Tian et al. Computational Urban Science (2024) 4:11
https://doi.org/10.1007/s43762-024-00122-4

Computational Urban Science

ORIGINAL PAPER Open Access

An analysis of agglomeration structure for Beijing, Tianjin, and Hebei based on spatial-temporal big data

Ying Tian[1,3], Changcheng Kan[2], Xiangyu Li[1] and Anrong Dang[1*]

Abstract

The Beijing-Tianjin-Hebei integration plan rose to the status of a national-level strategy in 2014. This paper provides a deep analysis of the Beijing-Tianjin-Hebei area's inter-city commuter big data. This research analyzed the overview of spatial structure, polycentric structure, hierarchical structure and clustering characteristics of the BTH based on network analysis methods. It reveals that the inter-city commuter network exhibits clear polycentric characteristics, with Beijing acting as the central hub. The degree of network correlation between cities in Tianjin and Hebei is notably low, indicating that the flow of people primarily revolves around Beijing, while interactions between other cities remain limited. Therefore, it is necessary to further decentralize Beijing's non-capital core functions. The level of connectedness among the areas surrounding the Bohai Rim is not very high, and it has not developed the coastal advantage. The cooperation could be strengthed among the cities within Bohai Rim. The polycentric structure has initially taken shape, but it exhibits obvious polarization characteristics. It is necessary to strengthen the interaction of talents between cities to form secondary central units in BTH.

Keywords Beijing-Tianjin-Hebei agglomeration, Inter-city commute, Network, Urban agglomeration structure, Multi-center pattern

1 Introduction

In the context of ongoing economic globalization, regional competition, urbanization, and the strengthening of the transmission of resources, cities are increasingly involved in more intensive forms of urban agglomerations through different "flows" to improve their competitiveness. Inter-city commuting is an intuitive reflection of the flow of multiple elements which reflects the degree of connection between cities. Among the five major regional development strategies in China, three are related to urban agglomeration, and the Beijing-Tianjin-Hebei integration strategy is one of them. The analysis of inter-city commuting is of great significance for studying urban network connections and the degree of integration of Beijing, Tianjin, and Hebei agglomeration (Dang et al., 2009). With the progress of advanced communication and transportation technologies, new data, such as mobile phones and the internet, makes it possible to analyze urban space directly through massive data on individual commuting activities (e.g., Dang et al., 2015). Castells (1996) has proposed the "space of flows" which form new city structure. With the support of infrastructures such as intercity railways, airline passenger transportation, the use of the internet, and so on, new spatial systems will replace the traditional site structures, and the material boundaries of cities will become increasingly blurred through the shrinking of spatial and temporal frameworks. As this trend, the functional activities that initially existed in

*Correspondence:
Anrong Dang
danganrong@126.com
[1] School of Architecture, Tsinghua University, Beijing 100084, China
[2] Baidu.com Times Technology (Beijing)Co. Ltd, Beijing 100085, China
[3] CAUPD Beijing Planning &Design Consultants Ltd, Beijing, China

one city will spread to other cities, and the connections between urban agglomerations must be afforded greater importance.

Therefore, how to construct a method that uses spatiotemporal big data to comprehensively evaluate the structural characteristics of urban agglomerations and apply the results to spatial planning has great practical significance. This study attempts to use commuter data and network analysis methods to understand the structural characteristics of urban agglomerations from the four aspects: network's overview structure, multi-center structure, hierarchical structure, and cluster characteristics. Then this research conduct comparative analysis between network analysis results and existing policies and plans of Beijing-Tianjin-Hebei urban agglomeration(BTH) to provide suggestions for the future development of this area.

We consider that our analysis is novel and useful in the following sense. Firstly, it provides a quantitative method for analyzing the overview of agglomeration's structure, multi-central structure, hierarchical structure, and clustering characteristics based on networks analysis using commuter data. Secondly, the structural characteristics of the urban agglomeration are analyzed through the case study of the Beijing-Tianjin-Hebei urban agglomeration; thirdly, for practical applications of commuter data, the network analysis results of the BTH are discussed with existing policies and plans to provide useful suggestions for the future development.

2 Literature review

It is generally believed that commuting reflects functional connections within or between cities (Rain, 1999). In recent years, the relationships between central cities and surrounding cities and towns have been analyzed in terms of work-life relationships reflected by commuting connections (Kloosterman & Lambregts, 2001; Vasanen, 2012; Parr, 2004). These researches have been studied according to two main aspects as following.

The first is the network structure of urban agglomeration based on cross-city commuting connections. The spatial structure, spatial interaction, and degree of connection for "megacities" have become a general approach to regional urban research. For example, De Goei et al. (2010) studied the multi-center structure and urban network development of cities in southeast England through a commuting model. Similarly, Limtanakool et al. (2009) studied the development speed and development characteristics of Dutch urban systems based on multi-year changes in commuting and leisure traffic. Hall and Pain (2006) also analyzed the commuting distributions of eight megacity areas in Europe, revealing the internal structure of "flow" among the megacity areas and urban cyberspace connections. In addition, Niu et al. (2017) identified the workplaces and residences of Shanghai residents using mobile phone signaling data. This allowed them to obtain connection data regarding employment density and commuting patterns, and they analyzed the development of nine new suburban towns in Shanghai from the perspective of the spatial relationship between workplaces and homes.Through an in-depth investigation and visit to the three new towns in Nanjing, such as Dongshan, Xianlin, and Jiangbei, Yang (2010) identified the spatial interaction mechanisms between the main cities and the new cities based on a questionnaire survey. The interaction relationships play a vital role in the spatial development of the new cities. At the same time, the structure and functional layout of the new cities also affect the spatial distribution of commuter flow. Wang (2023) et al. took the central urban area of Nanning as the research object, constructed the commuter flow on the grid-block-street scale, and systematically analyzed the distribution characteristics of commuting space. they explored the influencing factors of commuting volume through geographical detectors combined with multi-source spatio-temporal data.

Second, it is productive to study the development of the regional integration process manifested in terms of commuting patterns. Through a case study of European urban areas, Decoville et al. (2010) analysed the process of spatial integration in ten European cross-border metropolitan regions on the basis of three indicators, relating to flows of cross-border commuters, gross domestic product and the housing market. their work showed that strong economic interactions have an impact on the cross-border integration of communities. Möller et al. (2018) analyzed the difference between cross-border commuting and international commuting in Sweden, and how cross-border mobilities affected spatial integration. They conclude that the border region is characterized by integration through specialization. O'Clery et al. (2019) use commuter data to analyze the integration of labor markets. They develop a network-based model to find the relationship of commuting times and city scale.

Also there is now an increasing momentum in the analysis of networks and flows in cities using data that are collected routinely from digital sensors that pertain to the movements of travelers (Zhong et al., 2014). For instance, statistical analysis of different types of transportation data is being conducted in many cities (Liu et al. 2009, Munizaga & Palma, 2012). In cities such as London, real-time smart card (the 'Oyster-card' in London) data of individual person movements are analyzed to identify the polycentric structure and organization of the central city (Roth et al., 2011). Zhong et al. (2014) developed an integrated method based on network science and spatial analysis to detect changes in urban spatial structure by utilizing new data sources such as smart card data.

Generally speaking, research on dynamic commuter flow has focused on the network structure of urban agglomeration and the degree of connection within or between cities. In this respect, spatial distance is no longer the primary concern. Although there are many studies on urban structure based on commuter flow, most of them focused on identifying the center structure of urban agglomeration networks and judging the degree of regional integration development. There is a lack of systematic analysis methods and practical use of structure analysis using flow data. This study attempts to explore a complex network analysis method based on commuter flows, and uses the network analysis results to evaluate the spatial structure characteristics of urban agglomerations. At the same time, the analysis results are compared with the policies and spatial planning to provide new ideas for urban agglomeration's planning evaluation. This study takes the BTH as an example, treats it as a large region, studied its cross-city commuters' behavior, discovers spatial connection characteristics, and analyzes the implementation of large-scale regional planning. These works are important to the containment of "urban diseases"[1] in Beijing and the integrated development of the BTH.

