scientific data

OPEN

DATA DESCRIPTOR

An intercity investment network dataset of China based on the enterprise registration records (2000–2020)

Tianheng Shu[1,9], Shuo Yang[1,9], Taofang Yu[1]✉, Guangyu Cheng[2], Yitian Ren[3], Fangchen Shi[4], Ben Derudder[5,6,7] & Xia Liao[8]✉

Intercity investment activities among enterprises reflect the flow of capital between cities, thereby directly illustrating the economic connections between them. However, there is currently no publicly available dataset that captures this important feature. In this study, we introduce an intercity investment network (IIN) dataset for China, covering the period from 2000 to 2020, based on 17,273,411 large-scale enterprise registration records. The dataset represents 367 cities as nodes, with investment frequency between cities serving as edge weights to construct both directed and undirected networks. It captures the spatiotemporal patterns of China's IIN, highlighting dynamic changes in economic connectivity over time and space. The dataset aligns closely with urban networks formed by China's population mobility and the economic gravity model, is consistent with official records and existing research findings, and satisfies the distance decay effect, thus validating its scientific reliability. This dataset provides unique opportunities for exploring economic interactions and functional organization between cities, and advancing urban network research in China.

Background & Summary

As economic globalization and informatization progress, cities are becoming increasingly interconnected through the flow of people, information, goods, and capital, etc[1–3]. The emergence of the "space of flows" has shifted urban research from hierarchical systems toward a network-based approach that emphasizes spatial interactions between cities[4–6]. Enterprises, as key drivers of urban economic growth, play a crucial role in shaping economic connections between cities[2,7,8]. Constructing networks based on enterprise-enterprise linkage has become a major focus in recent urban network research. During China's rapid urbanization these decades, industries have expanded significantly, and a large number of enterprises have been established annually[9,10], forming increasingly close ties that reshape economic interactions and functional organizations between cities. To capture this dynamic process, there is an urgent need for accurate data to quantitatively describe the intercity network of economic connections, providing a scientific foundation for further research and policy development.

Traditional studies on urban networks have primarily relied on data such as population mobility, transportation flows, technology/knowledge exchanges, and logistics[11–17]. While valuable, these sources often overlook actual economic flows like capital investments. Current approaches for constructing urban networks based on enterprise data generally fall into several categories. One method uses interlocking network model, which establishes a service value matrix between cities to construct urban networks[6,18]. This method, although widely

[1]School of Architecture, Tsinghua University, Beijing, 100084, China. [2]Department of Real Estate and Construction, Faculty of Architecture, The University of Hong Kong, Hong Kong, 999077, China. [3]Department of Urban Planning and Design & Spatial Analysis and Policy Lab, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China. [4]School of Government, Peking University, Beijing, 100871, China. [5]Public Governance Institute, KU Leuven, Leuven, 3000, Belgium. [6]Department of Geography, Ghent University, Ghent, 9000, Belgium. [7]Department of Urban Studies and Sustainable Development, Nicolaus Copernicus University, Torun, 87-100, Poland. [8]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China. [9]These authors contributed equally: Tianheng Shu, Shuo Yang. ✉e-mail: yutaofang@mail.tsinghua.edu.cn; liaoxia_cq@163.com

used in world city network research, is complex and does not directly capture the city-to-city connections[19,20]. Another widely used approach focuses on the headquarter-branch relationship, emphasizing the control exerted by headquarters over branch locations in different cities[19,21–24]. While this approach effectively presents vertical and intra-enterprise connections, it is limited in scope as it focuses on enterprises with headquarters, neglecting the broader spectrum of intercity economic connections.

Recently, constructing IIN based on inter-enterprise investment activities has gained traction. These activities typically refer to cross-regional investments, where capital flows from one city to another[2]. By analyzing intercity investment behaviors and equity relationships, cities can establish complex capital connection networks that represent real economic ties. Inter-enterprise investments reflect not just capital flows, but also the transfer of information, technology, equipment, and human resources, etc[10,20]. As such, enterprise investment data provide a more accurate picture of intercity economic linkages. Several empirical studies have explored IIN, particularly in China. For example, *Li et al.*[2] analyzed the spatial patterns and influencing factors of the IIN in the Yangtze River Delta based on historical investment data from 3,698 listed firms. Similarly, *Guo et al.*[25] constructed an intercity investment network for 53 coastal cities in China from 2006 to 2016. Other studies, such as *Lu & Sun*[26], constructed national-scale networks based on headquarters-branch and inter-enterprise investments, while *Zhang et al.*[20] focused on the influence of institutions and markets on the network structure using large-scale enterprise investment records. Additionally, some studies have explored venture capital networks in Chinese cities[16,27].

