decision processes, 82*(1), 150–169.

Bai, X., Yang, P., & Shi, X. (2017). An overlapping community detection algorithm based on density peaks. *Neurocomputing, 226*, 7–15. https://doi.org/10.1016/j.neucom.2016.11.019

Balland, P.-A., & Boschma, R. (2021). Complementary interregional linkages and smart specialisation: An empirical study on European regions. *Regional Studies, 55*(6), 1059–1070. https://doi.org/10.1080/00343404.2020.1861240

Bassolas, A., Barbosa-Filho, H., Dickinson, B., Dotiwalla, X., Eastham, P., Gallotti, R., … Ramasco, J. J. (2019). Hierarchical organization of urban mobility and its connection with city livability. *Nature. Communications, 10*(1). https://doi.org/10.1038/s41467-019-12809-y. Article 1.

Bationo, F., d'Assises, B., Griffith-Jones, S. J., Murinde, V., Soumaré, I., & Tyson, J. (2023). Private capital flows and productivity in sub-Saharan Africa: Does the capital allocation puzzle matter? *Applied Economics, 0*(0), 1–18. https://doi.org/10.1080/00036846.2023.2212969

Bhimani, A., & Langfield-Smith, K. (2007). Structure, formality and the importance of financial and non-financial information in strategy development and implementation. *Management Accounting Research, 18*(1), 3–31. https://doi.org/10.1016/j.mar.2006.06.005

Cainelli, G., Giannini, V., & Iacobucci, D. (2022). How local geography shapes firm geography. *Entrep. Reg. Dev., 34*, 955–976. https://doi.org/10.1080/08985626.2022.2115559

Caldarelli, G., Arcaute, E., Barthelemy, M., Batty, M., Gershenson, C., Helbing, D., … Fernández-Villacañas, J. L. (2023). The role of complexity for digital twins of cities. *Nature Computational Science, 3*(5). https://doi.org/10.1038/s43588-023-00431-4. article 5.

Casado-Díaz, J. M. (2000). Local labour market areas in Spain: A case study. *Regional Studies, 34*(9), 843–856. https://doi.org/10.1080/00343400020002976

Davies, W. K. D., & Musson, T. C. (1978). Spatial patterns of commuting in South Wales, 1951–1971: A factor analysis definition. *Regional Studies, 12*(3), 353–366. https://doi.org/10.1080/09595237800185291

Derudder, B., & Taylor, P. J. (2018). Central flow theory: Comparative connectivities in the world-city network. *Regional Studies, 52*(8), 1029–1040. https://doi.org/10.1080/00343404.2017.1330538

Derudder, B., Taylor, P. J., Hoyler, M., Ni, P., Liu, X., Zhao, M., … Witlox, F. (2013). Measurement and interpretation of connectivity of Chinese cities in world city network, 2010. *Chinese Geographical Science, 23*(3), 261–273. https://doi.org/10.1007/s11769-013-0604-y

Dorugade, A. V. (2014). New ridge parameters for ridge regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences, 15*, 94–99. https://doi.org/10.1016/j.jaubas.2013.03.005

Elyasi, M., Meybodi, M., Rezvanian, A., & Haeri, M. A. (2016). A fast algorithm for overlapping community detection. In *2016 eighth international conference on Information and Knowledge Technology (IKT)* (pp. 221–226). https://doi.org/10.1109/IKT.2016.7777771

Fang, C., & Zhao, S. (2018). A comparative study of spatiotemporal patterns of urban expansion in six major cities of the Yangtze River Delta from 1980 to 2015. *Ecosystem Health and Sustainability, 4*(4), 95–114. https://doi.org/10.1080/20964129.2018.1469960

Farmer, C. J. Q., & Fotheringham, A. S. (2011). Network-based functional regions. *Environment and Planning A: Economy and Space, 43*(11), 2723–2741. https://doi.org/10.1068/a44136

Feng Jian, Z. Y. (2018). Socio-spatial reconstruction in Beijing (2000−2010). *Geographical Journal, 73*(4), 711–737. https://doi.org/10.11821/dlxb201804010

Fulong, W., & Yeh, A. G.-O. (1999). Urban spatial structure in a transitional economy. *Journal of the American Planning Association, 65*(4), 377–394. https://doi.org/10.1080/01944369908976069

Gehrig, T. (1998). Cities and the geography of financial centres. *CEPR Discussion Papers*, Article 1894 https://ideas.repec.org//p/cpr/ceprdp/1894.html.

Getis, A., & Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. In L. Anselin, & S. J. Rey (Eds.), *Perspectives on spatial data analysis* (pp. 127–145). Springer. https://doi.org/10.1007/978-3-642-01976-0_10.