3 Data and its characteristics

3.1 Study area and data

Our study area is Beijing-Tianjin-Hebei agglomeration(BTH) which have 2 municipalities (Beijing and Tianjin) directly under the Central Government, and 1 province Hebei including 12 prefecture-level cities. The land area of BTH is 218,000 km^2, and the population is more than 110.1 million (2021, year corresponds to the commuter data that follow). BTH is one of China's important economic, cultural and political centers, and its development prospects have attracted much attention. By analyzing the commuter network of BTH, we can obtain conclusions about the agglomeration's structure characteristics on commuting. This allows for a more robust assessment of the spatial structure, which may have broader implications on other large urban agglomerations.

The commuter flow data of 2021, is provided by Baidu Huiyan (https://huiyan.baidu.com/), Data accuracy is 1000*1000 grid. The home and workplace locations that make up the commuter flows in this dataset are inferred using a machine learning method based on millions of users' trajectory data collected from many smartphone apps including Baidu Maps, DiDi, etc. The approach to workplace identification is to analyze the long-term location service requests uploaded by users during working hours. The first step is to identify the locations with the longest occupation hours during daytime hours from Monday to Friday. We can then comprehensively judge whether the identified places are suitable for work using criteria such as urban functions (office spaces, commercial property, etc.) and population density. Our approach to identifying residential locations is similar. Since we can identify workplaces and homes for users, we can also infer the OD data for those users.

Table 1 Attributes description of commuter flow data

Field	Type	Description
ID	String	Unique grid id(home/work)
PrefectureLiving	String	Prefecture's name of residence
CountyLiving		County's name of Residence
Home_lng	Float	Longitude of residential grid center
Home_lat	Float	Latitude of residential grid center
PrefectureWorking	String	Prefecture's name of workplace
CountyWorking	String	County's name of workplace
Work_lng	Float	Longitude of workplace grid center
Work_lat	Float	Latitude of workplace grid center
Popu	Int	cross-city commuting population between residence and workplace

Commuting OD data are mined based on artificial intelligence to obtain the user's residence and work place, with an accuracy rate of more than 90%, and all steps of data processing are anonymized, and the output does not involve individual privacy, so this data product is highly reliable. (Lu et al., 2020) Each row of the data represents a pair of commuter flow, including fields such as the OD grids' ID, coordinates, the name of prefecture-level city, the name of county, and the corresponding number of commuters, as shown in Table 1. There are a total of 962,840 pieces of original data. Statistics are based on the names of counties a total of 22,597 pieces of data were processed, involving 8,992,380 commuters.

The inter-city commuting flow data were selected and defined as the person's residence and work place are across counties. 408,100 cross-city commuting populations were finally identified, forming a spatial interaction network with 199 nodes and 22,447 edges. This network is relatively dense.

3.2 Characteristics of inter-city commuter data

Work-life data is a typical example of spatiotemporal "big data" reflecting the degree of connection between people's residences and workplaces. The degree of

[1] Big cities in China are on the brink of a major shortfall in resources and infrastructure capacity, under a problem termed "urban diseases".

Table 2 Correlation of factors related to urban agglomeration networks and commuting data

	urban agglomeration network	Cross-city commuting data
Related factors	Centrality, Degree(in/out)etc.	Commuting distance, Commuting time, Density of Commuters, Inflow and outflow ratios, etc.
Spatial scale	Prefecture-level or county-level	Relationships between life and work
Research goal	Network structure and characteristics	Relationships between cities, work-life balance, etc.

Table 3 Part of summary data

Name	ID	Outflow Density	Inflow Density	Ratio of Inflow and Outflow	Outflow mean distance	Inflow mean distance	Weighted degree	permanent population 2021(Thousand)	Capita GDP
Beijing Yanqing	11	0.050174	0.039737	0.791984	25.018599	20.727778	32,098	35.7	5.470308
Beijing Xicheng	202	0.162821	0.356371	2.188729	88.85429	229.629683	590,322	113.7	44.03958
Beijing Tongzhou	5	0.185435	0.103103	0.556006	276.094465	182.499763	483,301	167.5	6.323582
Beijing Shunyi	6	0.120345	0.084793	0.704583	141.086276	114.21666	251,910	122.8	16.22883

connection between cities can be reflected through the work-life transits and, therefore, it can show the structure of urban agglomeration. Cross-city work-life information has the following characteristics:

First, in contrast with cross-city migration or regional flow data, cross-city commuting big data accurately reflects the cross-city flow of people with repeatability and definite purpose over a discrete period of time, which has the characteristic of stability. Residential and work attributes in the data highlight the different needs of people for life and work functions in the region. It is an intuitive expression of mutual communication and coordination among urban clusters. Commuter data can, to some extent, represent the stable urban agglomeration structure.

Second, cross-city work-life data represents a network composed of the inflow and outflow of people working and living cross-city. This cross-city communication network has the characteristics of a shared data network. As nodes of the data network, cities, and towns usually have attributes such as in-degree, out-degree, centrality, etc. As edges between nodes, connection relationships usually have characteristics of differences in terms of weight, and these factors are often introduced into the analysis and research of data networks (Büttner et al., 2015; De Montis et al., 2007).

Third, as the data networks in a real geographic coordinate system, the spatial distances between nodes have practical significance. Commuting time based on real road distances can reflect the cost of cross-city commuting, which is a significant factor for work-life commuting networks (Table 2).

3.3 Data preparation

Based on the commuter flow network, with counties as statistical units, the inflow density, outflow density, inflow-outflow ratio, inflow mean distance, outflow mean distance, weighted centrality degree of each unit are calculated respectively, as well as the number of permanent residents and capita GDP. Part of the summery results are showed in Table 3. Visualize the above table data, as shown in Fig. 1.

4 Methods

Considering the development needs of urban agglomerations and the characteristics of relevant data, our approach includes four steps (Fig. 2):

(1) Delimiting spatial units. It is essential to select appropriate spatial units based on research objectives and needs and to divide other multi-source data, such as cross-city work-life data, across each spatial unit. The division of units should fully ensure comparability. In analyzing the Beijing-Tianjin-Hebei agglomeration, this paper mainly selects each city's district and county administrative divisions as the spatial units. In order to better reflect internal relationships, we selected district and county units with finer spatial granularity where the statistical data were easy to