Despite the increasing recognition of the importance of IIN in China, existing research still faces several limitations, primarily due to data availability and quality. Many studies either focus on developed regions such as the Yangtze River Delta, which has a high degree of marketization[28], or on specific groups such as listed companies or venture capital networks, as these datasets are more easily accessible. Even the very few studies that have constructed urban networks based on large-scale enterprise investment data often lack a focus on the data itself, do not include validation, and fail to make the data publicly available. Moreover, the limited temporal scope of existing datasets restricts the ability to depict the long-term structural dynamics of IIN, especially in rapidly urbanizing regions like China. In reality, a comprehensive characterization of IIN requires micro-level data on investment activities between enterprises across all types of industries and regions[20], which entails an enormous volume of data. However, due to the challenges of accessing enterprise-level data and concerns over privacy, there is currently no publicly available dataset capable of capturing the spatiotemporal dynamics of IIN. This is the critical issue that our study seeks to address.

Our study addresses these gaps by providing a publicly available and validated dataset of China's IIN. Using a large-scale enterprise registration dataset covering the period from 2000 to 2020, we extracted 11,954,035 inter-enterprise investment records from a total of 17,273,411 enterprise entries. These enterprise-enterprise investment records were then transformed into enterprise-city connections, and ultimately aggregated into city-city investment connections. The resulting IIN represents cities as nodes, with investment frequency between cities serving as weighted edges. Our dataset includes 367 cities nationwide, with 134,666 directed and 92,994 undirected city dyads, ensuring its representativeness. The dataset consists of directed and undirected weighted IINs for every five-year interval from 2000 to 2020, alongside key network metrics such as indegree, outdegree, and node degree, and we also provide an additional version of the IIN that incorporates intracity investment. All data are publicly available for download on the Figshare platform[29]. To ensure data quality, we conducted several technical validations, including random sampling and manual verification with official records, comparing the dataset with urban networks derived from China's population mobility data and economic gravity model, testing the distance decay effect of intercity investment, and cross-referencing results from previous studies.

Methods

Overview. Our research involves several steps to establish the IIN, as illustrated in Fig. 1, which include five main stages: (1) Collecting original enterprise registration records from an enterprise registration information platform; (2) Extracting records of investor and investee enterprises involved in investment activities; (3) Adding information on the cities where the investor and investee enterprises are located; (4) Aggregating investment frequency between enterprises to the city level and removing records where both the investor and investee enterprises belong to the same city; and (5) Constructing IIN dataset and validating it from four aspects.

Data sources and cleaning. The National Enterprise Credit Information Publicity System (NECIPS) provides detailed registration information for all enterprises established in China, including rich historical information such as enterprise name, address, registered capital, industry classification, and outward investment records[30]. The outward investment records contain detailed information such as name of the investor enterprise, name of the investee enterprise, registered capital, paid-in capital, and investment date. However, due to privacy restrictions, it is no longer possible to obtain large-scale data directly from the official system in bulk. Therefore, we alternatively used another third-party query platform, Qichacha (www.qcc.com), to collect raw data. Qichacha is one of China's most authoritative enterprise information platforms, having obtained certification from the People's Bank of China for enterprise credit investigation and being officially registered with NECIPS[31]. Its fundamental data originate from NECIPS, where all enterprise registrations in China must be filed. Beyond this baseline, Qichacha employs advanced technologies including big data mining and artificial intelligence to process and structure massive amounts of additional information from sources such as enterprise annual reports, bidding documents, and other official records, with automated timely updates (see http://www.ixy360.com/). The platform has accumulated registration information for over 200 million enterprises across 8,000 industries (see http://www.ixy360.com/), effectively covering nearly all registered businesses in mainland China. This comprehensive

