



OPEN Investigating the impacts of urban agglomeration spatial structures on carbon emissions based on spatial networks of cities

Ge Tan¹, Xiuyuan Zhang², Haoyu Wang¹, Shuping Xiong¹, Zihao Sun¹, Yichen Lei¹ & Shihong Du²✉

Quantifying the relationships between urban agglomeration spatial structures and carbon emissions yields insights for developing sustainable, low-carbon urban agglomerations. Previous studies characterizing urban agglomeration spatial structures usually rely solely on the attribution data of individual cities, but overlook inter-city linkages, i.e., the interactions between cities within urban agglomerations. However, the essence of urban agglomerations roots in the intimate connections between cities through the flows of people, logistics, information, etc. Therefore, it is crucial to accurately characterize the pattern of inter-city connections in order to completely express urban agglomeration spatial structures. However, previous studies have not explored this aspect in depth. To address this gap, this study innovatively utilizes train schedule data in 2010, 2015 and 2020 to measure inter-city connections and construct spatial networks of Chinese cities, thereby modeling the patterns of inter-city linkages. Subsequently, indicators characterizing urban agglomeration spatial structures were developed on the basis of spatial networks of cities. Panel data regression was then employed to investigate the relationships between these indicators and carbon emissions. Experimental results indicate that (1) monocentric spatial structure is conducive to reducing carbon emissions; (2) morphological polycentricity has a more significant effect on carbon emissions compared to functional polycentricity; and (3) network disparity exhibits a positive correlation with carbon emissions. These findings provide essential support for the formulation of carbon emission reduction policies at urban agglomeration level.

As the world's largest carbon emitter, China is actively pursuing its "dual carbon" targets, namely achieving carbon peaking and carbon neutrality^{1,2}. Urban areas account for approximately 85% of the country's carbon emissions^{3,4}, with urban agglomerations contributing more than 60% of emissions⁵. Therefore, for reaching the dual-carbon goal, emission reduction efforts should be focused on urban areas, especially urban agglomerations⁶.

Urban agglomerations are compact, highly economically integrated clusters of cities, typically anchored by one megacity and comprising a core region formed by at least three metropolitan areas or large cities. Through well-developed transportation and infrastructure networks, the core region maintains close connections with surrounding areas accessible for daily commutes. As these urban agglomerations evolve, they will emerge as comprehensive, integrated entities encompassing industrial configuration, infrastructure development, regional market formation, urban-rural planning, environmental management, and the provision of public services. Ultimately, they will form economic and social communities characterized by shared interests and a collective future⁷⁻⁹. Urban agglomerations constitute a novel regional unit actively engaging in global competition and the international division of labor. They are widely regarded as the highest form of spatial organization achieved at the mature stage of urban development and represent an inevitable phase in the trajectory of urbanization in China^{10,11}.

The spatial structures are the direct reflections of the development of urban agglomerations in the spatial dimension¹², which refer to the distribution and spatial combination of various resources, elements and socio-economic activities within the geographical scope of urban agglomerations, and reflect the function, location and socio-economically related structures of cities in specific regions^{13,14}.

¹Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China. ²College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. ✉email: dshgis@hotmail.com

Unlike individual cities, which affect carbon emissions by influencing transportation and residential energy consumption within the city¹⁵, the spatial structures of urban agglomerations profoundly affect carbon emissions through functional specialization and complementarity among cities, as well as through the spatial layout of key elements (e.g., industrial clustering)^{16,17}. The degree of industrial concentration or dispersion may influence production efficiency and the energy required to produce the same level of output, thereby affecting industrial energy use and ultimately carbon emissions^{18,19}. In addition, functional specialization and coordination among cities, together with the spatial organization of supply chains, may alter inter-city transportation energy demand, which could further shape carbon emissions. These possible pathways suggest that different spatial configurations could exert significant impacts on carbon outcomes²⁰. Precisely because these mechanisms remain insufficiently understood, it is important to systematically examine how the spatial structures of urban agglomerations affect carbon emissions—an issue this study seeks to address empirically²¹.

Given the key position and important role of urban agglomerations in the process of regional coordinated development and low-carbon transformation^{6,17,22}, in-depth research on urban agglomeration spatial structures and its effect on carbon emissions is of urgent significance in promoting regional coordination of emission reduction and realizing high-quality development at the practical level^{8,23}.

When exploring the relationships between urban agglomeration spatial structures and carbon emissions, measuring and expressing the spatial structures is the core topic of research. Among these, polycentricity, which refers to the balanced hierarchical relationships among centers within a regional system²⁴, is regarded as a key indicator of the spatial organization patterns of urban agglomerations⁹. Studies have analyzed the polycentricity of urban agglomerations and its evolution process from multiple dimensions and scales.

For example, the monocentric and polycentric characteristics of urban agglomerations have been measured using the primacy index and the rank-size index based on resident population data, and the evolution trends and influencing factors of spatial structures have been explored in the spatio-temporal dimension^{25,26}. The results indicate significant differences in the spatial and temporal distributions of urban agglomeration structures across regions. The spatial structure evolution of 20 urban agglomerations in China during 1992–2012 has also been examined using nighttime light data as a proxy for census information, revealing a transition from monocentric to polycentric forms and identifying the major driving forces behind this transformation²⁷. Furthermore, long-term time-series changes in regional polycentricity across China during 1997–2013 at the provincial scale have been analyzed, offering empirical evidence to better understand the spatiotemporal dynamics of urban agglomeration structures at the macro level^{28,29}.

Several empirical studies have drawn divergent conclusions regarding the spatial structures of urban agglomerations and their economic, social, and environmental implications. An analysis of U.S. metropolitan areas has shown that polycentric spatial structures are positively associated with higher labor productivity relative to monocentric ones³⁰. This finding supported the notion that polycentric structure could mitigate agglomeration diseconomies and facilitate the sharing agglomeration externalities³¹. However, the study also indicated that a network of geographically adjacent small cities remained an insufficient substitute for the urbanization externalities provided by a single large city. The spatial structure of six major urban agglomerations in China was quantified using demographic and economic data based on the rank-size rule, and its spatiotemporal impact on carbon emissions was assessed through a geographically and temporally weighted regression (GTWR) model⁹. The findings indicated a regional shift in the relationship between spatial structure and carbon emissions—from positive to negative in areas such as the Yangtze River Delta and Pearl River Delta. The effects of polycentric structures were found to vary by context, with both promoting and mitigating carbon emissions under different conditions. Further analysis of green development efficiency (GDE) across 19 Chinese urban agglomerations between 2007 and 2020 revealed a negative association between monocentric spatial structures and GDE¹⁴. Another study, taking the Shandong Peninsula Urban Agglomeration as an example¹⁶, measured spatial structures in four dimensions: monocentricity and polycentricity, centralization and diffusion, spatial compactness, and transportation accessibility, and assessed its impacts on carbon emissions. The results indicated that the impacts of polycentric characteristics on carbon emissions are not yet significant, concentration and accessibility have a positive impact on carbon emissions, and spatial compactness has a negative impact on carbon emissions. This indicates that, in the process of urban agglomeration planning, attention should be paid to the urban-rural integration pattern and the construction of low-carbon transportation system.

As mentioned above, while existing studies have explored the characteristics of urban agglomeration spatial structures and their carbon emission effects from various perspectives and scales, significant limitations persist:

- Overreliance on attribute data of individual cities for measuring urban agglomeration spatial structures, lacking a perspective of spatial networks of cities. In characterizing urban agglomeration spatial structures, previous studies predominantly have relied on the attribute data of individual cities within urban agglomerations while overlooked the role of intercity connection data. This approach fails to capture the essence of urban agglomerations as tightly interconnected networks formed through flows of people, goods and information. Such interconnected networks are not only fundamental to understanding the spatial development of urban agglomerations but also essential for capturing their spatial structures. Unfortunately, these are frequently omitted in existing studies, leading to an incomplete understanding of urban agglomeration spatial structures.
- Overconcentration on morphological polycentricity for representing urban agglomeration spatial structures. Polycentricity, as the most widely applied spatial structural feature, is used to assess whether an urban agglomeration is single-centered or multi-centered. However, its roles have been frequently restricted to morphological dimensions. The assessment of morphological polycentricity primarily relies on scale data (such as population and GDP) of cities within an urban agglomeration to quantify the degree of spatial concentration. However, city size does not always accurately reflect its actual importance or centrality³². Certain small and medium-sized cities may play pivotal roles in industrial chains, serving as key nodes for connectivity within

urban agglomerations³³. Therefore, metrics like population and GDP alone are insufficient to reflect these cities' true "central" roles within urban agglomerations, and simultaneously, morphological polycentricity derived from these scale indicators also fails to comprehensively reveal the actual polycentricity of the urban agglomeration.

These gaps resulted in previous studies failing to fully understand and reveal the complex characteristics of urban agglomeration spatial structures and their profound impacts on carbon emissions, thus making it difficult to provide a scientific basis for emission reduction policies at the urban agglomeration scale. To fill these gaps, several approaches are proposed in this study:

- Introducing connection data to construct spatial networks of cities, expressing urban agglomeration spatial structures from the perspective of spatial networks of cities. Train schedule data and gravity model³⁴ are utilized to precisely measure the substantive connections between cities. Subsequently, spatial networks of cities are constructed with cities represented as nodes and connections as edges. Graph theory and network analysis methods are employed to systematically quantify and analyze the internal network structure characteristics of urban agglomerations, with a particular focus on the influence of these structure characteristics (as components of urban agglomeration spatial structures) on carbon emissions.
- Introducing "functional polycentricity" metric. To address the limitations of morphological polycentricity metrics, this study introduces the concept of functional polycentricity^{35,36}. This metric evaluates the functional importance of a city based on its role within the interconnected intercity network and measures the centrality of urban agglomerations by evaluating the spatial concentration of functional importance across multiple cities within the agglomeration.
- Systematically investigating the impacts of urban agglomeration spatial structures on carbon emissions by taking 19 urban agglomerations in China as case studies.

In summary, the main contributions of this paper are as follows. (1) Spatial networks of cities for the years 2010, 2015, and 2020 are mapped to provide a dynamic depiction of intercity connections over time. (2) The spatial structures of urban agglomerations are measured from the network perspective using connection data, which provides a new analytical method for understanding urban agglomeration spatial structures. (3) For the first time, to the best of our knowledge, functional polycentricity and morphological polycentricity are combined to compare the differences in their impacts on carbon emissions, revealing the mechanisms by which different polycentric structures contribute to the carbon emissions of urban agglomerations. And finally, (4) the impacts of spatial structures of urban agglomerations on carbon emissions are explored by taking 19 Chinese urban agglomerations as case study, and carbon emission reduction policy recommendations at the regional level are proposed based on the results of the study. In one word, this study provides a novel perspective and theoretical basis for understanding urban agglomeration spatial structures and its carbon emissions effects.

As urban agglomerations represent an inevitable stage in China's urban development, and as the world's largest carbon emitter, China faces immense pressure to reduce emissions from its cities and urban agglomerations. This study holds significant theoretical and practical value, providing a scientific foundation for the construction of low-carbon urban agglomerations and the promotion of regional sustainable development.

Study area and used data

Study area

Based on China's new urbanization plan (2014–2020), development planning documents for urban agglomerations, and related studies^{5,8,14,37}, 19 urban agglomerations involving 246 cities (Fig. 1) were selected, including the Beijing-Tianjin-Hebei urban agglomeration (BTH-UA), Middle Reaches of the Yangtze River urban agglomeration (MRZR-UA), Pearl River Delta urban agglomeration (PRD-UA), Yangtze River Delta urban agglomeration (YRD-UA), Central Plains urban agglomeration (CP-UA), Chengdu-Chongqing urban agglomeration (CY-UA), Harbin-Changchun urban agglomeration (HC-UA), Guanzhong Plain urban agglomeration (GZP-UA), Central and Southern Liaoning urban agglomeration (CSL-UA), Shandong Peninsula urban agglomeration (SDP-UA), Beibu Gulf urban agglomeration (BG-UA), Lanzhou-Xining urban agglomeration (LX-UA), Northern Slope of Tianshan Mountains urban agglomeration (NSTM-UA), Hu-Bao-E-Yu urban agglomeration (HBZE-UA), Central Shanxi urban agglomeration (CS-UA), Ningxia along the Yellow River urban agglomeration (NYR-UA), Dianzhong urban agglomeration (DZ-UA), Qianzhong urban agglomeration (QZ-UA), and Guangdong Fujian Zhejiang Coastal urban agglomeration (GFZC-UA).

It is worth noting that the clear definitions of the geographic boundaries (or included cities) have not been provided for these urban agglomerations, leaving substantial ambiguity regarding their specific scope. In this study, we reviewed official documents and relevant academic materials and found that certain cities may belong to multiple urban agglomerations, and inconsistencies in the delineation of urban agglomeration boundaries exist across different studies. Given that urban agglomerations are not fixed geographic entities and their boundaries are inherently dynamic and uncertain, this study adopts an inclusive approach by using the "union" of disputed urban agglomeration boundaries as the research scope^{5,8,32}. In cases of disagreement, we have made an effort to include contested cities within the corresponding urban agglomerations. Consequently, our definition of urban agglomeration boundaries may be broader than those in other studies, and instances of cities belonging to multiple urban agglomerations are particularly noticeable in the eastern coastal regions (Table 1). Table 2 presents the socio-economic characteristics of 19 urban agglomerations, including total population and GDP data for three years.