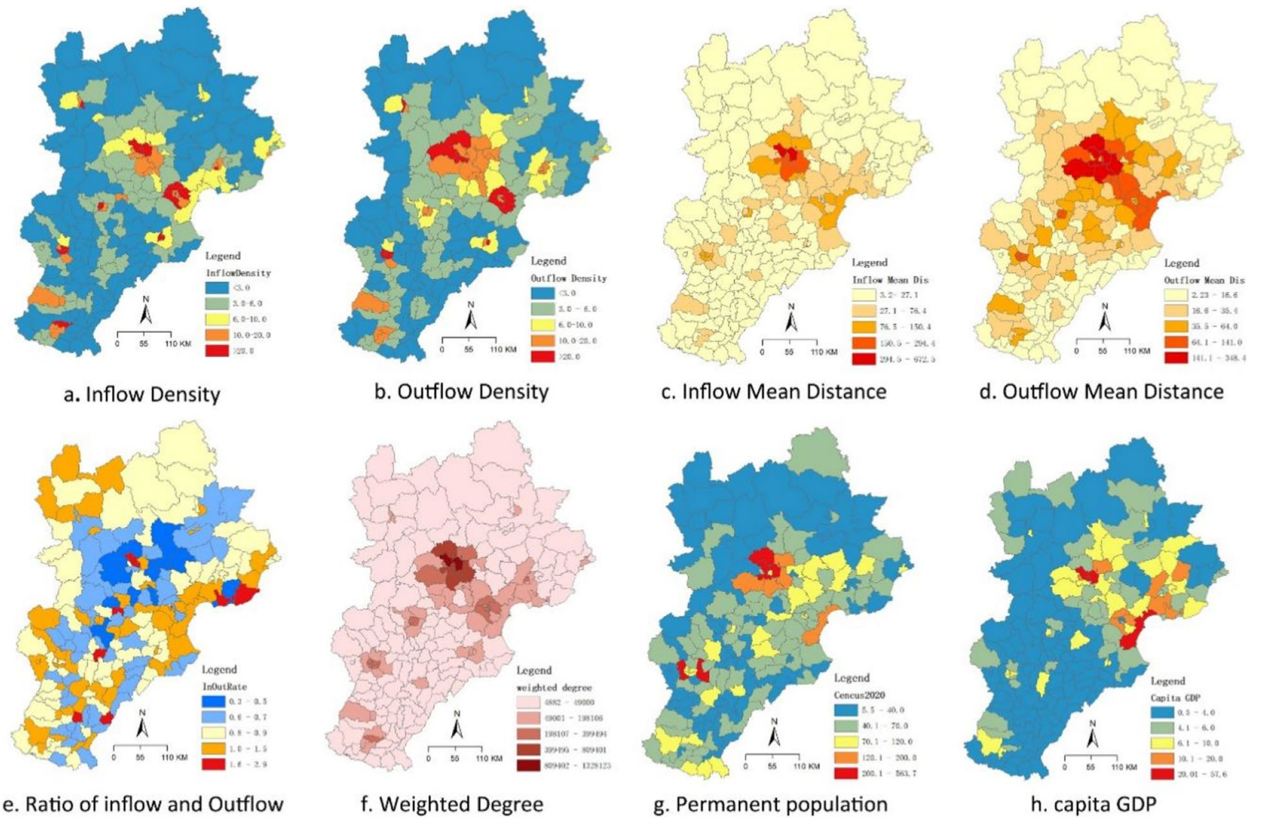

Fig. 1 Distribution of each factor

obtain. This study identified the workplaces and living spaces of individuals crossing districts and counties as cross-city commuting populations.

2) Selecting the analysis factors. It is necessary to summarize the analysis factors selected in previous cross-city commuting studies and to select data with relevant factors in combination with the characteristics of inter-city commuter data and network structure indices. Finally, this study utilizes the total inflow, total outflow, input distance, output distance, and centrality as preliminary analysis factors. We define the migration behavior of those living locally and working outside as outflow and the migration behavior of those living elsewhere and working locally as inflow. The concept of our analysis factors is as follows: the total inflow and outflow, respectively, represent the total input population of a spatial unit (living in other spatial units but working locally) and the total output population (living in a local spatial unit but working elsewhere). The input and output distances respectively represent the linear distance of starting point of the migrant input to a spatial unit and the linear distance of starting point for the migrant output from a spatial unit. This factor represents the influence range of the input and output of spatial units across the city. Outflow and inflow ratios refer to the ratio of the total outflow to the total inflow for a spatial unit, indicating that the spatial unit is preferred by those living or working cross-city. Centrality (i.e. the degree centrality in the context of network analysis) indicates the proportion of total inputs and outputs in relation to a spatial unit to the total number of connections for the whole region. This factor represents the importance of particular spatial units within the overall migration network.

Considering that the forms of the spatial units selected in this study are not unified, however, the scale of the unit itself has a great impact on the factor level. Further standardization can be used to obtain six factors: inflow density, outflow density, outflow and inflow ratio, weighted centrality, inflow mean distance, and outflow mean distance.

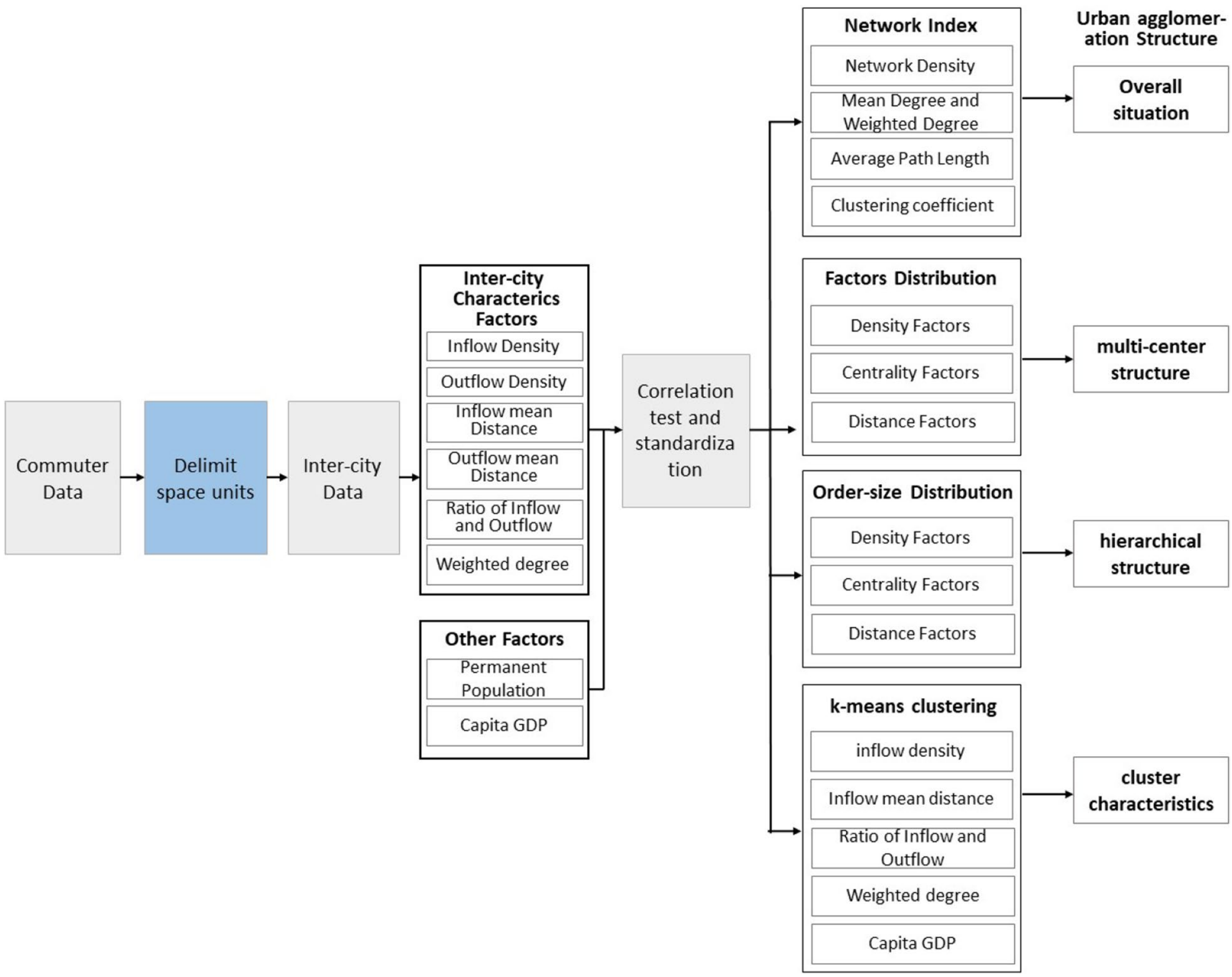

Fig. 2 Research framework

In addition, it is necessary to consider the factors reflecting the other characteristics of spatial units, generally referring to factors such as economy, culture, and public services. Therefore, it is crucial to select economic developments and the other factors reflecting spatial units to reflect the characteristics of the spatial units at an overall level.

(3) By calculating selected factors to analyze the structure of agglomeration, including overall situation, polycentric features, hierarchical and clustering features.

(4) We compare the network analysis result with the spatial planning and policies to promote suggestions for urban agglomeration development.

5 Results

5.1 Overview of spatial structure

We analyzed network structures using the social network analysis software Gephi (results shown in Table 4). Through the general analysis of the network structure, we conclude that the scale of the Beijing-Tianjin-Hebei commuting network is 22,447, and the network density is 0.57. The Beijing-Tianjin-Hebei area is a spatial region with 199 nodes, and the maximum network scale that can be achieved is 39,601, indicating that the network density of the Beijing-Tianjin-Hebei commuter flow is relatively high.

The centrality of the network for Beijing-Tianjin-Hebei commuter flow is 112.8, which means that a point has a connection with an average of 113 other points in the network. This indicates that the interaction between networks is greater.