Griffith, D. A., & Jones, K. G. (1980). Explorations into the relationship between spatial structure and spatial interaction. *Environment and Planning A: Economy and Space, 12*(2), 187–201. https://doi.org/10.1068/a120187

He, X., Cao, Y., & Zhou, C. (2021). Evaluation of polycentric spatial structure in the urban agglomeration of the Pearl River Delta (PRD) based on multi-source big data fusion. *Remote Sensing, 13*(18). https://doi.org/10.3390/rs13183639. Article 18.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoerl, A. E., & Kennard, R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 42*(1), 80–86. https://doi.org/10.2307/1271436

Hu, L., Yang, J., Yang, T., Tu, Y., & Zhu, J. (2020). Urban spatial structure and travel in China. *Journal of Planning Literature, 35*(1), 6–24. https://doi.org/10.1177/0885412219853259

Lee, B. (2007). "Edge" or "edgeless" cities? Urban spatial structure in U.S. metropolitan areas, 1980 to 2000*. *Journal of Regional Science, 47*(3), 479–515. https://doi.org/10.1111/j.1467-9787.2007.00517.x

Lei, Y., Zhou, Y., & Shi, J. (2019). Overlapping communities detection of social network based on hybrid C-means clustering algorithm. *Sustainable Cities and Society, 47*, Article 101436. https://doi.org/10.1016/j.scs.2019.101436

Li, M., Guo, W., Guo, R., He, B., Li, Z., Li, X., … Fan, Y. (2022). Urban network spatial connection and structure in China based on railway passenger flow big data. *Land, 11*(2). https://doi.org/10.3390/land11020225. Article 2.

Lim, M. (2014). Seeing spatially: People, networks and movements in digital and urban spaces. *International Development Planning Review, 36*(1), 51–72. https://doi.org/10.3828/idpr.2014.4

McDonald, G. C. (2009). Ridge regression. *WIREs Computational Statistics, 1*(1), 93–100. https://doi.org/10.1002/wics.14

McMillen, D. P. (2001). Nonparametric employment subcenter identification. *Journal of Urban Economics, 50*(3), 448–473. https://doi.org/10.1006/juec.2001.2228

Milbourne, P., & Kitchen, L. (2014). Rural mobilities: Connecting movement and fixity in rural places. *Journal of Rural Studies, 34*, 326–336. https://doi.org/10.1016/j.jrurstud.2014.01.004

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J., & Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E, 68*(5), Article 056110. https://doi.org/10.1103/PhysRevE.68.056110

Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature, 435*(7043), 814–818. https://doi.org/10.1038/nature03607

Parr, J. B. (1979). Regional economic change and regional spatial structure: Some interrelationships. *Environment and Planning A: Economy and Space, 11*(7), 825–837. https://doi.org/10.1068/a110825

Parr, J. B. (1985). A population-density approach to regional spatial structure. *Urban Studies, 22*(4), 289–303.

Pavesi, A., Vianelli, A., Chirico, N., Bao, Y., Blinkova, O., Belshaw, R., … Karlin, D. (2018). Overlapping genes and the proteins they encode differ significantly in their sequence composition from non-overlapping genes. *PLoS One, 13*(10), Article e0202513. https://doi.org/10.1371/journal.pone.0202513

Peel, L., Larremore, D. B., & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances, 3*(5), Article e1602548. https://doi.org/10.1126/sciadv.1602548

Peixoto, T. P. (2021). Revealing consensus and dissensus between network partitions. *Physical Review X, 11*(2), Article 021003. https://doi.org/10.1103/PhysRevX.11.021003

Priemus, H., & Hall, P. (2004). Multifunctional urban planning of mega-city-regions. *Built Environment, 30*(4), 338–349. https://doi.org/10.2148/benv.30.4.338.57154

Riitters, K. H., O'Neill, R. V., Hunsaker, C. T., Wickham, J. D., Yankee, D. H., Timmins, S. P., … Jackson, B. L. (1995). A factor analysis of landscape pattern and structure metrics. *Landscape Ecology, 10*(1), 23–39. https://doi.org/10.1007/BF00158551

Ross, C. (2012). *Megaregions: Planning for global competitiveness*. Island Press.

Shevky, E. (1972). *The social areas of Los Angeles: Analysis and typology, (New edition)*. Greenwood Press.

Sohn, J. (2005). Are commuting patterns a good indicator of urban spatial structure? *Journal of Transport Geography, 13*(4), 306–317. https://doi.org/10.1016/j.jtrangeo.2004.07.005

Spirin, V., & Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences, 100*(21), 12123–12128. https://doi.org/10.1073/pnas.2032324100

Tang, Y., Li, L., & Liu, X. (2021). State-of-the-art development of complex systems and their simulation methods. *Complex System Modeling and Simulation, 1*(4), 271–290. https://doi.org/10.23919/CSMS.2021.0025

Tchamyou, V. S. (2019). The role of information sharing in modulating the effect of financial access on inequality. *Journal of African Business, 20*(3), 317–338. https://doi.org/10.1080/15228916.2019.1584262

de la Torre, S. R., Kalda, J., Kitt, R., & Engelbrecht, J. (2018). *Detecting overlapping community structure: Estonian network of payments (arXiv:1812.04250)*. arXiv. https://doi.org/10.48550/arXiv.1812.04250.

van der Waerden, P., Bérénos, M., & Wets, G. (2019). Chapter one—Communication and its relationship with digital and physical mobility patterns—A review. In E. Ben-Elia (Ed.), *Vol. 3. Advances in transport policy and planning* (pp. 3–27). Academic Press. https://doi.org/10.1016/bs.atpp.2019.05.001.