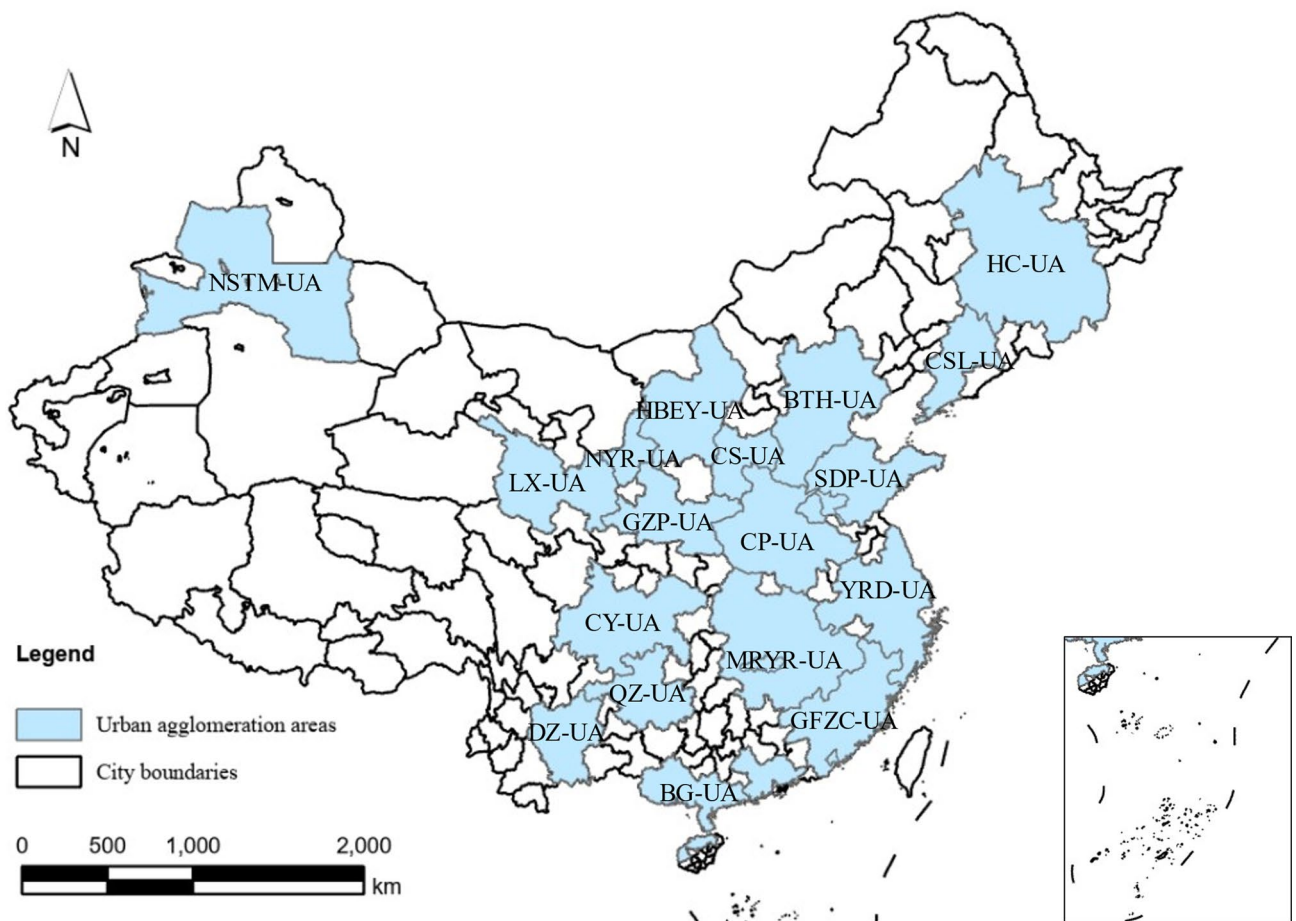


Fig. 1. The 19 urban agglomerations in China.

Data sources and preprocessing

Carbon emissions data

Carbon emissions data serve as a critical dependent variable in regression analysis, with their quality directly influencing the reliability of regression results and the validity of policy recommendations. Therefore, ensuring the accuracy and comprehensiveness of the data is essential. This study primarily utilizes the China City CO₂ Emissions Dataset, developed by the China City Greenhouse Gas Working Group (CCG), as the core data source. This dataset includes carbon emissions data for the majority of Chinese cities for the years 2005, 2010, 2015, and 2020 (available at <https://www.cityghg.com/>).

The dataset is constructed by integrating the following three primary data sources: (1) CHRED 3.0 Database, providing city-level carbon emissions information³⁸; (2) official data derived from statistical yearbooks, government documents, and relevant survey reports at the city level; and (3) data collected through field investigations, interviews, telephone consultations, and official requests to relevant government departments conducted by the CCG.

For cities not included in the dataset, calculations were estimated using ODIAC gridded carbon emissions data. The Open Source Data Inventory of Anthropogenic Carbon Dioxide (ODIAC) is a global, high spatial resolution gridded carbon emissions data product for estimating carbon dioxide (CO₂) emissions from fossil fuel combustion³⁹. The spatial distribution of its emissions is estimated at a spatial resolution of 1 km based on power plant information (e.g., emission intensity and geographic location) as well as satellite-observed nighttime light data on land⁴⁰. ODIAC provides gridded CO₂ emission data distributed by month. In this study, ODIAC data were processed in ArcGIS to generate annual gridded carbon emission data for the corresponding year, and the total carbon emissions within the administrative boundaries of the corresponding cities are further aggregated to ensure the completeness of the carbon emission data for the cities that are not covered by the original dataset.

Train schedule data

Train schedule data serves as a critical source for constructing spatial networks of cities⁴¹, typically encompassing train numbers, station names, stop sequences, arrival and departure times, stop durations, and cumulative distances from the points of origin (Fig. 2). The schedule is typically adjusted by China State Railway Group Co., Ltd. (formerly the Ministry of Railways) in early January (prior to the Spring Festival travel period) and early July (before the summer travel peak), resulting in notable variations in train operations (available at <http://www.ch>

Urban agglomerations	Abbreviation	Cities
Beijing-Tianjin-Hebei	BTH-UA	Baoding, Beijing, Cangzhou, Chengde, Handan, Hengshui, Langfang, Qinhuangdao, Shijiazhuang, Tangshan, Tianjin, Xingtai, Zhangjiakou
Middle Reaches of the Yangtze River urban	MRYR-UA	Changde, Ezhou, Fuzhou (Jiangxi), Hengyang, Huanggang, Huangshi, Jian, Jingmen, Jingzhou, Jingdezhen, Jiujiang, Nanchang, Qianjiang, Shangrao, Tianmen, Wuhan, Xiantao, Xianning, Xiangyang, Xiaogan, Xinyu, Yichang, Yingtan
Pearl River Delta	PRD-UA	Dongguan, Foshan, Guangzhou, Huizhou, Jiangmen, Shenzhen, Zhaoqing, Zhongshan, Zhuhai
Yangtze River Delta	YRD-UA	Anqing, Changzhou, Chizhou, Chuzhou, Hangzhou, Hefei, Huzhou, Jiaying, Jinhua, Maanshan, Nanjing, Nantong, Ningbo, Shanghai, Shaoxing, Suzhou (Jiangsu), Taizhou (Zhejiang), Taizhou (Jiangsu), Tongling, Wenzhou, Wuxi, Wuhu, Xuancheng, Yancheng, Yangzhou, Zhenjiang, Zhoushan
Central Plains	CP-UA	Anyang, Bengbu, Bozhou, Fuyang, Handan, Heze, Hebi, Huaibei, Jiyuan, Jiaozuo, Jincheng, Kaifeng, Liaocheng, Luoyang, Luohe, Nanyang, Pingdingshan, Puyang, Sanmenxia, Shangqiu, Suzhou (Anhui), Xinxiang, Xinyang, Xingtai, Xuchang, Yuncheng, Changzhi, Zhengzhou, Zhoukou, Zhumadian
Chengdu-Chongqing	CY-UA	Chengdu, Dazhou, Deyang, Guangan, Leshan, Luzhou, Meishan, Mianyang, Nanchong, Neijiang, Suining, Yaan, Yibin, Chongqing, Ziyang, Zigong
Harbin-Changchun	HC-UA	Daqing, Haerbin, Liaoyuan, Mudanjiang, Qiqihaer, Siping, Songyuan, Suihua, Changchun
Guanzhong Plain	GZP-UA	Baoji, Linfen, Pingliang, Shangluo, Tianshui, Tongchuan, Weinan, Xian, Xianyang, Yuncheng
Central and Southern Liaoning	CSL-UA	Anshan, Benxi, Dalian, Fushun, Liaoyang, Panjin, Shenyang, Tieling, Yingkou
Shandong Peninsula	SDP-UA	Binzhou, Dezhou, Dongying, Jinan, Liaocheng, Qingdao, Rizhao, Weihai, Weifang, Yantai, Zibo
Beibu Gulf	BG-UA	Beihai, Changjianglizu, Chengmai, Chongzuo, Danzhou, Dongfang, Fangchenggang, Haikou, Lingao, Maoming, Nanning, Qinzhou, Yangjiang, Yulin (Guangxi), Zhanjiang
Lanzhou-Xining	LX-UA	Baiyin, Dingxi, Haibeizangzu, Haidong, Hainanzangzu, Huangnanzangzu, Lanzhou, Linxiahuizu, Xining
Northern Slope of Tianshan Mountains	NSTM-UA	Changjiuizu, Kelamayi, Tacheng, Tulufan, Wulumuqi, Yilihasake
Hu-Bao-E-Yu	HBEY-UA	Baotou, Eerduosi, Huhehaote, Yulin (Shanxi)
Central Shanxi	CS-UA	Jinzhong, Linfen, Lüliang, Taiyuan, Yangquan
Ningxia along the Yellow River	NYR-UA	Shizuishan, Wuzhong, Yinchuan, Zhongwei
Dianzhong	DZ-UA	Chuxiongqiyu, Honghehanizuyizu, Kunming, Qujing, Yuxi
Qianzhong	QZ-UA	Anshun, Bijie, Guiyang, Qiandongnanmiaozudongzu, Qiannanbuyizumiao, Zunyi
Guangdong Fujian Zhejiang Coastal	GFZC-UA	Fuzhou (Fujian), Fuzhou (Jiangxi), Ganzhou, Lishui, Longyan, Nanping, Ningde, Putian, Quzhou, Quanzhou, Sanming, Xiamen, Shangrao, Wenzhou, Yingtan

Table 1. 19 urban agglomerations in China: abbreviations and constituent cities.

Urban agglomerations	Population (2010)	GDP (2010)	Population (2015)	GDP (2015)	Population (2020)	GDP (2020)
BTH-UA	104.47	4392.97	111.61	7004.68	107.47	8556.03
MRYR-UA	121.20	3438.44	124.59	6481.17	123.35	9156.14
PRD-UA	56.18	3834.39	58.73	6329.23	77.99	9039.18
YRD-UA	147.72	8764.21	162.84	14249.92	179.10	21394.41
CP-UA	157.60	3478.90	161.30	5565.90	164.05	8060.79
CY-UA	95.76	2320.15	98.24	4401.19	102.73	6811.01
HC-UA	48.95	1698.54	48.67	2487.37	42.60	2118.93
GZP-UA	42.56	918.92	43.35	1597.82	43.41	2262.07
CSL-UA	30.68	1742.92	31.06	2420.79	30.70	2111.68
SDP-UA	94.58	3965.08	97.14	6247.98	101.53	7311.55
BG-UA	36.92	839.73	39.52	1542.76	42.42	2066.44
LX-UA	14.55	264.34	15.00	487.23	15.73	629.06
NSTM-UA	11.27	401.26	11.61	682.85	11.04	798.29
HBEY-UA	10.81	872.38	11.34	1351.74	11.93	1320.96
CS-UA	16.86	470.52	17.32	649.93	17.38	941.00
NYR-UA	5.19	145.84	5.47	269.79	6.11	356.77
DZ-UA	21.73	491.74	22.46	882.37	23.38	1549.54
QZ-UA	26.03	353.43	26.46	886.90	29.21	1379.37
GFZC-UA	88.13	2511.20	90.77	4383.84	90.20	6997.07

Table 2. Social and economic characteristics of 19 urban agglomerations in China. Population is in millions, GDP is in billions of CNY.

Train No.	Stop No.	Station Name	Arrival Time	Departure Time	Stop Duration (min)	Distance from Start (km)
G1	1	Beijing South	—	07:00	0	0
G1	2	Tianjin South	07:30	07:32	2	120
G1	3	Cangzhou West	08:07	08:09	2	235
G1	4	Dezhou East	08:42	08:44	2	345
G1	5	Jinan West	09:14	09:17	3	405
G1	6	Tai'an	09:44	09:46	2	485
G1	7	Xuzhou East	10:26	10:28	2	680
G1	8	Nanjing South	11:28	11:31	3	1020
G1	9	Shanghai Hongqiao	12:00	—	0	1318

Fig. 2. Train schedule data.

www.12306.cn/). For the remainder of the year, train schedules remain largely consistent⁴². Therefore, this study selects data from June 30th of each year as the representative dataset, sourced from the 12,306 (available at <https://www.12306.cn/>).

Foundational urban statistical data

The foundational urban statistical data, such as population and economic indicators, are sourced from the *Urban Statistical Yearbook of China*. This yearbook systematically compiles socio-economic statistics for cities across various years, including key indicators such as population size, GDP, industrial structure, and land use, providing a reliable basis for urban spatial network research. Due to missing population data for certain cities in specific years, we addressed this issue by using the WorldPop dataset⁴³ for the corresponding years to fill in the gaps. Specifically, we aggregated the gridded WorldPop data to the administrative boundaries of the cities to obtain the complete population data for each region. This approach ensures that the population data used in the analysis is both consistent and comprehensive.

Administrative boundary data

The administrative boundary data used in this study are derived from the Database of Global Administrative Areas (GADM) (available at <http://www.gadm.org/>). After projection transformation and topology checking, the administrative division data of different scales are used in the preprocessing, calculation, and analysis of other experimental data.

Methodology

This study aims to explore the impact of urban agglomeration spatial structures on carbon emissions. Hence, a methodological framework is proposed, consisting of three parts: (1) constructing spatial networks of cities; (2) presenting an indicator system for measuring urban agglomeration spatial structures; and (3) quantifying the influences of urban agglomeration spatial structures on carbon emissions by panel data regression analysis, supplemented by robustness analysis.