Table 4 Network index of commuter flow

Index	value	Description
Network Density	0.57	An important indicator to measure the relationship between nodes
Mean network degree	112.8	The average of centrality degree of all nodes, reflecting the average connectivity of the nodes
Mean network weighted degree	45,180	The average of weighted degree of all nodes, reflects the node centrality of the network
average path length	1.42	The average of shortest distance between two nodes, reflects the average dispersion of the network and the "small world effect"
Clustering coefficient	0.73	Reflects the clustering characteristics of the network and measures the degree of close connection of nodes in the network

The mean in-degree centrality and out-degree centrality are 113.1 and 146.6 respectively, and their coefficients of variation are 0.29 and 0.46. The difference of in-degree centrality is relatively small, indicating that the network node traffic outflow is relatively balanced. The coefficient of variation of out-degree centrality is high, indicating that the outflow of each node in the network is obviously concentrated, and presenting characteristics of polarization.

Weighted centrality degree reflects the importance of a node in the network, and the weighted average degree of the commuter network is 45,180. The weighted in-degree centrality and weighted out-degree centrality are 44,045 and 60,358 respectively, and their coefficients of variation are 1.93 and 1.94 respectively. The weighted in-degree centrality reflects the importance of a node as a work point in the commuter network, and the weighted out-degree centrality reflects the importance of a node as a residential point in the commuter network. The coefficients of variation indicate there are big difference of volume among nodes as a cross-city work or residential place.

The average path length of the commuter network is 1.42, which means most nodes are connected through 1 to 2 intermediate nodes. Therefore, commuter connections are mainly short-distance node connections, reflecting structural characteristics of high accessibility and small separation between nodes.

The clustering coefficient of the commuter network is 0.73[2]. The larger the clustering coefficient, the more likely the network is to contain a small group structure. This indicator shows that the commuter network in Beijing, Tianjin and Hebei exhibits obvious clustering characteristics.

5.2 Formation of multi-center structure

The Beijing-Tianjin-Hebei urban agglomeration presents multi-center and two-way functional connections. The balanced distribution of functional connections between cities represents balanced development. If the two-way functional connection between cities is relatively balanced, it can be regarded as functional equalization. Multi-centered networks, while obviously imbalanced, can be regarded as single flows of factors and single centers of functions (Burger et al., 2011).

First, Beijing-Tianjin-Hebei urban agglomeration has gradually formed a structure with Beijing as the main core and prefecture-level cities such as Tianjin, Shijiazhuang, and Baoding as secondary cores. This view can be proved by the following three aspects of analysis.

(1) From the perspective of density distribution, areas with high cross-city inflow and outflow population density show high spatial agglomeration and polycentric characteristics. High-density areas are concentrated in the central urban areas of prefecture-level cities such as Beijing, Tianjin, Shijiazhuang, Zhangjiakou, Qinhuangdao, Tangshan, Baoding, etc. (Fig. 2a and b).
(2) From the total number of cross-city work-living functional connections among counties and the weighted central degree of counties within the Beijing-Tianjin-Hebei area, the structure of urban agglomeration can be examined. Cross-city commuting with a high level of flow exists in a small number of cities. Beijing, Tianjin, and Shijiazhuang are the areas with the most intensive cross-city flow, and a multi-center structure is forming there. Cross-city commuting within Beijing is particularly obvious and, secondarily, there is greater cross-district commuting population between the central urban areas of Tianjin, Shijiazhuang, Baoding, Handan (Fig. 3). In addition, the connection between Beijing and Langfang, Shijiazhuang and Cangzhou is obviously closer than others. Weighted in-degree centrality reflects the importance of a node as a work place in the commuting network. The top three weighted in-degree centralities are Beijing Chaoyang District, Beijing Xicheng District,

[2] The value range is between 0 and 1. The closer the value is to 1, the stronger the clustering between nodes. The closer the value is to 0, the weaker the clustering between nodes.

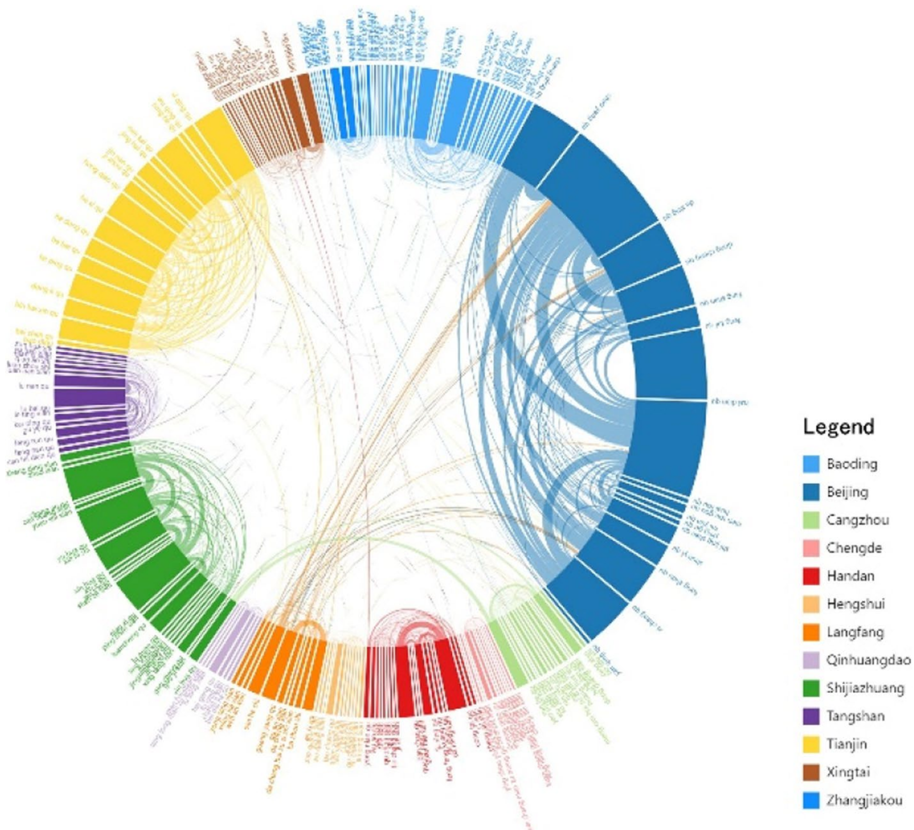

Fig. 3 Inter-city commute relation

and Beijing Haidian District. They are the three districts and counties that attract the largest number of jobs, with weighted in-degree values accounting for 9.23%, 4.62%, and 4.20 respectively. Weighted out-degree centrality reflects the importance of a node as a residence. The top three are Beijing Haidian District, Beijing Chaoyang District, and Beijing Tongzhou District. It shows that these nodes play an important role in resource allocation and are mainly concentrated in the central area of Beijing (Fig. 4). But the network structure is not obvious. There is a low level of connection between sub-centers, such as between the counties of Beijing and either Tianjin or Hebei.

(3) From the view of commuting distance. The average inter-city commuting distance is mainly within 40 km (please see Fig. 5). The distance and quantity of moves in cross-city commuting a significant negative correlation with a noticeable apparent downward trend. The moving distance for cross-city commutes is less than 40 km, and cross-city working-living tends to carry out short-distance travel activities from the overall level. Short-distance moving behaviors with distances less than 20 km mainly occur between urban districts. Taking Beijing's Haidian, Chaoyang, Xi'cheng, Fangshan, Fengtai, and Tongzhou as examples, the short-distance cross-city commutes between districts and counties within the same prefecture-level city are obvious (Fig. 5). Long-distance moves of greater than the standard quartile within the full sample mainly occur between the cities of Handan and Tianjin or Beijing, indicating that, as the core city of the Beijing-Tianjin-Hebei area, Beijing, Tianjin attracts more people from more distant cities.