Wang, H., Yang, G., & Qin, J. (2020). City centrality, migrants and green Inovation efficiency: Evidence from 106 cities in the Yangtze River Economic Belt of China. *International Journal of Environmental Research and Public Health, 17*(2). https://doi.org/10.3390/ijerph17020652. Article 2.

Wang, Z., Fu, H., Liu, H., & Liao, C. (2023). Urban development sustainability, industrial structure adjustment, and land use efficiency in China. *Sustainable Cities and Society, 89*, Article 104338. https://doi.org/10.1016/j.scs.2022.104338

Xu, R., Yang, G., Qu, Z., Chen, Y., Liu, J., Shang, L., Liu, S., Ge, Y., & Chang, J. (2020). City components–area relationship and diversity pattern: Towards a better understanding of urban structure. *Sustainable Cities and Society, 60*, Article 102272. https://doi.org/10.1016/j.scs.2020.102272

Yin, J., Soliman, A., Yin, D., & Wang, S. (2017). Depicting urban boundaries from a mobility network of spatial interactions: A case study of Great Britain with geo-located twitter data. *International Journal of Geographical Information Science, 31*(7), 1293–1313. https://doi.org/10.1080/13658816.2017.1282615

Yu, Q., Li, W., Yang, D., & Zhang, H. (2020). Mobile phone data in urban commuting: A network community detection-based framework to unveil the spatial structure of commuting demand. *Journal of Advanced Transportation, 2020*, 1–15. https://doi.org/10.1155/2020/8835981

Zhang, P., Zhao, Y., Zhu, X., Cai, Z., Xu, J., & Shi, S. (2020). Spatial structure of urban agglomeration under the impact of high-speed railway construction: Based on the social network analysis. *Sustainable Cities and Society, 62*, Article 102404. https://doi.org/10.1016/j.scs.2020.102404

Zhang, & Thill, J.-C. (2017). Detecting and visualizing cohesive activity-travel patterns: A network analysis approach. *Computers, Environment and Urban Systems, 66*, 117–129. https://doi.org/10.1016/j.compenvurbsys.2017.08.004

Zhang, & Thill, J.-C. (2019). Mesoscale structures in World City networks. *Annals of the American Association of Geographers, 109*(3), 887–908. https://doi.org/10.1080/24694452.2018.1484684

Zhang, W., Fang, C., Zhou, L., & Zhu, J. (2020). Measuring megaregional structure in the Pearl River Delta by mobile phone signaling data: A complex network approach. *Cities, 104*, Article 102809. https://doi.org/10.1016/j.cities.2020.102809

Zhao, M., Derudder, B., & Huang, J. (2017). Examining the transition processes in the Pearl River Delta polycentric mega-city region through the lens of corporate networks. *Cities, 60*, 147–155. https://doi.org/10.1016/j.cities.2016.08.015

Zhen, F., Qin, X., Ye, X., Sun, H., & Luosang, Z. (2019). Analyzing urban development patterns based on the flow analysis method. *Cities, 86*, 178–197. https://doi.org/10.1016/j.cities.2018.09.015

Zheng, Q., Li, Y., Min, T., Chen, F., & Liu, K. (2022). Spatial structures and driving factors of express logistics networks between urban agglomerations in China. *Int. J. Logist. Res. Appl., 0*, 1–21. https://doi.org/10.1080/13675567.2022.2153357

Zhong, B., Hei, Y., Li, H., Rose, T., & Luo, H. (2019). Patent cooperative patterns and development trends of Chinese construction enterprises: A network analysis. *Journal of Civil Engineering and Management, 25*(3). https://doi.org/10.3846/jcem.2019.8137. Article 3.

Zhou, L., & de Vries, W. T. (2022). Collective action for the market-based reform of land element in China: The role of trust. *Land, 11*(6). https://doi.org/10.3390/land11060926. Article 6.

Zhou, L., Zhang, W., Fang, C., Sun, H., & Lin, J. (2020). Actors and network in the marketization of rural collectively-owned commercial construction land (RCOCCL) in China: A pilot case of Langfa, Beijing. *Land Use Policy, 99*, Article 104990. https://doi.org/10.1016/j.landusepol.2020.104990

Zhou, L., Zhou, Y., De Vries, W. T., Liu, Z., & Sun, H. (2024). Collective action dilemmas of sustainable natural resource management: A case study on land marketization in rural China. *Journal of Cleaner Production, 439*, Article 140872. https://doi.org/10.1016/j.jclepro.2024.140872

Zhou, M., Yue, Y., Li, Q., & Wang, D. (2016). Portraying temporal dynamics of urban spatial divisions with Mobile phone positioning data: A complex network approach. *ISPRS International Journal of Geo-Information, 5*(12). https://doi.org/10.3390/ijgi5120240. Article 12.