Construction of spatial networks of cities

The key to constructing spatial networks of cities lies in measuring the connection strength between cities. We used train schedule data as the primary data source for constructing spatial networks of cities, which only provides information on the names of departure stations, terminal stations, intermediate stops, and their operating frequencies (Fig. 2). However, quantifying the strength of connections between cities based on this data poses a significant challenge. To address this, a four-step approach is used in our study:

- (1) **Mapping Stations to Prefecture-Level Cities.** Since the research units are urban agglomerations composed of cities, the smallest spatial unit should correspond to prefecture-level cities. However, train schedule data are provided at the station level, and a single city may have multiple train stations. To ensure consistency, we aggregate all train stations within the same city to the prefecture-level scale, enabling unified spatial analysis.
- (2) **Generating City Pairs.** For each train route, all possible city pairs are generated based on the sequence of cities along its path, as these cities are likely connected through the flows of people and goods facilitated by the train. For example, if a train route passes through cities $A \rightarrow B \rightarrow C$, **this results in** three city pairs $A \rightarrow B$, $B \rightarrow C$ and $A \rightarrow C$. For the via n cities, the number of city pairs of the train is

$$\text{Number of City Pairs} = \frac{n \times (n - 1)}{2} \quad (1)$$

(3) Weight Assignment

At this stage, by applying the aforementioned city pair generation steps (1) and (2) to all train routes and aggregating the results at the city-pair level, the train frequency of all city pairs can be obtained. Some studies directly use this frequency to represent the connection strength between cities^{42,44}; however, this method implicitly assumes that the passenger flow between any two cities along the same train route is equal—an assumption that is inconsistent with reality. In fact, passenger flows between two cities on the same route are influenced by multiple factors⁴⁵. To illustrate this, two examples are provided as follows:

- (1) On a train route departing from Beijing, terminating in Guangzhou, and passing through Shijiazhuang, Zhengzhou, and Changsha, it can be reasonably assumed that the number of passengers traveling from Beijing to Shijiazhuang or Zhengzhou is significantly larger than the number of passengers traveling to Guangzhou. This is primarily because shorter distances tend to produce higher travel demand, whereas long-distance trips are relatively less frequent.
- (2) On another route departing from Beijing, passing through Hami (a small city with a relatively low population), and ultimately arriving in Urumqi, it is clear that the number of passengers traveling from Beijing to Urumqi is significantly larger than the number of passengers traveling to Hami, despite the shorter distance between Beijing and Hami. This discrepancy arises from the smaller scale of Hami, where there is limited demand for visits, resulting in weaker actual connections with Beijing.

These examples collectively demonstrate that assuming identical passenger flow between any two cities on the same train route is not justified. Consequently, relying solely on train frequency to represent intercity connection strength may fail to fully capture the actual magnitude of intercity interactions. In order to more accurately assess the strength of connections between city pairs, multiple factors need to be taken into account, the most critical of which are the city size and the geographic distance⁴⁶. Specifically, city pairs characterized by shorter distances and larger city size tend to have an advantage in terms of passenger flows and naturally have a stronger connection between them, while the actual demand for travel between cities that are farther away or smaller in size will be relatively weaker. The gravity model is an effective solution to this issue. It draws on the concept of gravity in physics, by incorporating both city size (e.g., population or economic indicators) and geographic distance to assign weights between different pairs of cities within the same train route⁴⁷.

The relative weight W_{ij} between city i and city j on the same train route is:

$$W_{ij} = \frac{P_i \cdot P_j}{d_{ij}^2} \quad (2)$$

where P_i is Population size or economic scale of city i , P_j is Population size or economic scale of city j , d_{ij} is the Geographic distance between cities i and j . Subsequently, normalization is performed to obtain the absolute weight W'_{ij} .

$$W'_{ij} = \frac{W_{ij}}{\sum_i \sum_j W_{ij}} \quad (3)$$

- (4) **Aggregating to Calculate Comprehensive Connection Strength.** The absolute weights of city pairs generated from all train routes are aggregated to obtain the comprehensive connection strength between city pairs, which serves as a quantitative measure of the links between nodes in spatial networks of cities.

It is worth noting that the connection strength between cities measured by train schedules mainly reflects passenger flows. While passenger train schedules provide an effective proxy for intercity connections by reflecting the relative magnitude of passenger flows across cities, they inevitably present certain limitations. Specifically, this dataset does not incorporate other important modes of intercity linkages such as road, air, or freight transportation, which also play significant roles in shaping the overall spatial structure of urban agglomerations. Therefore, the connectivity measured in this study should be interpreted as passenger-based relative linkages, which capture a substantial but not exhaustive dimension of urban interactions. This limitation is acknowledged and discussed to remind readers of the representativeness of the results.

Indicator system for measuring urban agglomeration spatial structures

This study proposes an indicator system based on spatial networks of cities to quantitatively characterize the urban agglomeration spatial structures and more accurately analyze and evaluate their impact on carbon emissions. The system consists of four components: (1) morphological polycentricity, (2) functional polycentricity, (3) network structure, and (4) control variables.

Morphological polycentricity

Polycentricity is commonly used to assess whether an urban agglomeration exhibits a single-center or multi-center structure^{17,48}. However, in-depth analysis of polycentricity extends beyond the simple question of “whether multiple centers exist.” Instead, it delves deeper into the hierarchical relationships among cities (or centers) within the urban agglomeration²⁴. Consequently, the key to quantifying polycentricity lies in characterizing the hierarchical ranking relationships of cities, which are determined by evaluating their relative status in terms of size, function, and influence. Based on the indicators used to quantify the relative status of cities, polycentricity can be classified into morphological polycentricity and functional polycentricity^{9,35}. Morphological polycentricity is typically assessed using indicators that reflect city size, such as population or GDP⁴⁹, while functional polycentricity emphasizes the strength of functional linkages and interactions among cities⁵⁰.

In this study, the morphological polycentricity of urban agglomerations is assessed by representing the relative status of cities through their population size and quantifying the hierarchical relationships between cities using the rank-size index⁵¹. The rank-size index is an important index used to characterize the size distribution of cities within urban agglomerations, which is based on the Rank-Size Rule and reveals the relationships between the sizes of different cities in urban agglomerations and their rankings through a mathematical formula. The formula of rank-size distribution can be expressed as⁵²:

$$P(r) = P_1 \cdot r^{-q} \quad (4)$$

where $P(r)$ represents the population size of the city ranked r (e.g., population, GDP, or total nighttime light intensity), P_1 denotes the population size of the largest city (i.e., the city ranked 1), r is the rank of the city, and q is the scaling parameter, which reflects the degree of polycentricity within the urban agglomeration. To fit the value of q , this study employs the least squares method, substituting the actual city population size and rank data into the formula and applying a logarithmic transformation to enable linear fitting:

$$\ln P(r) = \ln P_1 - q \ln r \quad (5)$$

A larger q indicates that the size distribution within the urban agglomeration tends toward a monocentric structure, where the largest city dominates significantly, with its size far exceeding that of other cities, forming a single-core urban agglomeration. Conversely, a smaller q suggests a shift toward a polycentric structure, characterized by more balanced size distributions among cities, with multiple cities playing relatively equivalent roles in the urban agglomeration.

Functional polycentricity

City size does not always accurately reflect its actual importance or centrality³². Some small and medium-sized cities may play pivotal roles in industrial chains, serving as key nodes for connectivity within urban agglomerations³³. The size indicators such as population and GDP fail to fully capture the true “central” role of these cities in urban agglomerations. As a result, morphological polycentricity based solely on scale indicators has certain limitations. To address this, this study introduces functional polycentricity, reevaluating the relative status of cities from the perspective of their importance in the interconnection network of the urban agglomeration.

We conceptualize urban agglomerations as a weighted network, where cities are represented as nodes, connections between nodes as edges, and edge weights signify the strength of connections between cities. The relative status of a city (node) is measured by its weighted degree centrality, calculated using the following formula⁵³:

$$C_i = \sum_j W_{ij} \quad (6)$$

where C_i represents the weighted degree centrality of city i , which quantifies the overall importance (relative status) of city i within the network. W_{ij} denotes the weight of the edge between city i and city j , reflecting the strength of the connection between these two cities. This formula calculates the total strength of connections for city i by summing the weights of all edges linked to it, thereby assessing its overall importance (relative status) within the urban agglomeration.

Based on this, hierarchical relationships between cities are constructed. The rank-size index is then utilized to quantify these hierarchical relationships⁵⁴, applied through Eqs. (4) and (5), to ultimately derive the evaluation of functional polycentricity.

Network structures

Considering that urban agglomerations are essentially a tightly interwoven network of people, logistics and information flows, their network characteristics serve as key measures of their spatial structures⁵. This paper portrays the characteristics of network structures within urban agglomerations from the perspective of graph

theory by constructing spatial networks of cities⁵⁵. In summary, we selected the three metrics, Weighted Density, Weighted Clustering Coefficient, and Network Disparity^{56,57}, to measure the network structures of urban agglomerations.

- (1) **Weighted Density.** Weighted density reflects the overall connectivity strength and the intensity of resource flow within the network. It is used to evaluate the interconnectivity of urban agglomerations at the macro level. The formula is as follows:

$$D_w = \frac{2 * \sum_{i,j \in N} w_{ij}}{n(n-1)} \quad (7)$$

where w_{ij} is the edge weight between nodes i and j , n is the total number of nodes in the network, $\sum_{i,j \in N} w_{ij}$ represents the total weight of all edges in the network, $n(n-1)$ denotes the maximum possible number of edges in a fully connected network.

- (2) **Weighted Clustering Coefficient.** The clustering coefficient measures the density of connections within a node's local neighborhood, reflecting the clustering tendencies of the network. In the case of urban agglomeration networks, the calculation of the clustering coefficient presents several challenges due to the network's unique characteristics. Firstly, these networks are weighted graphs with significant variation in edge weights. Secondly, they are close to complete graphs, meaning that nearly all nodes are connected, but some edge weights are extremely small, approaching zero. Without any processing, these low-weight edges may lead to an overestimation of the clustering coefficient, causing it to approach the maximum value of 1 and obscuring the network's actual clustering characteristics^{53,56}. To mitigate this issue, we implemented the following steps: First, we filtered out edges with weights below the national median edge weight for the given year. Let w_{median} represent the national median edge weight, and the filtered edge set is defined as:

$$E' = \{(i, j) \in E | w_{ij} > w_{\text{median}}\} \quad (8)$$

where E is the original edge set, and w_{ij} denotes the weight of edge (i, j) . This step removes edges with low weights, which could distort the clustering coefficient calculation, while retaining all nodes. Next, the filtered graph is treated as an unweighted graph for the clustering coefficient calculation. In an unweighted graph, the clustering coefficient is defined as⁵⁵:

$$C(v) = \frac{2 \times E_v}{k_v(k_v - 1)} \quad (9)$$

$$C_{\text{avg}} = \frac{1}{N} \sum_{v \in V} C(v) \quad (10)$$

where $C(v)$ represents the clustering coefficient of node v , k_v is the degree of node v , E_v is the number of edges between the neighbors of node v . N is the total number of nodes, and

V is the set of all nodes, C_{avg} is the weighted clustering coefficient averaged over all vertices of the network.

- (3) **Network Disparity.** Disparity quantifies the overall uneven distribution of edge weights across all nodes in a network, reflecting the global structural heterogeneity in the entire system⁵⁷. Theil Index is employed to measure this characteristic, defined as⁵⁸:

$$T = \frac{1}{M} \sum_{e=1}^M \frac{w_e}{\bar{w}} \ln \left(\frac{w_e}{\bar{w}} \right) \quad (11)$$

where w_e denotes the edge weight between nodes i and j , \bar{w} is the average edge weight, M represents the total number of edges in the network.

Control variables

This study incorporates both the industrial structure and the overall size of urban agglomerations as control variables. The industrial structure is quantified by the proportion of the added value of the secondary industry to GDP. Industrial structure is a key determinant of carbon emissions⁵⁹, with energy-intensive, industry-dominated economies typically associated with higher emissions, whereas an increase in the share of cleaner, low-energy tertiary sectors is linked to a reduction in emissions⁶⁰. In addition, the overall size of urban agglomerations, represented by the total GDP of all cities within the agglomeration, is also included as a control variable. This inclusion accounts for scale effects that could potentially confound the relationship between the independent and dependent variables. By incorporating both industrial structure and agglomeration size, the study seeks to

mitigate the influence of non-research-related factors, thereby enhancing the precision of the analysis of the impacts of the independent variables on carbon emissions.

Panel data regression model and robustness analysis

Panel data regression model

This study employs panel data regression, which offers significant advantages over conventional cross-sectional or time-series models⁶¹, to explore the relationships between urban agglomeration spatial structures and carbon emissions. Panel data regression can effectively control unobserved heterogeneity across entities through fixed or random effects, thereby reducing omitted variable bias. Moreover, it can integrate temporal and cross-sectional data to comprehensively analyze the dynamic changes and long-term trends of variables, enhancing the explanatory power and reliability of the model⁶². Owing to these advantages, panel data regression has been widely applied in studies on the relationship between urban spatial structures and carbon emissions⁶³, as it provides a more precise understanding of the complex and multidimensional interactions between them⁶⁴.

The basic formula for panel data regression is:

$$y_{it} = \alpha + \beta X_{it} + \mu_i + \epsilon_{it} \quad (12)$$

where y_{it} refers to the dependent variable for entity i at time t (in this study, the carbon emissions of urban agglomeration i at time t), X_{it} denotes the explanatory variables (the spatial structure characteristics of urban agglomeration i at time t), α is the intercept, μ_i captures the individual effects (either fixed or random), and ϵ_{it} is the random error term. The core of panel data regression lies in how the individual effects μ_i are treated, which can be done by two models: the Fixed Effects Model (FEM) and the Random Effects Model (REM).