The average value of the inflow and outflow ratio for cities with large cross-city commuting volumes in Beijing, Tianjin, and Hebei is 0.96 (please see Fig. 2e). The two-way commuting between other cities is relatively similar. The commuting inflow is greater than the

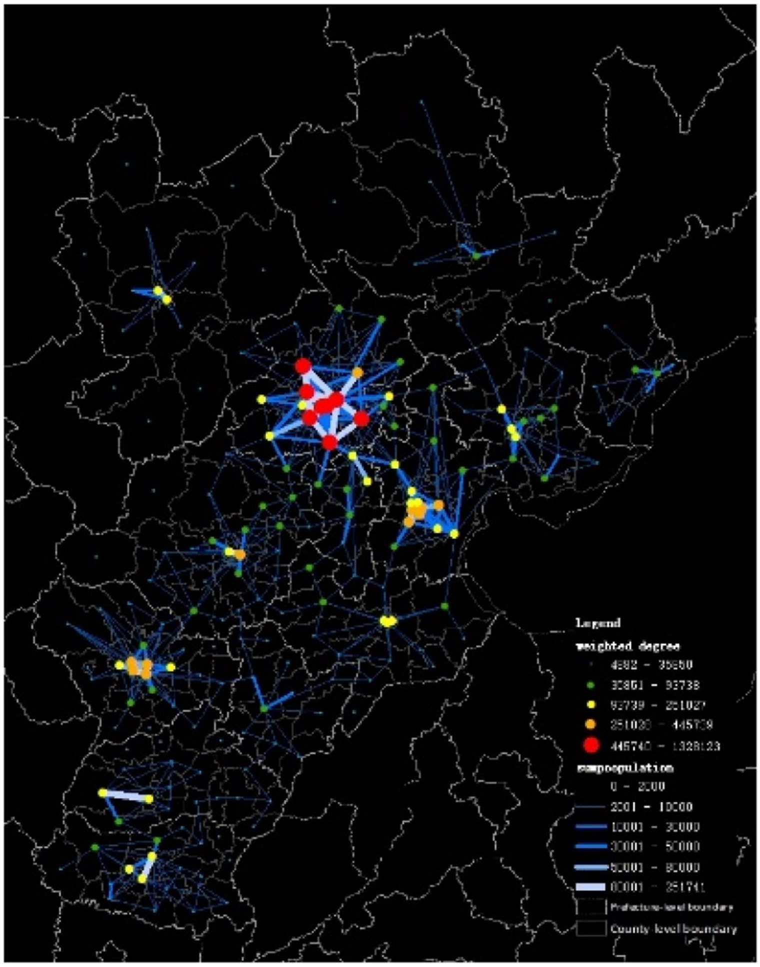

Fig. 4 Weighted degree and total commute population

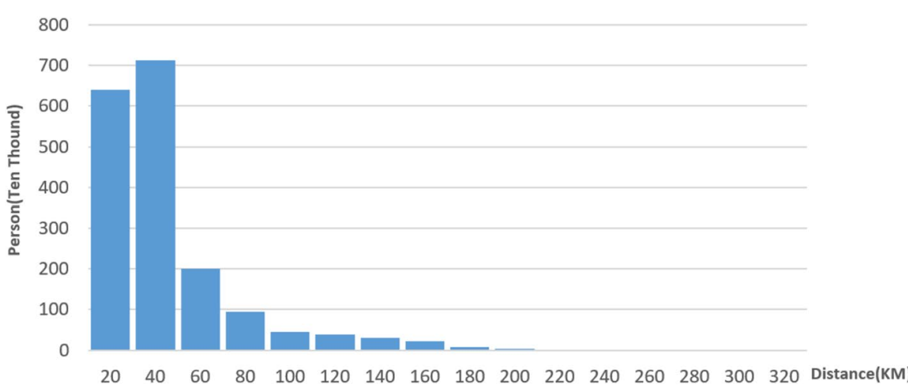

Fig. 5 Histogram of commuting distance between Beijing-Tianjin-Hebei cities

commuting outflow, indicating that the cities with close cross-city commuting connections are not providing either employment or residence alone but rather there are two-way resource exchanges between cities. The inflow and outflow ratio for Dongcheng District, Xicheng District, etc. in Beijing, Rongcheng City in Baoding, and Caofeidian City in Tangshan is greater than 1.5, indicating that there is more of an inflow of people in these areas and more people tend to work there. Outflow predominates in Changping District, Mentougou District,

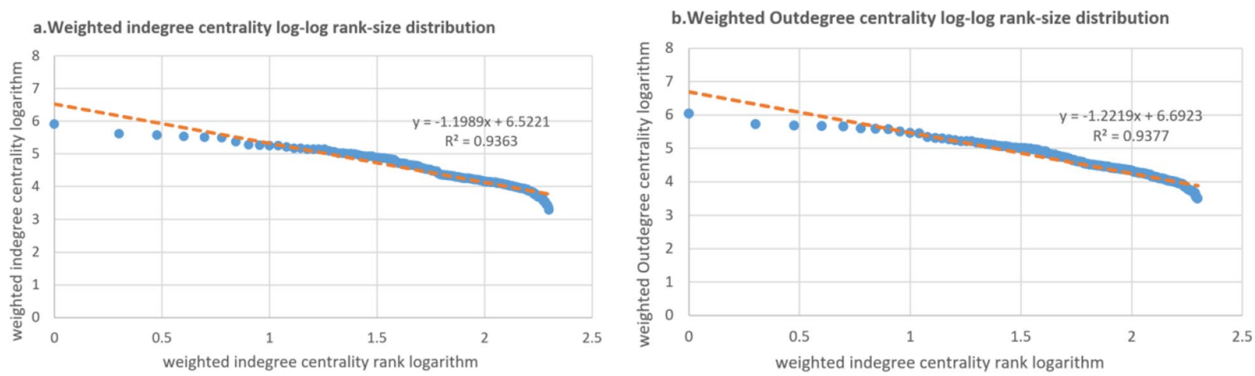

Fig. 6 Weighted in-degree and out-degree logarithmic order-size distribution of commuting network

Fangshan District, Tongzhou District in Beijing, Dingxing City in Baoding, and Gu'an City in Langfang, indicating that more people tend to live here. Large cities mainly provide more employment opportunities. For example, Beijing and Shijiazhuang is relatively higher which means they provide more working opportunities. At the same time, Langfang and Cangzhou are dominated by residential use, with a greater outflow of population than inflow (Fig. 2e).

5.3 The network is highly hierarchical and exhibits non-homogeneous development

In order to analyze the hierarchical structure of the commuter network, the weighted in-degree centrality and weighted out-degree centrality of each node in the network and their order were taken as logarithms. Regression analysis of rank-size for weighted in-degree centrality of commuter networks was performed. The scale goodness of fit (R^2) is 0.94, and the value of |a| is 1.20 (Fig. 6-a); the scale goodness of fit (R^2) of out-degree rank-size regression analysis is 0.94, | The values of a| are 1.22 respectively (Fig. 6-b). It can be seen that |a| is both greater than 1, indicating that the weighted in-degree and out-degree centrality distributions of nodes in the network have obvious hierarchical characteristics. The in-degree centrality of core nodes plays an important role in the commuter network. The homogenization phenomenon is obvious.

5.4 The network has obvious clustering characteristics and presents a small community structure

Combined with the data characteristics of cross-city commuting behavior and the structural characteristics of network data, this paper analyzes factors such as inflow density, outflow density, mean inflow distance, mean outflow distance, weighted degree of centrality, ratio of inflow to outflow, and capital GDP. In testing the correlation between factors including, inflow density, and outflow density, mean inflow distance and mean outflow distance is, and outflow density, mean inflow distance, and mean outflow distance all have strong relationships, so only inflow factors are retained. The distribution of the five retained factors is shown in Table 5.

The cluster analysis of 199 spatial units in the Beijing-Tianjin-Hebei agglomeration was done using the k-means clustering method. We found that when clustering was fixed to 4 categories, the aggregation coefficient tends to be stable, and the characteristics of each category were the most clear-cut. Therefore, this study selects four categories for classification. The classification results are named A-D. The values of each classification cluster are shown in Table 5, and the distribution status is shown in Fig. 7.

Class A is the area to be developed in the Beijing-Tianjin-Hebei agglomeration, accounting for most of the spatial units. The per capita GDP in this unit is relatively low and overall, the region is underdeveloped. The proportion of cross-city inflow is among the lowest, and the ratio of inflow to outflow is relatively balanced. In addition, the weighted centrality level is extremely low. This class's level of exchange with the surrounding units is limited, and cross-city commuting is also limited to short distances. This context cannot attract people from distant units to work or live here.