- (1) **Fixed Effects Model (FEM).** The FEM assumes that the individual effects μ_i are fixed and may correlate with the explanatory variables X_{it} . By introducing entity-specific intercepts, FEM controls for unobserved, time-invariant characteristics of each entity, thereby reducing omitted variable bias. This model is particularly suitable for studying within-entity variations over time.
- (2) **Random Effects Model (REM).** The REM assumes that the individual effects μ_i are random and uncorrelated with the explanatory variables X_{it} . By treating μ_i as a random variable, REM leverages both cross-sectional and time-series variations, improving the efficiency of parameter estimation. This model is appropriate when the focus is on overall patterns across entities with minimal individual heterogeneity.

The choice between FEM and REM is typically determined by the Hausman test⁶⁵, which evaluates whether the individual effects are correlated with the explanatory variables, guiding the selection of the appropriate model.

Robustness analysis

The panel regression model inherently relies on linear assumptions, which may limit its ability to capture potential nonlinear effects and complex interactions between spatial structures and carbon emissions. As a result, its explanatory power may be constrained. To address this limitation, this study incorporates the Random Forest (RF) model⁶⁶ and its interpretative tool, SHAP (Shapley Additive Explanations)⁶⁷, as complementary methods.

Random Forest model, a non-parametric machine learning method, does not rely on linear assumptions and is capable of automatically learning nonlinear relationships between variables and carbon emissions, while also identifying high-dimensional interaction effects. However, its drawback lies in the absence of traditional significance tests and coefficient interpretations typically associated with regression models⁶⁶.

To enhance the interpretability of the model, we employ SHAP (Shapley Additive Explanations) on top of the Random Forest model⁶⁷. The SHAP method, based on the Shapley value principle from cooperative game theory, decomposes the model's predictions into the marginal contributions of each explanatory variable, thus enabling transparent interpretations within a complex nonlinear framework. Its core concept involves calculating the average marginal contribution of a variable across all possible feature combinations to assess its importance in the prediction. Compared to traditional feature importance methods, SHAP offers theoretical advantages in terms of consistency and local accuracy. It ensures that when the effect of a variable on the model's prediction increases, its importance measure does not decrease, and that the model's predictions can be accurately reproduced by the linear summation of the contributions of individual variables. Furthermore, SHAP can describe the average contribution of different variables to carbon emissions at the macro level, as well as explain the prediction for individual urban agglomerations at specific time points, providing a combined micro-macro explanatory power.

Results

Spatial network of cities within urban agglomerations

Six representative urban agglomerations were selected to reflect diverse geographical regions, development stages, and spatial structure characteristics. Based on data from 2020, these urban agglomerations are analyzed through geographic and topological maps. The selected urban agglomerations include the Beijing-Tianjin-Hebei Urban Agglomeration (BTH-UA), the Chengdu-Chongqing Urban Agglomeration (CY-UA), the Middle Reaches of the Yangtze River Urban Agglomeration (MRYR-UA), the Northern Slope of Tianshan Mountains

Urban Agglomeration (NSTM-UA), the Beibu Gulf Urban Agglomeration (BG-UA), and the Harbin-Changchun Urban Agglomeration (HC-UA).

As shown in Fig. 3, Beijing and Tianjin serve as central nodes, dominating the spatial connections within the BTH-UA. Cities such as Baoding and Shijiazhuang act as secondary nodes, playing significant roles in bridging and linking the network, while other cities exhibit relatively sparse connections. Overall, the BTH-UA demonstrates a distinct “core-periphery” spatial structure.

Figure 4 illustrates the network structure of the MRYR-UA, where Wuhan, Changsha, and Nanchang serve as core nodes, demonstrating dense spatial connections. Among them, Changsha exhibits the strongest radiation effect, with higher connection intensity, occupying a central position and playing a dominant role in the MRYR-UA network. Wuhan and Nanchang follow closely, maintaining strong connectivity with multiple cities, and together with Changsha, form the core nodes of the network. The connectivity of most other cities is relatively balanced.

The CY-UA exhibits significantly lower overall connection intensity (Fig. 5) than the BTH-UA and MRYR-UA, reflecting the relatively weaker network integration of urban clusters in western China. Chengdu occupies a dominant position within the spatial network, characterized by denser and stronger connections with surrounding cities, highlighting its substantial radiation capacity and leading role. In contrast, Chongqing, while an important node in the region, demonstrates notably weaker connection intensity and a smaller radiation range compared to Chengdu, resulting in an imbalanced dual-core structure. Overall, the spatial network of the Chengdu-Chongqing urban agglomeration exhibits a Chengdu-centered radial structure, with limited horizontal connections between other cities and relatively poor network balance.

The spatial connections within the NSTM-UA are relatively sparse (Fig. 6), with a limited number of connections between cities and weak connection intensity. This indicates a low degree of regional network integration and insufficient spatial connectivity. The absence of a central city with strong radiation capacity, along with the near absence of connections for some peripheral cities, has resulted in an extremely unbalanced and fragmented network structure, failing to establish an efficient spatial connection pattern.

The HC-UA located in Northeast China (Fig. 7) is characterized by Harbin and Changchun as regional core nodes. These two cities exhibit the densest spatial connections, playing a leading and radiative role in establishing stable linkages with surrounding cities. Jilin follows, maintaining relatively strong connections with the core cities but demonstrating weaker connectivity with other cities. The remaining cities exhibit relatively balanced interconnections, though their overall connectivity is significantly lower than that of the three core cities.

The BG-UA, as the only urban agglomeration spanning the strait, exhibits significant cross-regional connectivity (Fig. 8). The distribution of network connections among cities is uneven, showing clear regional clustering patterns. Nanning, as the core node on the northern side of the strait, demonstrates strong radiation capacity, whereas the cities on the southern side have relatively sparse and balanced connections, with no prominent core node emerging.

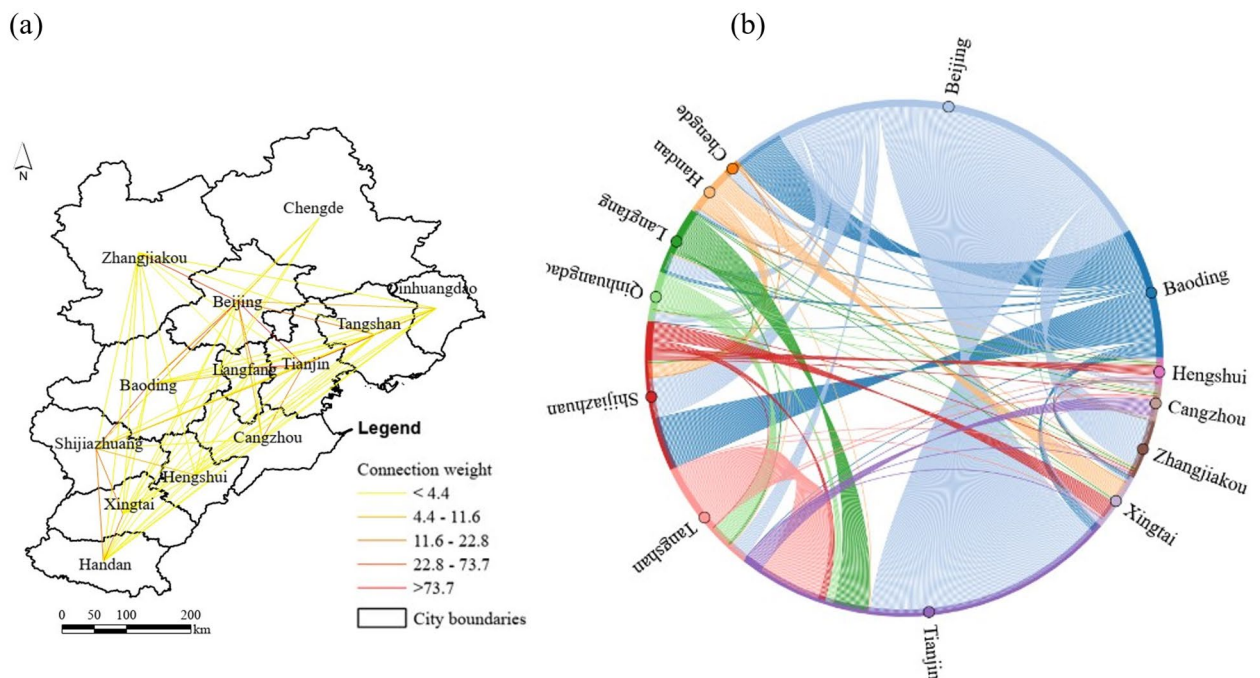


Fig. 3. Spatial network of BTH-UA. (a) Map generated by the authors using ArcGIS Pro 3.1 (Esri). (b) Chord diagram produced in Python 3.9 using the HoloViews library for visualization, based on network data constructed in this study.

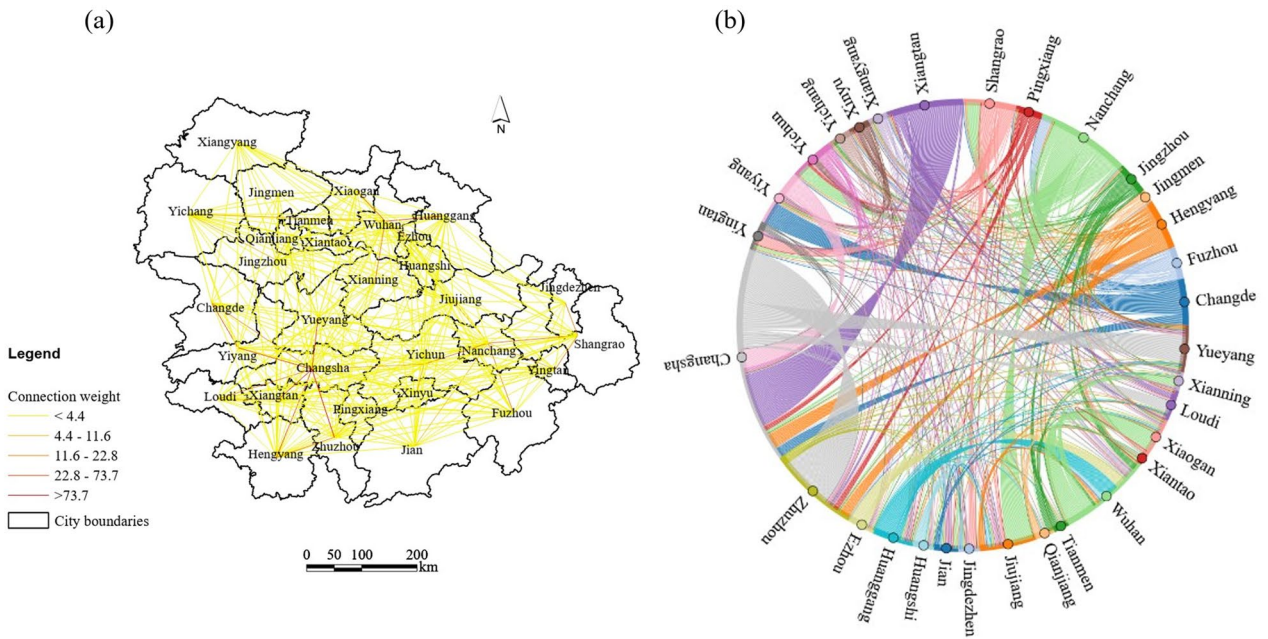


Fig. 4. Spatial network of MRYR-UA. (a) Map generated by the authors using ArcGIS Pro 3.1 (Esri). (b) Chord diagram produced in Python 3.9 using the HoloViews library for visualization, based on network data constructed in this study.

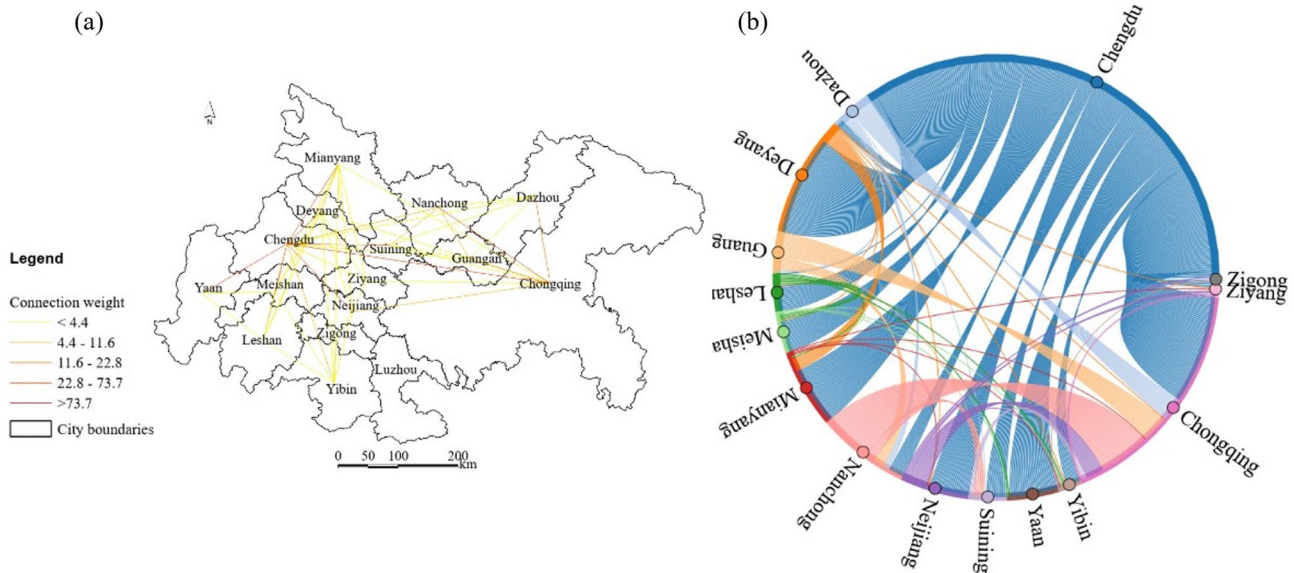


Fig. 5. Spatial network of CY-UA. (a) Map generated by the authors using ArcGIS Pro 3.1 (Esri). (b) Chord diagram produced in Python 3.9 using the HoloViews library for visualization, based on network data constructed in this study.

Carbon emissions of urban agglomerations

The total carbon emissions for the 19 urban agglomerations in 2010, 2015, and 2020 were calculated using the methods outlined in Sect. 2.2.1 (Fig. 9). The carbon emissions of the 19 UAs show an overall upward trend in 2010, 2015, and 2020. The YRD-UA exhibits the highest emissions with a significant increase, reflecting intensive economic activity and high energy consumption, followed by the BTH-UA, CP-UA, and SDP-UA. Notably, the DZ-UA is the only UA with a decline in emissions, while the NSTM-UA shows the largest increase. Additionally, the LX-UA, GZP-UA, HC-UA, CS-UA, CY-UA, and PRD-UA demonstrate a “rise-then-fall” pattern.

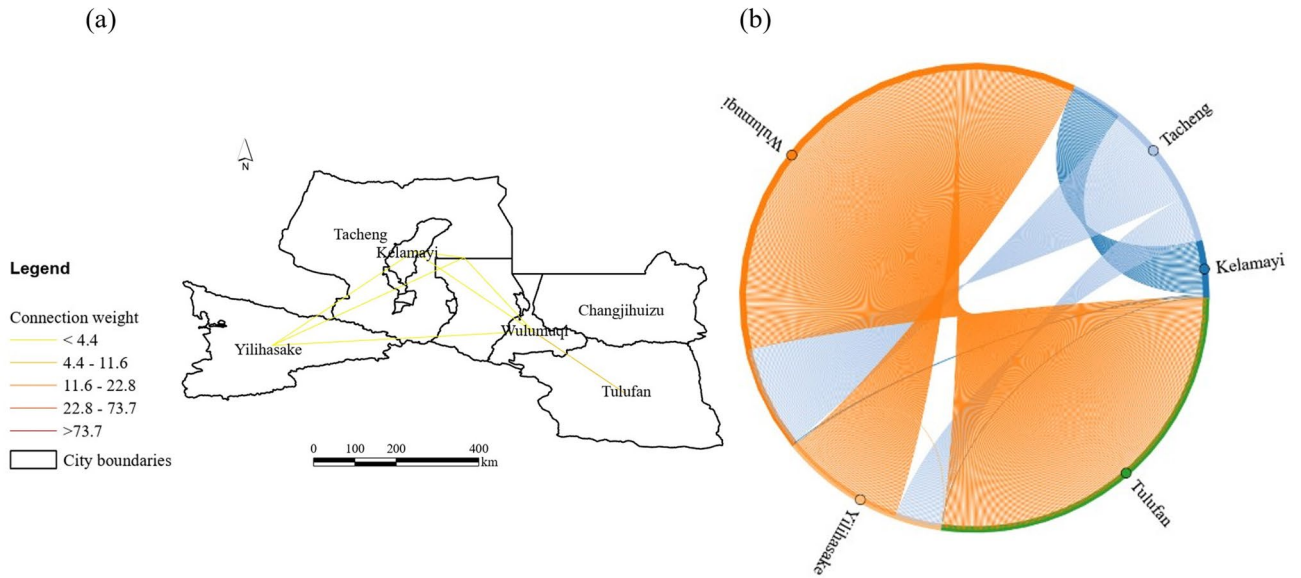


Fig. 6. Spatial network of NSTM-UA. **(a)** Map generated by the authors using ArcGIS Pro 3.1 (Esri). **(b)** Chord diagram produced in Python 3.9 using the HoloViews library for visualization, based on network data constructed in this study.

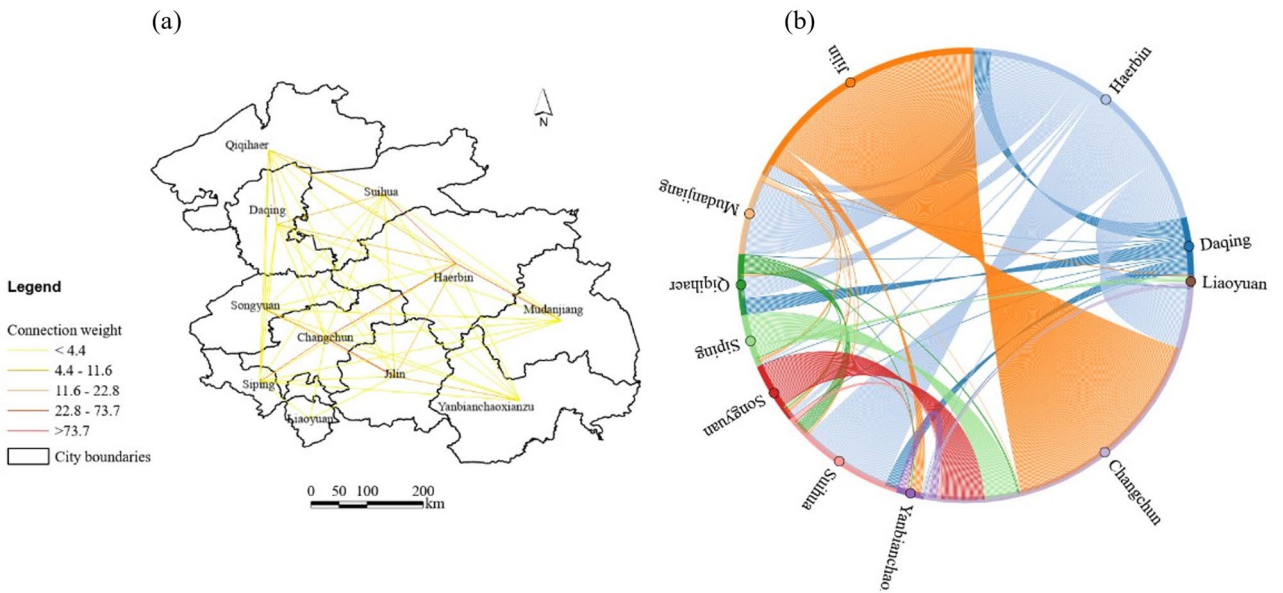


Fig. 7. Spatial network of HC-UA. **(a)** Map generated by the authors using ArcGIS Pro 3.1 (Esri). **(b)** Chord diagram produced in Python 3.9 using the HoloViews library for visualization, based on network data constructed in this study.

The spatial structure indicators of the 19 urban agglomerations

The spatial structure indicators for the 19 urban agglomerations were computed based on the indicator system described in Sect. 3.2 and spatial networks of cities outlined in Sect. 4.1 for the years 2010, 2015, and 2020.

Figure 10 illustrates the morphological polycentricity (a high value indicates a strong tendency toward monocentric structures) of the 19 urban agglomerations in 2010, 2015, and 2020. Overall, the evolution of polycentricity varies significantly across the urban agglomerations, exhibiting no clear general trend. The results encompass monocentric development, polycentric development, and relatively stable, moderate changes over time. The NSTM-UA records the highest morphological polycentricity, indicating a highly monocentric spatial structure where human and socioeconomic activities are predominantly concentrated in the core city,

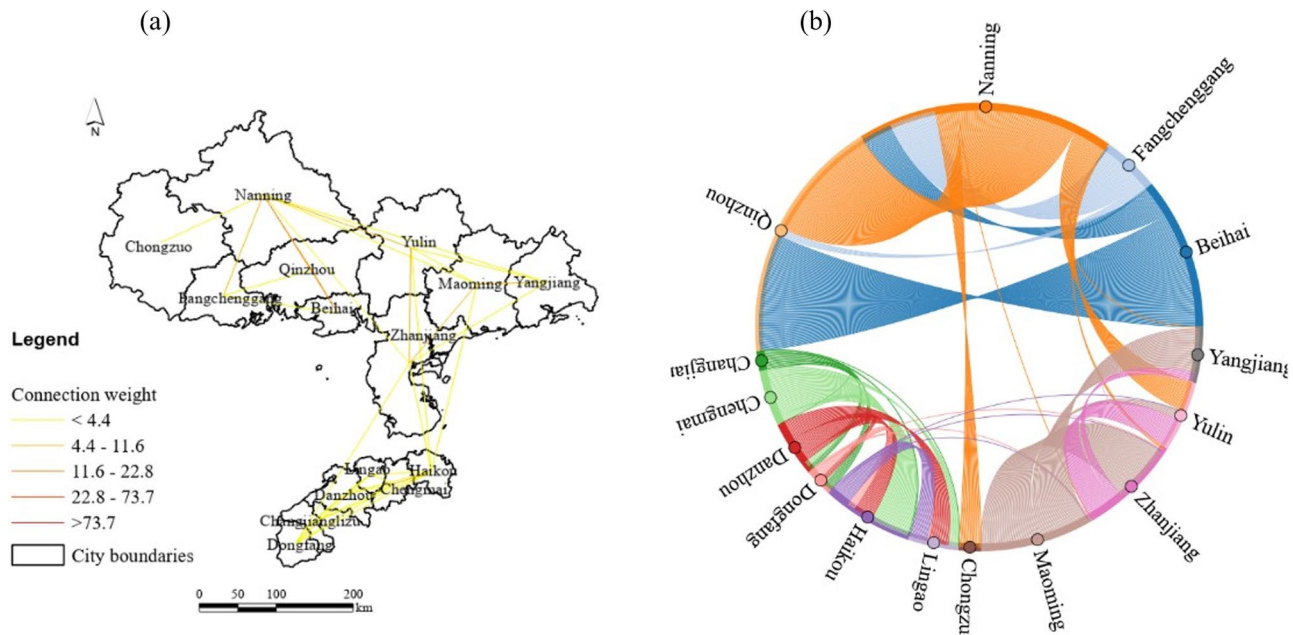


Fig. 8. Spatial network of BG-UA. (a) Map generated by the authors using ArcGIS Pro 3.1 (Esri). (b) Chord diagram produced in Python 3.9 using the HoloViews library for visualization, based on network data constructed in this study.

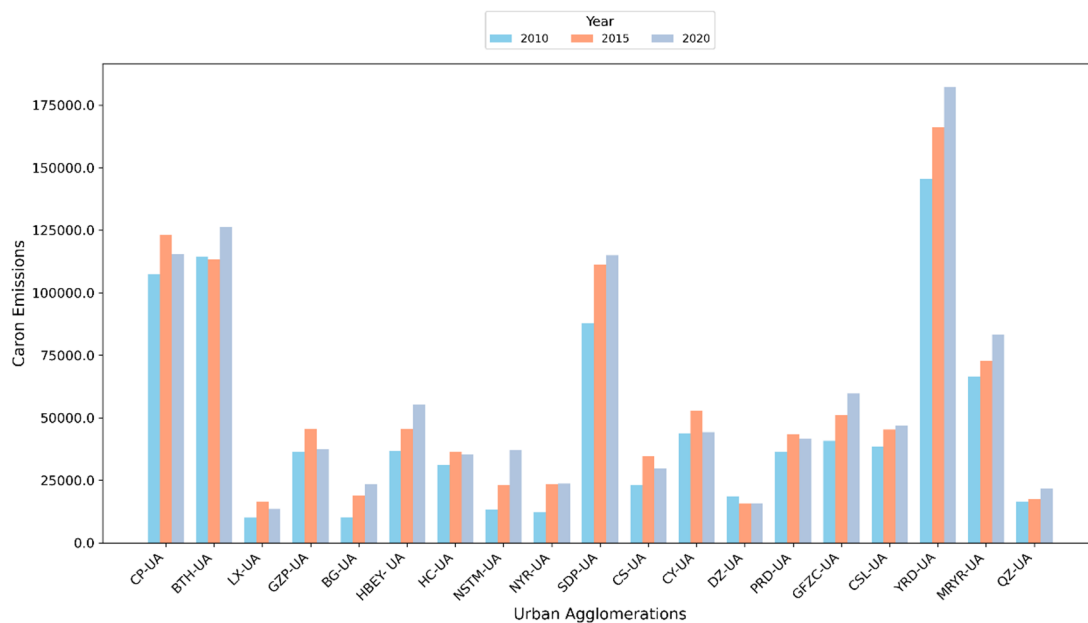


Fig. 9. Carbon emissions of the 19 urban agglomerations.

Urumqi, while surrounding areas remain underdeveloped. The CY-UA ranks second, which contrasts with common perceptions. Although the CY-UA is typically recognized as a dual-core urban agglomeration, the significantly larger population and area of Chongqing, as China’s largest municipality, result in an overall pronounced monocentric pattern. In contrast, the HBEY-UA, SDP-UA, and CP-UA exhibit clear polycentric characteristics, with no dominant core city and a relatively balanced distribution of cities in terms of scale and spatial configuration.