Table 5 Final cluster centers

	Cluster			
	A	B	C	D
Inflow Density	0.04	0.23	0.18	0.20
Capital GDP(ten thousand RMB/person)	4.08	22.49	9.19	17.56
Inflow Mean Distance(KM)	16.50	600.50	77.70	215.24
The ratio of Inflow and Outflow	0.93	1.76	1.04	1.15
weighted degree	34,451	1,210,152	239,226	575,390

Class B and D units belong to the core areas, which are obviously concentrated in Beijing. The per capita GDP of this unit is significantly higher than that of other units. It has a developed economy, high weighted centrality, and long migration inflow and outflow distances, which shows that people are attracted from distant spatial units to live or work here. Theimmigration into Class B units is significantly greater than emigration, indicating that there are more jobs to attract more people living in the surrounding areas to work here. As commercial and scientific research is concentrated in areas in the country, Haidian and Chaoyang also exhibit these features. Immigration and emigration for Class D are relatively balanced, showing that people are attracted to live and work here by virtue of the class's advantages.

Class C has a high level of connection to be developed in the Beijing-Tianjin-Hebei area, mainly in terms of a prefecture-level downtown area and a new national area, Xiong'an. The per capita GDP here is higher than surrounding units, and the cross-city work-life input and output density is higher. The average distance between immigration and emigration is mostly at the middle and lower levels, which are only capable of attracting the surrounding population with a low-weight centrality. This class belongs to the secondary development centers in urban agglomerations and needs to be further enhanced in terms of connections with surrounding regions.

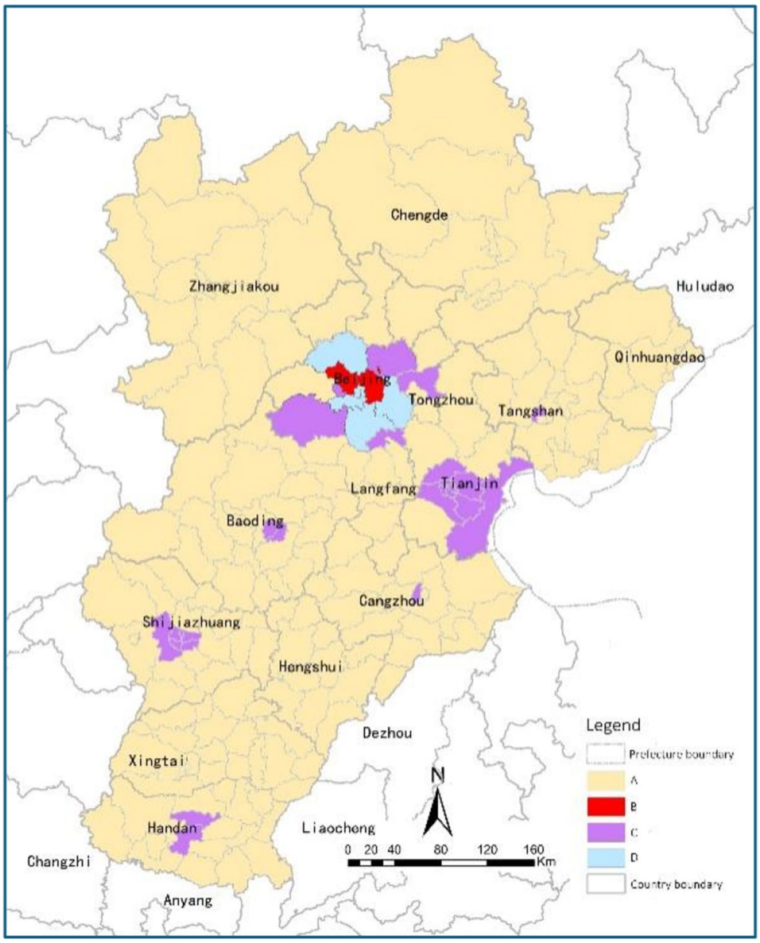

Fig. 7 Cluster divisions

6 Discussion and conclusions

The coordinated development of Beijing, Tianjin and Hebei is a major regional strategy deployed and promoted by Chinese General Secretary Xi Jinping. It first occurred in the government work report of March 5, 2014. The purpose is to strengthen economic cooperation around the Bohai and Beijing-Tianjin-Hebei areas. The Political Bureau of the CPC Central Committee held a meeting on April 30, 2015, to review and adopt the planning outline for the coordinated development of Beijing,

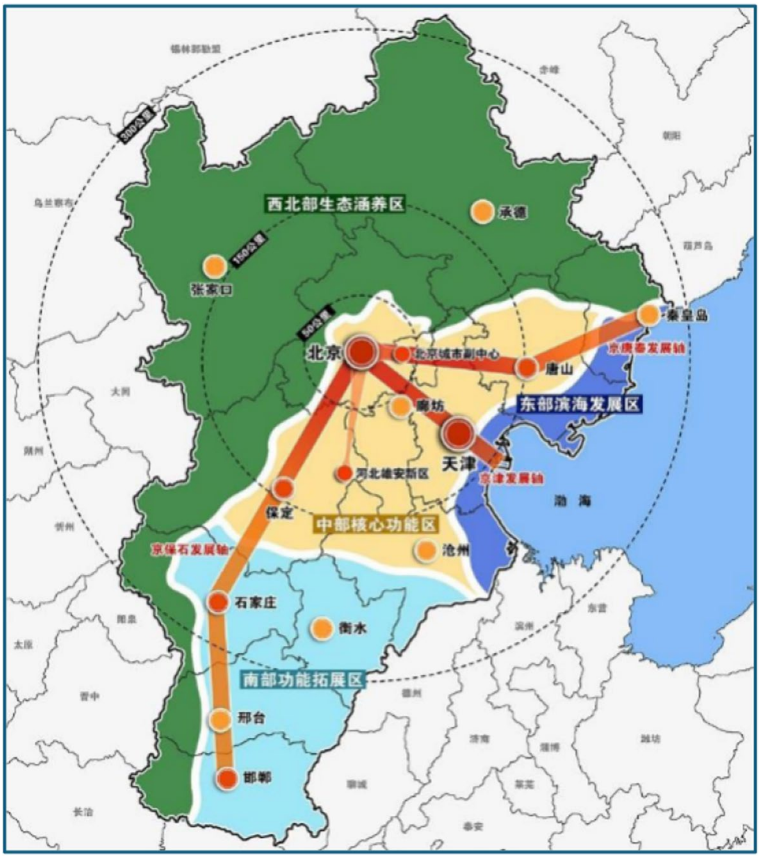

Fig. 8 Policy spatial structure

Tianjin, and Hebei. In this outline, they adopted the concepts of "one core, two cities, three axes, four districts, and multiple nodes" in supporting the spatial layout of the Beijing-Tianjin-Hebei area. The "one core" refers to Beijing, and the "two cities" refers to Beijing and Tianjin, which are the main engines for the coordinated development of Beijing, Tianjin, and Hebei. "Multiple nodes" includes regional central cities such as Shijiazhuang and node cities such as Qinhuangdao. This approach makes the comparison between planned spatial structures and the network formed by the connection of workplaces and residences, and in discussing structural optimization, we provide the following suggestions for Beijing-Tianjin-Hebei urban agglomeration:

6.1 "One core and two cites" has been formed, though non-capital functions need further evacuation

According to the analysis of the existing planning, Beijing is the core city of Beijing-Tianjin-Hebei Urban Agglomeration, and the "Beijing Overall Plan (2016–2035)" is also determined to build a world-class urban agglomeration with the capital as its core (please see Figs. 8 and 9). Through network analysis, we have found that Beijing has an obvious core position in the Beijing-Tianjin-Hebei area, which has the most intensive workplace-home population exchange and the widest radiation range.