Figure 11 shows the functional polycentricity of the 19 urban agglomerations in 2010, 2015 and 2020. Compared with morphological polycentricity, the values of functional polycentricity are generally high. This indicates that, compared with scale factors such as population and GDP, the functional factors within urban agglomerations are more concentrated, the monocentric structure is more pronounced, and the dominant role

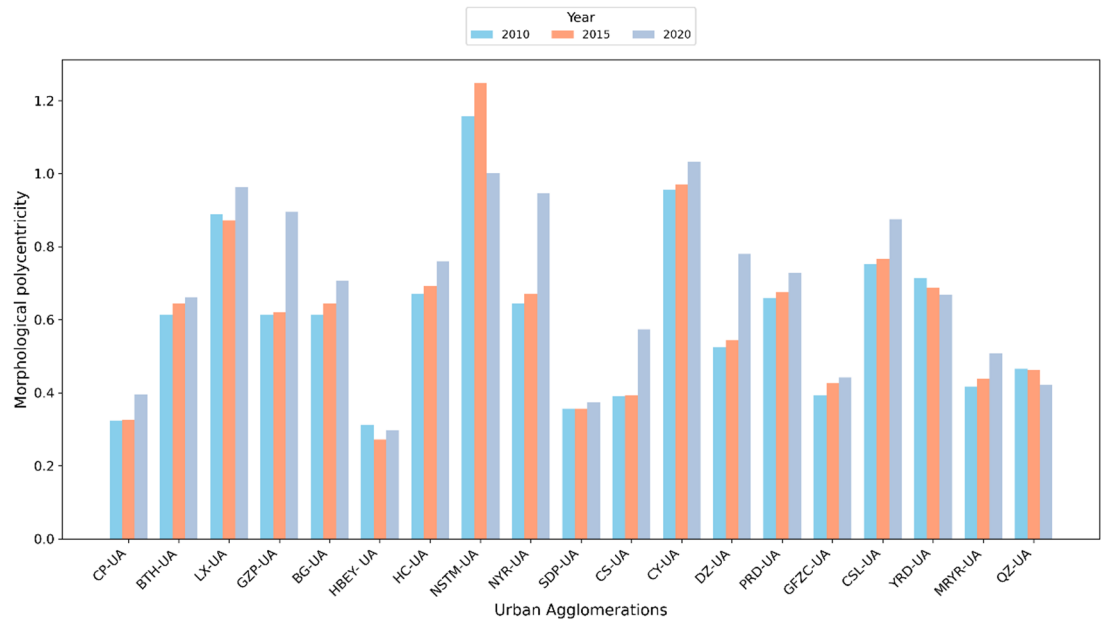


Fig. 10. Morphological polycentricity of the 19 urban agglomerations.

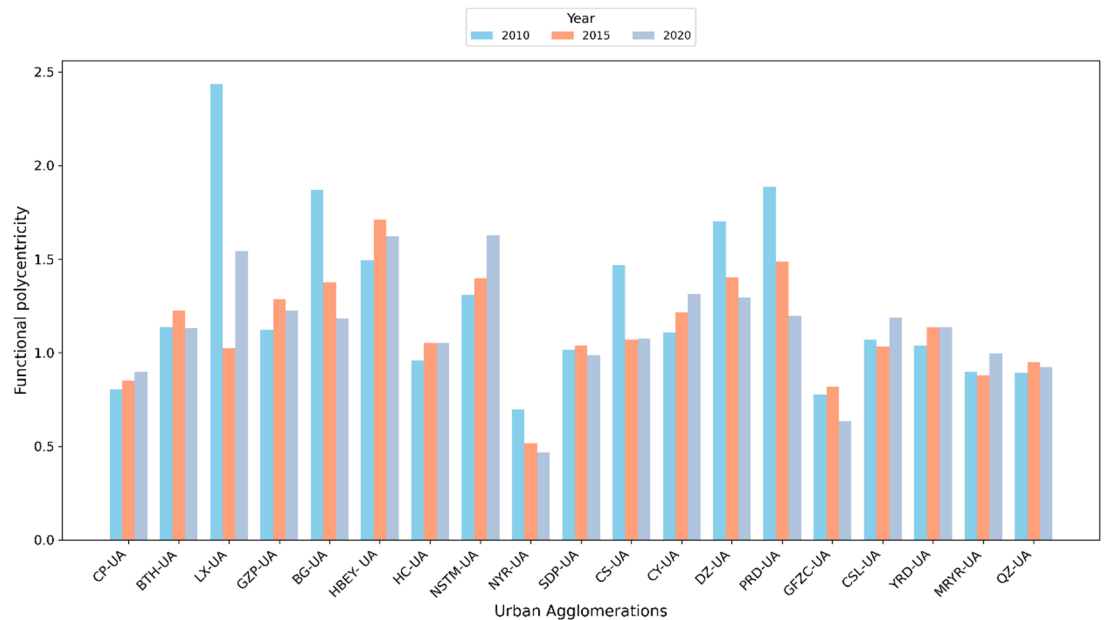


Fig. 11. Functional polycentricity of 19 urban agglomerations.

of the core city is more prominent. Both the NYR-UA and the GFZC-UA have lower functional polycentricity, reflecting that the distribution of functions in these urban agglomerations is more balanced, and they show a certain level of functional polycentric structures. However, even so, the value of functional polycentricity is still higher than that of morphological polycentricity, which further indicates that the concentration of functional elements is relatively more obvious within urban agglomerations. The functional polycentricity of the PRD-UA, BG-UA, and DZ-UA decreases over time, indicating a gradual decentralization of functional elements toward multiple cities and a more balanced functional allocation across the urban agglomerations. This trend suggests increasingly evident functional coordination within these urban agglomerations. On the contrary, the functional polycentricities of the NSTM-UA, CY-UA, and CP-UA increase year by year, which indicates that the degree of functional concentration of these urban agglomerations is increasing, and their monocentric structure is becoming more prominent, with the core city's functional dominance being further strengthened.

As is shown in Fig. 12, the weighted density within urban agglomerations has shown a consistent upward trend over the years, indicating a significant enhancement in intercity connectivity. However, notable regional

disparities persist. The BTH-UA, PRD-UA, and YRD-UA exhibit significantly higher weighted densities compared to other urban agglomerations, with the fastest growth rates, reflecting strong intercity linkages and a high level of regional integration. The SDP-UA, CY-UA, MRYR-UA, HC-UA, and GZP-UA form the second tier, where network densities are steadily increasing, indicating a gradual strengthening of intercity connections within these regions. In contrast, urban agglomerations in the northwest, such as NSTM-UA and NYR-UA, exhibit relatively low densities and slower growth, reflecting a lag in regional development and highlighting the need for improved intercity connectivity.

As is illustrated in Fig. 13, urban agglomerations such as BTH-UA, SDP-UA, CS-UA, PRD-UA, CSL-UA, and YRD-UA consistently exhibit high weighted clustering coefficients in 2010, 2015, and 2020, indicating strong local connectivity and intensive interactions among neighboring cities within these agglomerations. These urban agglomerations demonstrate high cohesion, reflecting trends of collaboration and centralization. In contrast, the weighted clustering coefficients of the LX, CP, DZ, CY and NSTM urban agglomerations are low, indicating weak cohesion, loose overall connections, and a lack of strong collaborative relationships. The network structures of these urban agglomerations require further optimization.

Disparity quantifies the overall uneven distribution of edge weights across all nodes in the network⁵³. In the context of urban agglomerations, network disparity quantifies the uneven distribution of these connections' strengths (edge weights) across all cities in the urban agglomeration. Higher disparity indicates that connections are concentrated on a few dominant edges, meaning that certain cities have disproportionately strong ties with specific cities, while other connections remain weak. The network disparity of each urban agglomeration shows some volatility and regional differences in different years (Fig. 14). The NYR-UA exhibits the lowest network disparity, with a continuous decline over time, indicating that the connection strength among cities within the agglomeration has become more evenly distributed. The CP-UA, QZ-UA, BTH-UA, MRYR-UA, and GZP-UA show minimal changes in network disparity, reflecting a relatively stable connectivity structure within these regions. In contrast, the BG-UA, PRD-UA, and CSL-UA exhibit a significant decline in network disparity, likely driven by improved connection strength among peripheral cities. On the other hand, the HC-UA, YRD-UA, and CY-UA demonstrate a notable increase in network disparity, which may result from the strengthening of connections among core cities while the connectivity among peripheral cities remains insufficiently improved.

Results of panel data regression and robustness analysis

Results of panel data regression

Before applying panel data regression, logarithmic transformation is performed on both independent and dependent variables to mitigate heteroskedasticity⁶⁸. Subsequently, the variables are standardized to eliminate dimensional differences, facilitate comparison between variables, and enhance the interpretability of regression coefficients. Based on the results of the Hausman test ($\chi^2 = 4.339$, $p = 0.502$), the random effects model was employed for the regression analysis (Sect. 3.3) and the results are reported in Table 3. The goodness of fit for the model is $R^2=0.76$, indicating that 76% of the variation in carbon emissions is explained by the independent variables.

$$CE = -0.26MP - 0.13FP - 0.01WD + 0.58UAS + 0.15ND - 0.05WC + 0.05IS + \epsilon \quad (13)$$

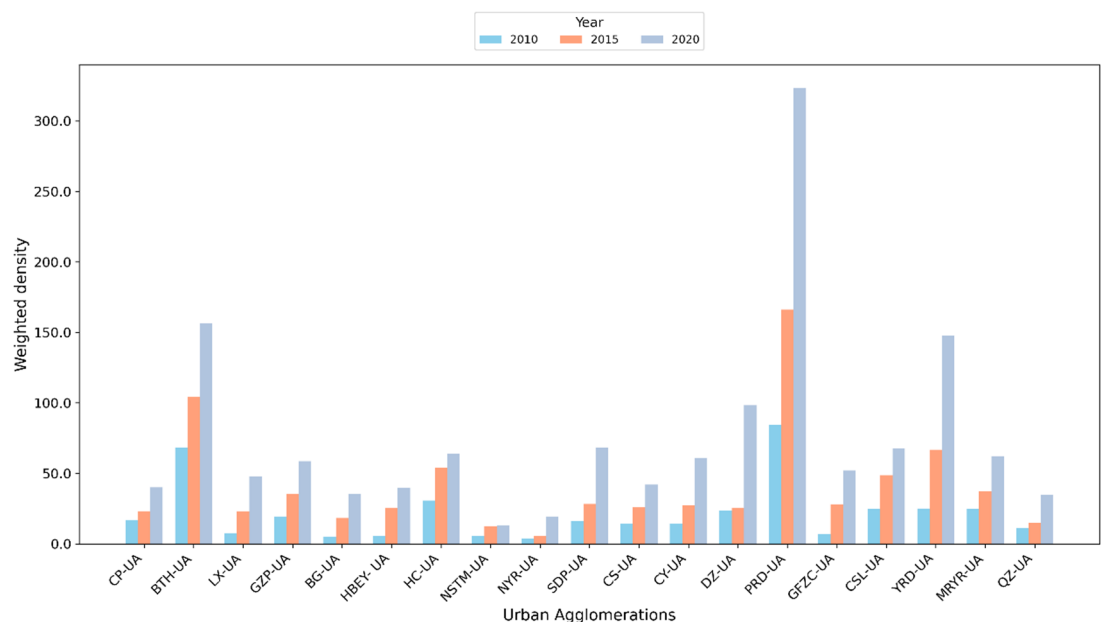


Fig. 12. Weighted density of the 19 urban agglomerations.

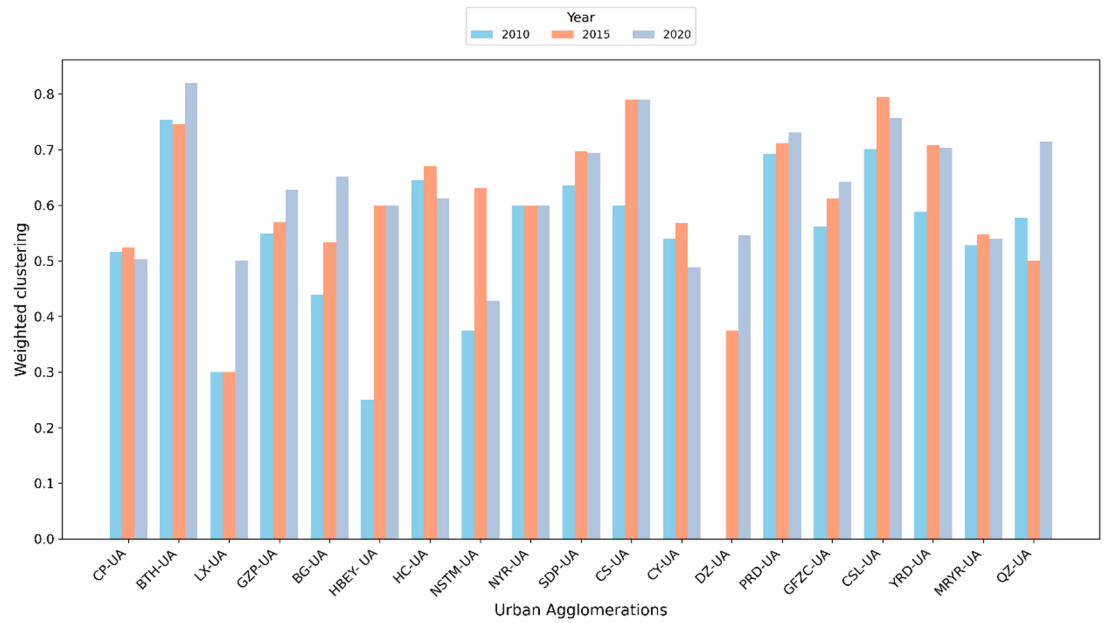


Fig. 13. Weighted clustering coefficient of the 19 urban agglomerations.

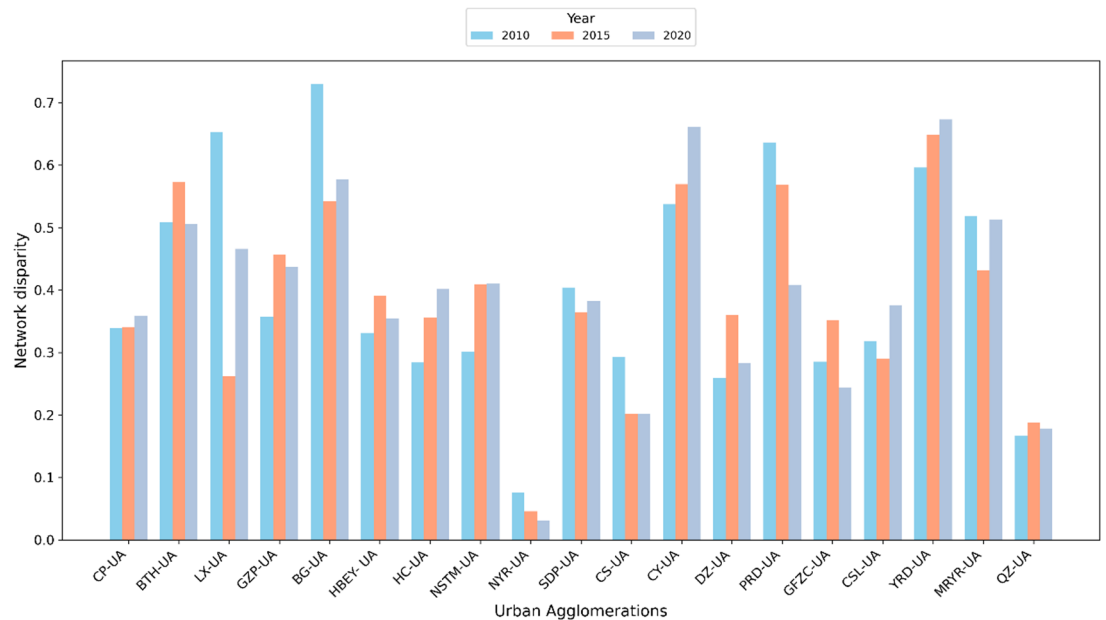


Fig. 14. Network disparity of 19 urban agglomerations.