Attention should be paid that the coordinated development of Beijing-Tianjin-Hebei is to relieve Beijing's non-capital functions. Focusing on solving the problem of "big city disease" in Beijing, a plan and policy system have formed for alleviation, and promoted the construction of the first batch of universities, hospitals, and state-owned enterprise headquarters in Xiong'an New Area. At the same time, we have also issued regulations on household registration, investment, and alleviation to Xiongan's policies on wages and income of employees or companies. On the other hand, the construction of Beijing's sub-center city was promoted, and Tongzhou will build a city sub-center administrative office area with supporting education, medical, cultural and other facilities. It is expected to take over and relieve some of the functions of Beijing. It can be seen from the commuter network structure that Xiongan has developed rapidly since its construction in 2019, and has formed a regional commuter node by 2021. Xiong'an includes Xiongxian, Anxin, and Rongcheng, with weighted centralities of

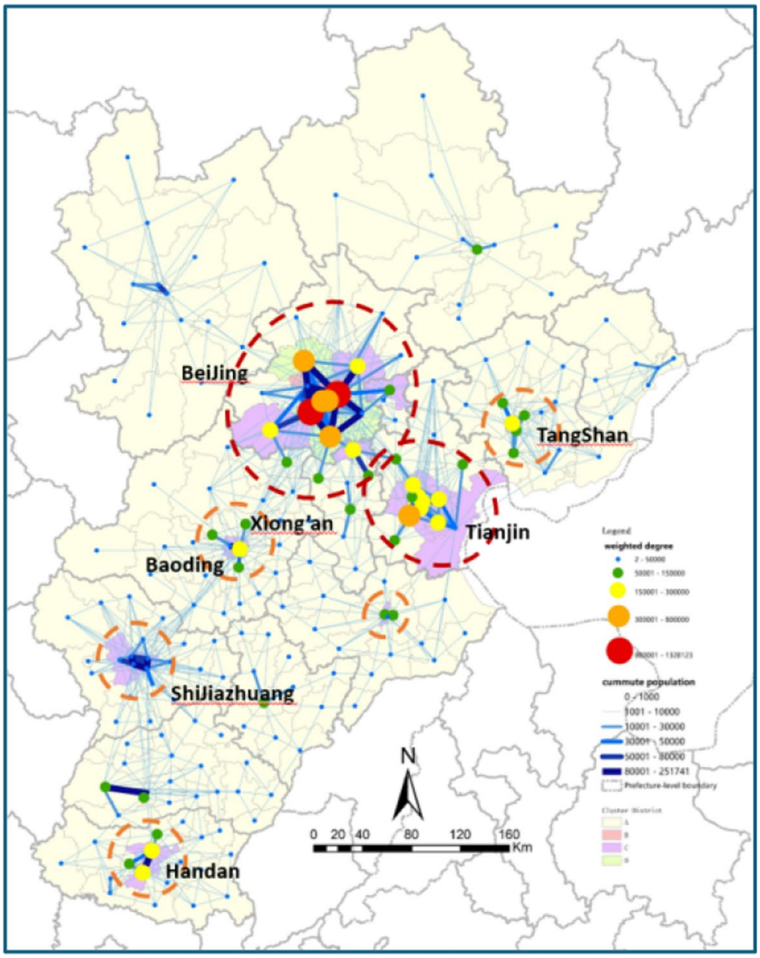

Fig. 9 Commuter network structure

32,486, 32,401 and 48,051, both in the top 100 among 199 counties. The Tongzhou Sub-center began planning and construction in 2012. After nearly ten years of development, it gained some achievements. The weighted centrality is 483,301, ranking seventh in the whole network, with an outflow weighted centrality of 310,604 and an inflow weighted centrality of 172,697. It ranks higher among the agglomeration, tending to provide rensidences. In addition, we also found from the results of our cluster analysis that the most important places providing work are Haidian and Chaoyang of Beijing. To better balance the structure of urban agglomeration, we should strengthen Tianjin and Hebei's capability of undertaking non-capital functions for Beijing, creating more employment opportunities.

Tianjin has also basically taken on this pattern, however, compared with Beijing, it still has defects in terms of scale and connection scope, and its connection with Beijing is weaker. Tianjin's level of integration must be further improved, as well as the degree of cooperation among its public services and industries.

6.2 The strength and breadth of ties in the Bohai Rim area must be improved

Apart from Beijing serving as the core of the Beijing-Tianjin-Hebei area, the coordinated development of Bohai Rim is also particularly important. The Bohai Rim area is the subject of future planning, however, from the perspective of existing network connections, it mainly serves as an axis connection between the Tianjin urban area and Beijing (please see Fig. 10). There are no network connections along the Bohai Rim, and there is a low level of personnel exchange. In contrast, within the Bohai Rim, there is no city with a more significantly weighted center in the Beijing-Tianjin-Hebei, indicating that the region also lacks regional nodes. Therefore, from a comprehensive perspective, this coastal area has not formed a sound network structure, and the cooperative

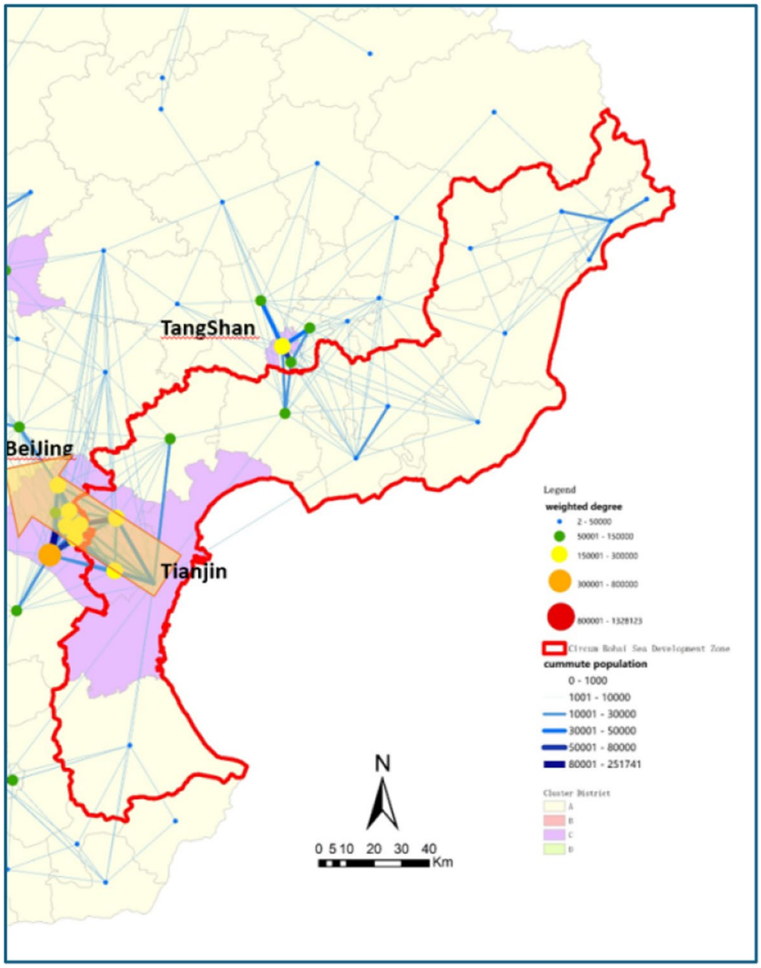

Fig. 10 Bohai rim structure

connections within the Bohai Rim in the Beijing-Tianjin-Hebei region should be strengthened.

6.3 The polycentric structure has initially taken shape, but exchange of talents between regional central cities should strengthen

The network density of 0.57 for the Beijing-Tianjin-Hebei is relatively high, which stems from the preliminary formation of its network structure. The regional personnel flow between nodes is also two-way and relatively balanced. However, the position of certain regions' central cities in urban agglomerations is not prominent, and the cities in Hebei have a relatively lower level of flattening. Except Beijing and Tianjin, most of the other urban areas have a low weighted centrality in the network. Regarding planning, cities like Shijiazhuang, Tanshan, Baoding, Handan, etc., are the regional centers, and cities like Qinhuangdao, Cangzhou, Xingtai, etc., are node cities. Regarding our network results, Baoding, including Xiong'an, increasingly demonstrates its position in the urban agglomeration, and Tangshan and Handan have also formed a connected network around its central urban areas. As the provincial capital city, Shijiazhuang did not match its weighted centrality in the urban agglomeration, which should depend on the Beijing-Baoding-Shijiazhuang Development Axis to accelerate cooperation with surrounding areas, especially with Beijing. Cities like Qinhuangdao, Cangzhou, and Xingtai, which have not formed prominent central units in the urban agglomeration, should enhance the attractiveness of the city through improving the urban environment, the quality of facilities, and cultural construction, and speed up talent cooperation with surrounding cities.