The results from the panel data regression show that the coefficient for Urban agglomeration size (UAS) is 0.58, which is statistically significant at the 1% level ($p < 0.01$), indicating that larger urban agglomerations are associated with increased carbon emissions. The coefficient of Network Disparity (ND) is 0.15 and is statistically significant at the 5% level ($p < 0.05$), suggesting that greater network disparity is associated with higher carbon emissions. In contrast, both Morphological Polycentricity (MP) and Functional Polycentricity (FP) have negative coefficients of -0.26 and -0.13, respectively, and are statistically significant at the 5% level ($p < 0.05$). The coefficients for Weighted Clustering (WC), Industrial Structure (IS), and Weighted Density (WD) are -0.05, 0.05, and -0.01, respectively, but none of these factors pass the significance tests ($p > 0.05$), suggesting that their effects on carbon emissions are not statistically significant.

Results of robustness analysis

To further validate the robustness of our regression results, we employed a random forest model enhanced with SHAP for interpretability. It is important to note that the independent and dependent variables in the random

Indicators	Coefficient	0.05	0.95	0.025	0.975	0.005	0.995
Morphological polycentricity	-0.26**	-0.43	-0.09	-0.47	-0.06	-0.53	0.01
Functional polycentricity	-0.13**	-0.23	-0.03	-0.25	-0.01	-0.29	0.03
Weighted density	-0.01	-0.21	0.18	-0.25	0.22	-0.33	0.30
Urban agglomeration size	0.58***	0.38	0.78	0.34	0.82	0.27	0.89
Network disparity	0.15**	0.03	0.27	0.01	0.29	-0.04	0.33
Weighted clustering	-0.05	-0.13	0.03	-0.14	0.04	-0.17	0.07
Industrial structure	0.05	-0.07	0.17	-0.09	0.19	-0.14	0.24

Table 3. The results of panel data regression. ** Significant at 0.05, *** Significant at 0.01. Substituting the coefficients, the regression model is specified as follows:

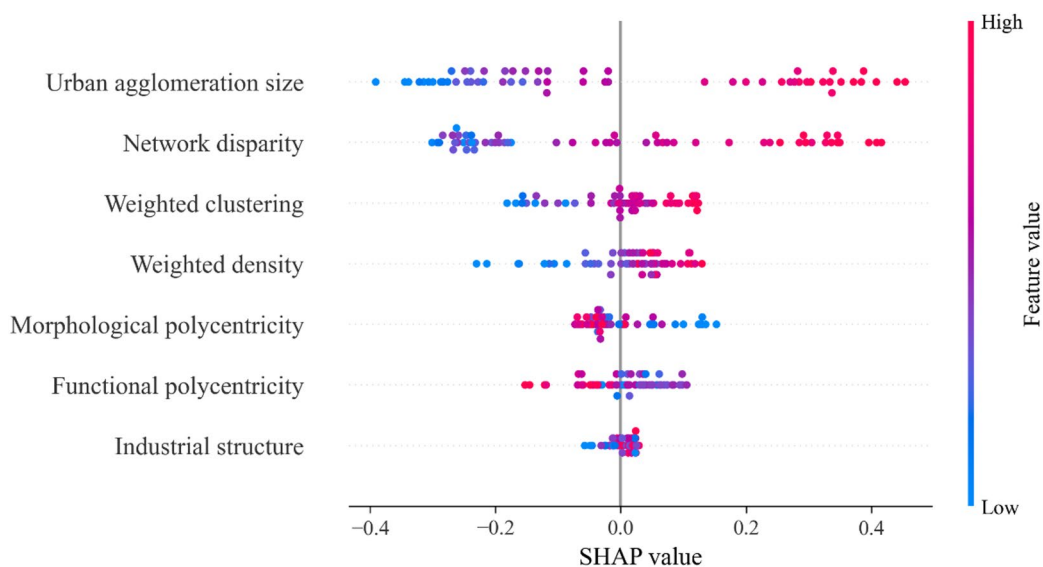


Fig. 15. SHAP impact of urban agglomeration spatial structures on carbon emissions.

forest model are consistent with those used in the panel data regression analysis. The random forest model consisted of 800 decision trees, with the number of features at each split determined by the “sqrt” method. Out-Of-Bag (OOB) validation was incorporated, and a logarithmic transformation was applied to the variables to improve the model’s robustness. The OOB R^2 score of 0.72 indicates that the model effectively captures the relationship between the independent and dependent variables.

Figure 15 presents the SHAP impact of various urban spatial structure features on carbon emissions. The SHAP values represent the contribution of each feature to the model’s prediction of carbon emissions. The x-axis represents the SHAP values for each feature, where positive values indicate a positive contribution to carbon emissions and negative values indicate a negative contribution. The y-axis lists the different features, such as “urban agglomeration size,” “network disparity,” “weighted clustering,” and “weighted density,” reflecting their impact on carbon emissions. Each point represents a sample’s feature value and its corresponding SHAP value, showing the individual contribution of the feature for each sample. The color of the points ranges from blue to red, representing the range of feature values, where blue indicates lower values and red indicates higher values.

As is illustrated in Fig. 15, urban agglomeration size shows a clear positive impact on carbon emissions, with the highest contribution. As urban agglomeration size increases, the positive contribution to carbon emissions becomes more significant, making this feature the most influential in the model. Network disparity also exhibits a positive impact, with an increase in network disparity leading to higher carbon emissions. Similarly, both weighted clustering and weighted density show a positive impact, indicating that higher clustering and density in urban areas are associated with increased carbon emissions. In contrast, morphological polycentricity and functional polycentricity both present negative impacts, suggesting that a more polycentric urban form is linked to reduced carbon emissions. Morphological polycentricity has a slightly stronger negative impact compared to functional polycentricity. Regarding industrial structure, the points are clustered near the median line, suggesting that this feature has a limited impact on carbon emissions. This indicates that industrial structure has a relatively limited contribution to the model’s predictions.

Discussion

Monocentric or polycentric?

The results from the panel data regression model and robustness analysis indicate that both morphological polycentricity and functional polycentricity have a negative impact on carbon emissions. Notably, as the values of these indicators increase, the urban agglomeration structure shifts towards a monocentric configuration, moving away from a polycentric structure. Therefore, our study reveals that both morphological and functional polycentric structures contribute to increased carbon emissions in urban agglomerations which differs greatly from existing finds at individual cities⁴⁸. At city scales, polycentric structures can disperse population and economic functions, thereby shortening travel distances, reducing transportation demand, and lowering fossil fuel consumption, ultimately decreasing carbon emissions⁶⁹. However, at urban agglomeration scales, the situation is different. Our findings suggest that the monocentric structures, characterized by the concentration of GDP, population, and industries in a core city, has a suppressive effect on carbon emissions. This suppressive effect may be attributed to the following factors:

- (1) **Industrial Agglomeration.** In monocentric urban agglomerations, firms are concentrated in specific cities, achieving economies of scale through division of labor and collaboration. This agglomeration reduces unit production costs, minimizes transportation energy consumption, and improves energy efficiency, thereby reducing carbon emissions to some extent.
- (2) **Agglomeration-Induced Green Technological Innovation.** Industrial agglomeration accelerates the dissemination of knowledge and technology among firms, fostering innovation spillover effects. Particularly in high-tech industrial clusters, learning and cooperation among firms promote technological upgrades and innovation. The adoption of advanced technologies contributes to reducing carbon emissions.
- (3) **Infrastructure Sharing.** Within agglomerated areas, firms and residents can share public infrastructure (e.g., transportation, communication, and electricity), effectively lowering individual usage costs and improving resource utilization efficiency, which in turn reduces energy consumption and carbon emissions.

However, agglomeration can also exacerbate carbon emissions. For instance, over-concentration in core cities often leads to traffic congestion, prolonging travel times and increasing transportation-related carbon emissions, thereby undermining the environmental benefits of agglomeration effects. In conclusion, while monocentric structures at the urban agglomeration scale exhibit certain potential for carbon reduction, agglomeration can also promote carbon emissions under specific conditions. The relationship is complex and warrants further empirical research to gain deeper insights.

Morphological versus functional polycentric

The importance of cities is determined not just by their absolute size, but more significantly by how they are connected with other cities⁷⁰. Based on this recognition, we define the importance of cities through their network connections with other cities and quantify the spatial distribution (or degree of agglomeration) of this importance, which we call functional polycentricity. Simultaneously, we define city importance based on absolute size indicators, and quantify its spatial distribution with the same approach, which we refer to as morphological polycentricity.

We further examined the impact of functional and morphological polycentricity on carbon emissions through panel data regression analysis. The results show that the absolute value of the coefficient for morphological polycentricity (-0.26^{***}) is slightly higher than that of functional polycentricity (-0.13^{***}). This finding suggests that, in terms of carbon emissions, the effect of morphological polycentricity is stronger than that of functional polycentricity. The robustness analysis confirms this result, indicating that the importance of morphological polycentricity slightly outweighs that of functional polycentricity, with both having a negative impact.

These results indicate that, at least in terms of carbon emissions, the significance of cities is more dependent on their absolute size.

Network structure and carbon emissions

From the perspective of resource and factor flows, the structures of urban agglomeration networks essentially determine, or at least reflect, how labor, capital, and other resources circulate among cities. The configuration of such flows has profound implications for carbon emissions, yet the effects are inherently complex and often involve trade-offs. For example, stronger intercity flows may increase transportation-related emissions, but at the same time they can enhance knowledge exchange and coordination, thereby improving production efficiency and reducing industrial energy consumption. Similar trade-offs may arise when factor agglomeration reduces duplicated construction and promotes economies of scale, while factor dispersion can extend production chains and increase logistics demand. Quantifying the exact magnitudes of these opposing effects is extremely difficult. Therefore, in this study we treat the internal mechanisms as a “black box” and focus on the net impacts, using explainable machine learning models and regression analysis to reveal how changes in network structures ultimately affect carbon emissions.

- **The Insignificant Impact of Clustering Coefficient and Weighted Density.** The clustering coefficient measures the local adjacency structure of nodes within a network, reflecting the strength of connectivity between cities in a local area. Weighted density represents the connectivity and strength of the network, indicating the extent and intensity of connections between cities. In the panel data regression, neither of these factors showed a significant impact on carbon emissions. For the weighted clustering coefficient, an increase in its value signifies a tighter intercity connection. Similarly, an increase in weighted density indicates an

enhancement in the strength of intercity linkages. While intuitively, their increase might lead to higher energy consumption in transportation between cities, as mentioned earlier, stronger intercity flow could indeed increase emissions related to transportation. However, it also enhances knowledge exchange and collaboration, improving production efficiency and reducing industrial energy consumption. As a result, the overall impact on carbon emissions is complex.

- **The Positive Impact of Network Disparity on Carbon Emissions.** Network disparity quantifies the distributional balance of edges within the network; a higher disparity indicates more uneven intercity connections. The positive coefficient (0.15**) reveals that network disparity has a significant positive impact on carbon emissions. Notably, although network disparity and node centrality share structural similarities in measuring the network, their relationships with carbon emissions are fundamentally opposite. While node centrality focuses on the number and strength of connections associated with nodes, disparity highlights the uneven distribution of edges. When strong connections exist among only a few cities, while others remain weakly connected (resulting in higher disparity), regional development becomes imbalanced, limiting the aggregation effects and exacerbating carbon emissions.

Policy recommendations

In urban agglomerations, concentrating key elements and resources in core cities is recommended as it can create agglomeration effects, thereby reducing carbon emissions. Policies should support the concentration of traffic, resources, and industrial configurations in core cities to maximize their potential for carbon reduction.

However, the importance of maintaining strong connections between smaller cities should not be overlooked. While the concentration of functions and forms in core cities can effectively reduce carbon emissions, overreliance on a few strong connections may lead to excessive network disparity. To address this, it is crucial to strengthen the links between core cities and surrounding smaller and peripheral cities, ensuring that weaker cities maintain close communication with the core cities. This approach will help balance regional development and ultimately reduce overall carbon emissions.

Furthermore, although tight and strong inter-city exchanges may be perceived as potentially increasing carbon emissions, their spillover effects (such as the flow of information and knowledge) could potentially counteract the associated negative impacts. Until strong evidence suggests otherwise, inter-city exchanges should not be hindered.

Conclusion

Previous studies mainly characterized urban agglomeration spatial structures through attribution data of cities; differently, this study innovatively regards urban agglomerations as spatial networks of cities, and expresses their spatial structures by using inter-city linkage data. In addition, to the best of our knowledge, this study is the first to investigate and compare the impacts of functional polycentricity and morphological polycentricity of urban agglomerations on carbon emissions. Accordingly, this study contributes not only to urban agglomerations planning but also to urban research community. Three conclusions can be drawn from this study.

Firstly, monocentric spatial structure of urban agglomerations contributes to reducing carbon emissions. In urban agglomeration planning, efforts should focus on enhancing the concentration of resources in core cities, promoting the development of monocentric structures, and improving resource allocation efficiency and agglomeration effects to effectively achieve carbon reduction goals.

Secondly, morphological polycentricity has a greater impact on carbon emissions than functional polycentricity, providing a new perspective for urban research community.

Thirdly, network disparity positively influences carbon emissions. Therefore, while strengthening the connectivity between core cities, it is also important to promote the coordinated development of peripheral cities and enhance overall network connectivity within urban agglomerations, thereby reducing network disparity and fostering regional coordination and low-carbon development.

This study primarily relied on passenger train flow data, but future research could expand the scope by exploring more comprehensive big data sources, incorporating other types of flows such as freight, information, financial, and road traffic flows. This would offer a more holistic understanding of urban agglomeration spatial structures and their impact on carbon emissions. Additionally, in the context of investigating the impact of urban agglomeration spatial structure on carbon emissions, the system operates at a large scale, involves numerous structural elements, and exhibits highly complex spatial interactions. The lack of ideal exogenous shock settings that satisfy the strict independence assumptions required for causal identification further poses substantial methodological and data constraints—challenges widely recognized in the international literature. Therefore, the conclusions should remain cautious and prudent with respect to causal interpretation. Future work will explore methodologically rigorous causal analysis to strengthen causal inference and mechanism validation of spatial structure on carbon emissions, thereby offering more credible evidence to better support policy-making for low-carbon urban development.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 22 May 2025; Accepted: 10 March 2026

Published online: 27 March 2026

References

- Tang, Z., Wang, Y., Fu, M. & Xue, J. The role of land use landscape patterns in the carbon emission reduction: Empirical evidence from China. *Ecol. Indic.* **156**, 111176 (2023).
- Xia, C. et al. Outsourced carbon mitigation efforts of Chinese cities from 2012 to 2017. *Nat. Cities*. **1**, 480–488 (2024).
- Dhakal, S. Urban energy use and carbon emissions from cities in China and policy implications. *Energy Policy*. **37**, 4208–4219 (2009).
- Dhakal, S. GHG emissions from urbanization and opportunities for urban carbon mitigation. *Curr. Opin. Environ. Sustain.* **2**, 277–283 (2010).
- Fang, C. & Yu, D. *China's Urban Agglomerations* (Springer, 2020).
- Xia, C. et al. Heterogeneity in carbon footprint trends and trade-induced emissions in China's urban agglomerations. *Commun. Earth Environ.* **6**, 723 (2025).
- Gottmann, J. Megalopolis or the urbanization of the northeastern seaboard. *Econ. Geogr.* **33**, 189–200 (1957).
- Fang, C. & Yu, D. Urban agglomeration: An evolving concept of an emerging phenomenon. *Landsc. Urban Plan.* **162**, 126–136 (2017).
- Wang, Y., Niu, Y., Li, M., Yu, Q. & Chen, W. Spatial structure and carbon emission of urban agglomerations: Spatiotemporal characteristics and driving forces. *Sustain. Cities Soc.* **78**, 103600 (2022).
- Portnov, B. A. Urban clustering, development similarity, and local growth: A case study of Canada. *Eur. Plann. Stud.* **14**, 1287–1314 (2006).
- Fang, C., Yu, X., Zhang, X., Fang, J. & Liu, H. Big data analysis on the spatial networks of urban agglomeration. *Cities* **102**, 102735 (2020).
- Zhang, B. & Yin, J. Exploring the impact of spatial structure on carbon emissions in Chinese urban agglomerations: Insights into polycentric and compact development patterns. *Urban Clim.* **62**, 102557 (2025).
- Liu, Y. et al. Quantitative structure and spatial pattern optimization of urban green space from the perspective of carbon balance: A case study in Beijing, China. *Ecol. Indic.* **148**, 110034 (2023).
- Liu, P., Zhong, F. & Han, N. Efficiency and equity: Effect of urban agglomerations' spatial structure on green development efficiency in China. *Sustain. Cities Soc.* **108**, 105504 (2024).
- Tan, G. et al. Assessing the impacts of urban functional form on anthropogenic carbon emissions: A case study of 31 major cities in China. *Ecol. Indic.* **167**, 112700 (2024).
- Liu, K., Xue, M., Peng, M. & Wang, C. Impact of spatial structure of urban agglomeration on carbon emissions: An T analysis of the Shandong Peninsula, China. *Technol. Forecast. Soc. Chang.* **161**, 120313 (2020).
- Hong, S., Hui, E. C. & Lin, Y. Relationship between urban spatial structure and carbon emissions: A literature review. *Ecol. Indic.* **144**, 109456 (2022).
- Wu, Z., Woo, S. H., Oh, J. H. & Lai, P. L. Temporal and spatial effects of manufacturing agglomeration on CO₂ emissions: evidence from South Korea. *Hum. Soc. Sci. Commun.* **11**, 790 (2024).
- Lu, H., Yao, Z., Cheng, Z. & Xue, A. The impact of innovation-driven industrial clusters on urban carbon emission efficiency: Empirical evidence from China. *Sustain. Cities Soc.* **121**, 106220 (2025).
- Dai, L. & Luo, J. Effects of spatial structure on carbon emissions of urban agglomerations in China. *Cities* **163**, 106021 (2025).
- Yang, Y., Zhao, Y. & Zhang, Y. The divergent effects of spatial structure of urban agglomerations on carbon emission reduction capacity. *Urban Clim.* **63**, 102602 (2025).
- Cao, F., Qiu, Y., Wang, Q. & Zou, Y. Urban form and function optimization for reducing carbon emissions based on crowd-sourced spatio-temporal data. *Int. J. Environ. Res. Public Health.* **19**, 10805 (2022).
- Wang, S., Wang, J., Fang, C. & Li, S. Estimating the impacts of urban form on CO₂ emission efficiency in the Pearl River Delta, China. *Cities* **85**, 117–129 (2019).
- Burgalassi, D. & Luzzati, T. Urban spatial structure and environmental emissions: A survey of the literature and some empirical evidence for Italian NUTS 3 regions. *Cities* **49**, 134–148 (2015).
- Bartosiewicz, B. & Marcińczak, S. Investigating polycentric urban regions: different measures—different results. *Cities* **105**, 102855 (2020).
- Xiao, Y., Yang, H., Kong, Q. & Chen, Y. Can the polycentric spatial structure achieve pollution and carbon emission reduction in China's urban agglomerations? *J. Environ. Manag.* **392**, 126732 (2025).
- Lan, F., Da, H., Wen, H. & Wang, Y. Spatial structure evolution of urban agglomerations and its driving factors in mainland China: From the monocentric to the polycentric dimension. *Sustainability* **11**, 610 (2019).
- Chen, X., Zhang, S. & Ruan, S. Polycentric structure and carbon dioxide emissions: Empirical analysis from provincial data in China. *J. Clean. Prod.* **278**, (2021).
- Xu, C., Bin, Q. & Shaoqin, S. Polycentric spatial structure and energy efficiency: Evidence from China's provincial panel data. *Energy Policy*. **149**, 112012 (2021).
- Meijers, E. J. & Burger, M. J. Spatial structure and productivity in US metropolitan areas. *Environ. Plann. A.* **42**, 1383–1402 (2010).
- Fujita, M. & Thisse, J. F. *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth* (Cambridge University Press, 2002).
- Batten, D. F. Network cities: creative urban agglomerations for the 21st century. *Urban Stud.* **32**, 313–327 (1995).
- Beverstock, J. V., Smith, R. G. & Taylor, P. J. World-city network: a new metageography? *Ann. Assoc. Am. Geogr.* **90**, 123–124 (2000).
- Anderson, J. E. The gravity model. *Annu. Rev. Econ.* **3**, 133–160 (2011).
- Burger, M. & Meijers, E. Form Follows Function? Linking Morphological and Functional Polycentricity. *Urban Stud.* **49**, 1127–1149 (2012).
- Yu, H., Yang, J., Li, T., Jin, Y. & Sun, D. Morphological and functional polycentric structure assessment of megacity: An integrated approach with spatial distribution and interaction. *Sustain. Cities Soc.* **80**, 103800 (2022).
- Wang, S. H., Huang, S. L. & Huang, P. J. Can spatial planning really mitigate carbon dioxide emissions in urban areas? A case study in Taipei, Taiwan. *Landsc. Urban Plan.* **169**, 22–36 (2018).
- Oda, T. & Maksyutov, S. A very high-resolution (1 km×1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmos. Chem. Phys.* **11**, 543–556 (2011).
- Oda, T. et al. October. Uncertainty associated with fossil fuel carbon dioxide (CO₂) gridded emission datasets. in *Proceedings, 4th International Workshop on Uncertainty in Atmospheric Emissions. 7–9 Krakow, Poland*, pp. 124–129 (Systems Research Institute, Polish Acad. Sci., 2015). (2015).
- Oda, T., Maksyutov, S. & Andres, R. J. The Open-source Data Inventory for Anthropogenic CO₂, version 2016 (ODIAC2016): a global monthly fossil fuel CO₂ gridded emissions data product for tracer transport simulations and surface flux inversions. *Earth Syst. Sci. Data.* **10**, 87–107 (2018).
- Lao, X., Zhang, X., Shen, T. & Skitmore, M. Comparing China's city transportation and economic networks. *Cities* **53**, 43–50 (2016).
- Yang, H., Du, D., Wang, J., Wang, X. & Zhang, F. Reshaping China's urban networks and their determinants: High-speed rail vs. air networks. *Transp. Policy.* **143**, 83–92 (2023).
- Tatem, A. J. WorldPop, open data for spatial demography. *Sci. Data.* **4**, 1–4 (2017).

44. Huang, Y. & Zong, H. The intercity railway connections in China: A comparative analysis of high-speed train and conventional train services. *Transp. Policy*. **120**, 89–103 (2022).
45. Belik, V., Geisel, T. & Brockmann, D. Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases. *Phys Rev. X* **1**, (2011).
46. Simini, F., González, M. C., Maritan, A. & Barabási A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
47. Jung, W. S., Wang, F. & Stanley, H. E. Gravity model in the Korean highway. *Europhys. Lett.* **81**, 48005 (2008).
48. Veneri, P. Urban Polycentricity and the Costs of Commuting: Evidence from Italian Metropolitan Areas. *Growth Change*. **41**, 403–429 (2010).
49. Sha, W., Chen, Y., Wu, J. & Wang, Z. Will polycentric cities cause more CO₂ emissions? A case study of 232 Chinese cities. *J. Environ. Sci.* **96**, 33–43 (2020).
50. Green, N. Functional polycentricity: A formal definition in terms of social network analysis. *Urban Stud.* **44**, 2077–2103 (2007).
51. Newman, M. E. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
52. Rosen, K. T. & Resnick, M. The size distribution of cities: an examination of the Pareto law and primacy. *J. Urban Econ.* **8**, 165–186 (1980).
53. Serrano, M. Á., Boguná, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. U S A.* **106**, 6483–6488 (2009).
54. Guérin-Pace, F. Rank-size distribution and the process of urban growth. *Urban Stud.* **32**, 551–562 (1995).
55. Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).
56. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U S A.* **101**, 3747–3752 (2004).
57. Goswami, S., Murthy, C. A. & Das, A. K. Sparsity measure of a network graph: Gini index. *Inf. Sci.* **462**, 16–39 (2018).
58. Foster, J. E. An axiomatic characterization of the Theil measure of income inequality. *J. Econ. Theory*. **31**, 105–121 (1983).
59. Jiang, J., Zhu, S., Gao, S., Aslam, B. & Wang, W. Impact of energy and industrial structure on environmental quality and urbanization: evidence from a panel of BRICS countries. *Environ. Sci. Pollut Res.* **30**, 114183–114200 (2023).
60. Xu, S. C., He, Z. X. & Long, R. Y. Factors that influence carbon emissions due to energy consumption in China: Decomposition analysis using LMDI. *Appl. Energy*. **127**, 182–193 (2014).
61. Baltagi, B. H., Song, S. H. & Koh, W. Testing panel data regression models with spatial error correlation. *J. Econom.* **117**, 123–150 (2003).
62. Du, L., Wei, C. & Cai, S. Economic development and carbon dioxide emissions in China: Provincial panel data analysis. *China Econ. Rev.* **23**, 371–384 (2012).
63. Al-mulali, U. Factors affecting CO₂ emission in the Middle East: A panel data analysis. *Energy* **44**, 564–569 (2012).
64. Fang, C., Wang, S. & Li, G. Changing urban forms and carbon dioxide emissions in China: A case study of 30 provincial capital cities. *Appl. Energy*. **158**, 519–531 (2015).
65. Chamberlain, G. Multivariate regression models for panel data. *J. Econom.* **18**, 5–46 (1982).
66. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760 (2018).
67. Lundberg, S. M. & Lee, S. I. A Unified Approach to Interpreting Model Predictions. in *Advances Neural Inform. Process. Systems* **30** (2017).
68. Astivia, O. L. O. & Zumbo, B. D. Heteroskedasticity in Multiple Regression Analysis: What it is, How to Detect it and How to Solve it with Applications in R and SPSS. *Practical Assess. Res. Evaluation*. **24**, n1 (2019).
69. Glaeser, M. E. & Kahn, E. L. The greenness of cities: Carbon dioxide emissions and urban development. *J. Urban Econ.* **67**, 404–418 (2010).
70. Taylor, P. J. Urban economics in thrall to Christaller: A misguided search for city hierarchies in external urban relations. *Environ. Plan. A.* **41**, 2550–2555 (2009).

Acknowledgements

The work presented in this paper is funded by the National Natural Science Foundation of China (No. 42330103) and the Ningbo Science and Technology Bureau (No.2022Z081).

Author contributions

G.T. conceived the study, designed the research framework and drafted the original manuscript. S.X. and X.Z. gathered and pre-processed the multi-source datasets. Z.S. and H.W. developed the analytical pipeline, performed the formal analyses and generated the visualisations. Y.L. reviewed the manuscript. S.D. supervised the project, secured funding and provided critical revisions. All authors discussed the results, reviewed and approved the final manuscript, and agree to be accountable for all aspects of the work.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.D.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026