This study uses the spatial-temporal big data of commuter flows, selects appropriate factors for network analysis, analyzed the overall structure, polycentric structure, hierarchical structure and clustering characteristics of the urban agglomeration. Then we compared it with the existing Beijing-Tianjin-Hebei development policies and planning plans. The analysis found three views. First, although the "one core, two cities" pattern has been basically formed, and the decentralization of non-capital functions has achieved certain results, Beijing is still the absolute core of the region in terms of employment and residence, and its functions of non-capital need to be further decentralized. Second, as the Bohai Rim region is an important economic coordinated development area in the country, the intensity and breadth of talent connections in its cities need to be improved. Third, although the polycentric structure of Beijing-Tianjin-Hebei has been initially formed, it is necessary to strengthen the interaction of talents between regional central cities and node cities to form secondary central units. However, this study is limited to the use of commuting flows and the discussion only at the urban agglomeration scale. There are research shows obvious differences in the regional spatial structure of the same flow at different scales (Taylor et al., 2008). Also, the regional centrality represented by the flow of people, capital or traffic survey data in the same region is different (Hall & Pain, 2006). In the future, different types of "flow" data are analyzed at different research scales, the results will remain to be discussed in depth.

Acknowledgements
The authors acknowledge support from the National Natural Science Foundation of China (Grant No. 52130804). All authors are thankful to the editor and anonymous reviewers for their constructive suggestions and comments.

Authors' contributions
Anrong Dang conceived and designed the study; Ying Tian performed the data analysis and took the lead in writing the manuscript; Changheng Kan contributed data and assisted in data analysis; Xiangyu Li polished the writing and participated in the data analysis. All authors read and approved the final manuscript.

Funding
This work was supported by the Key Project of the National Natural Science Foundation of China (Grant No. 52130804).

Availability of data and materials
The data that support this study are available from the Time Series Database of Baidu Huiyan. Due to privacy restrictions, data are made available upon request via https://huiyan.baidu.com/.

Declarations

Competing interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 15 February 2024 Revised: 3 April 2024 Accepted: 11 April 2024
Published online: 18 April 2024

References

Burger, M., De Goei, B., Van Der Laan, L., & Huisman, F. (2011). Heterogeneous development of metropolitan spatial structure: Evidence from commuting patterns in English and Welsh city-regions, 1981–2001. *Cities,28*(2), 160–170. https://doi.org/10.1016/j.cities.2010.11.006

Büttner, K., Scheffler, K., Czycholl, I., & Krieter, J. (2015). Social network analysis - centrality parameters and individual network positions of agonistic behavior in pigs over three different age levels. *Springerplus,4*(1), 185. https://doi.org/10.1186/s40064-015-0963-1

Castells, M. (2009). The Space of Flows. In The Rise of the Network Society, M. Castells (Ed.). https://doi.org/10.1002/9781444319514.ch6 .

Dang, A., Jian, X., Biao, T., Juan, L., & Fang, Q. (2015). Research progress of the application of big data in China's urban planning. *China City Planning Review,1*(24), 24–30.

Dang, A., Li, Y., & Qin, X. (2009). A study on spatial growth of Beijing's central urban area from 1950s to 2005 supported by spatial information technology. *China City Planning Review,18*(03), 17–21.

De Goei, B., Burger, M., Van Oort, F., & Kitson, M. (2010). Functional polycentrism and urban network development in the greater South East, United Kingdom: Evidence from commuting patterns, 1981–2001. *Regional Studies,44*(9), 1149–1170. https://doi.org/10.1080/00343400903365102

De Montis, A., Barthélemy, M., Chessa, A., & Vespignani, A. (2007). The structure of interurban traffic: A weighted network analysis. *Environment and Planning B: Planning and Design,34*(5), 905–924. https://doi.org/10.1068/b32128

Decoville, A., Durand, F., Sohn, C., & Walter, O. (2010). Spatial integration in European cross-border metropolitan regions: A comparative approach. *Social Science Research Network*. https://doi.org/10.2139/ssrn.1803992

Hall, P., & Pain, K. (2006). The Polycentric Metropolis: Learning from Mega-City Regions in Europe (1st ed.). Routledge. https://doi.org/10.4324/9781849773911 .

Kloosterman, R. C., & Lambregts, B. (2001). Clustering of economic activities in polycentric urban regions: The case of the Randstad. *Urban Studies,38*(4), 717–732. https://doi.org/10.1080/00420980120035303

Limtanakool, N., Schwanen, T., & Dijst, M. (2009). Developments in the Dutch urban system on the basis of flows. *Regional Studies,43*(2), 179–196. https://doi.org/10.1080/00343400701808832

Liu, L., Hou, A., Biderman, A. (2009). Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen, *12th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 1–6.

Lu, Q., Kan, C. C., & Wei, X. (2020). Commuting circle in twin cites of Tianjin based on location-based service data. *Urban Transport of China,18*(05), 45–53. https://doi.org/10.13813/j.cn11-5141/u.2020.0505

Möller, C., Alfredsson-Olsson, E., Ericsson, B., et al. (2018). The border as an engine for mobility and spatial integration: A study of commuting in a Swedish–Norwegian context. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography,272*(4), 217–233.

Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies,24*, 9–18. https://doi.org/10.1016/j.trc.2012.01.007

Niu, X., Ding, L., & Song, X. (2017). Analyzing suburban new town development in Shanghai from the perspective of jobs-housing spatial relationship. *Urban Planning,41*(8), 47–53. https://doi.org/10.11819/cpr20170807a

Niu, X., Wang, Y., Liu, J., & Feng, Y. (2018). Spatial structure of Shanghai conurbation area from perspective of inter-city functional links. *Urban Planning Forum,5*, 80–87. https://doi.org/10.16361/j.upf.201805009

O'Clery, N., Curiel, R. P., & Lora, E. (2019). Commuting times and the mobilisation of skills in emergent cities. *Applied Network Science,4*(1), 1–27.

Parr, J. B. (2004). The polycentric urban region: A closer inspection. *Regional Studies,38*(3), 231–240. https://doi.org/10.1080/003434042000211114

Rain, D. (1999). Commuting directionality, a functional measure for metropolitan and nonmetropolitan area standards. *Urban Geography,20*(8), 749–767. https://doi.org/10.2747/0272-3638.20.8.749

Roth, C., Kang, S. M., & Batty, M. (2011). Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS ONE,6*(1), e15923. https://doi.org/10.1371/journal.pone.0015923

Taylor, P. J., Evans, D. M., & Pain, K. (2008). Application of the interlocking network model to mega-city-regions: Measuring polycentricity within and beyond bity-regions. *Regional Studies,42*(8), 1079–1093.

Vasanen, A. (2012). Functional polycentricity: Examining metropolitan spatial structure through the connectivity of urban sub-centres. *Urban Studies,49*(16), 3627–3644. https://doi.org/10.1177/0042098012447000

Wang, J., Luo, G., Huang, Y., et al. (2023). Spatial characteristics and influencing factors of commuting in central urban areas using mobile phone data: A case study of nanning. *Sustainability,15*(12), 9648.
Yang, K. (2010). Residents΄ commuting space in new towns of Chinese metropolises: A case study in Nanjing. *Urban Development Studies,17*(2), 42–46.
Zhong, C., Arisona, S. M., U, L., 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.
Castells M .(1996). Rise of the Network Society: The Information Age: Economy, Society and Culture. Blackwell Publishers, Inc.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ying Tian PhD candidate in the Department of Urban Planning and Design School of Architecture, Tsinghua University, Beijing, China. Planner of CAUPD Beijing Planning &Design Consultants Ltd. Mainly engaged in digital twin city, big data analysis on urban planning.

Changcheng Kan Technical Architect of Baidu.com Times Technology (Beijing)Co. Ltd, Beijing, China. Mainly engaged in the application of machine learning, deep learning, and spatiotemporal big data in smart cities and industries.

Xiangyu Li Postdoctoral Fellow, Tsinghua University, Beijing, China. Mainly engaged in the use of remote sensing technology and geographic information systems on human settlements, regional planning, and urban and rural planning.

Anrong Dang Full Professor of Urban Planning and Design School of Architecture,Tsinghua University, Beijing, China. Mainly engaged in the use of remote sensing technology and geographic information systems on human settlements, regional planning, and urban and rural planning