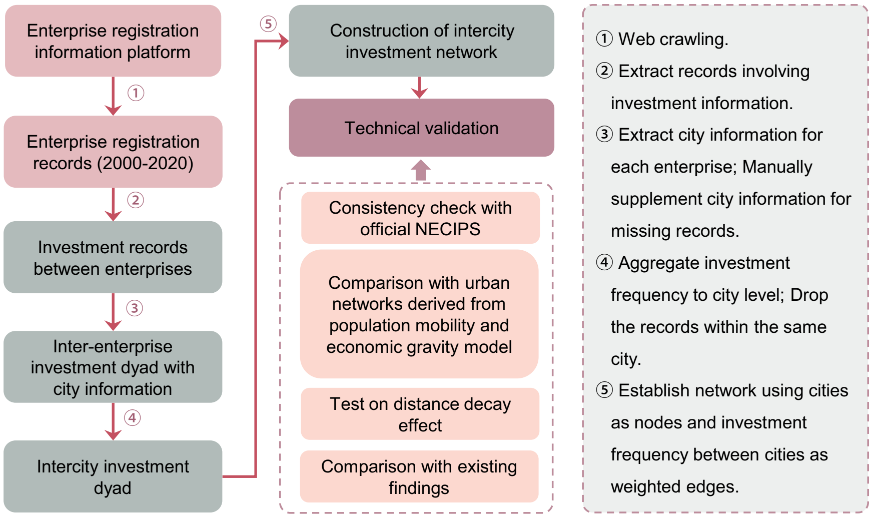

Fig. 1 Workflow of this study. NECIPS: National Enterprise Credit Information Publicity System.

approach results in broader coverage compared to NECIPS alone, while maintaining the reliability of the official system, making it particularly advantageous for our research purposes.

On the Qichacha platform, each enterprise has a dedicated webpage, where we can extract three key sections related to enterprise investment (see the example named "webpage example.pdf" in Figshare repository[29]). These three essential sections include basic registration information, outward investment records, and shareholder information. We specifically focus on enterprises with outward investment records, from which we can obtain all investee enterprise information through both current and historical outward investments, including enterprise names and place of registration. We then query the shareholder information of these investee enterprises to retrieve the actual paid-in capital amount and investment date. In summary, the integration of information across these three sections allows us to systematically construct inter-enterprise investment relationships. Section 1 provides information about the investor enterprise and its place of registration, Section 2 reveals the investee enterprise and its place of registration, and Section 3 helps us determine the investment year through paid-in capital dates. While the platform also provides investment amount data, we chose not to use it as network weights in our subsequent analysis. This decision was made because we discovered significant missing data, particularly in earlier years, and anomalous values that could potentially bias our dataset and compromise its representativeness (see detailed discussion in *Usage Notes* section).

Following the above procedure, we leveraged web crawling on the Qichacha platform to collect a total of 17,273,411 enterprise registration records involving investment activities across all industries in all cities of China from 2000 to 2020 at five-year intervals. We then removed duplicate records and excluded data from the registered cities outside mainland China. The next task was to fill in the information on the cities where the investor and investee enterprises were registered. In our study, cities are defined based on China's administrative divisions as of 2020, covering 367 cities, including 4 municipalities, 293 prefecture-level cities, 30 county-level cities or counties under direct provincial control, 30 autonomous prefectures, 7 regions, and 3 leagues[32]. While the majority of enterprises had their registration city directly available in their records (place of registration), some lacked this information. To address these cases, we implemented a three-step strategy: (1) For records with address information, we first used Python's "cpca" package[33] to extract city names; if this failed, we used the address and enterprise name to call the Amap API for geocoding[34], thereby obtaining city information; (2) If both methods failed or the original data lack address, we manually searched for the enterprise name on Internet to fill in missing field. Records that could not be resolved using the aforementioned methods were deleted. To avoid issues with city name duplication, all city data were processed using their respective unique administrative division codes (see "CityInfo.xlsx" in the data repository[29]). Ultimately, we obtained 11,954,035 inter-enterprise investment records, with 754,438 in 2000, 1,103,421 in 2005, 1,551,400 in 2010, 2,649,989 in 2015, and 5,894,787 in 2020. Considering privacy concerns, we provide a sample of anonymized enterprise names in the file "Inter-enterprise investment records (sample).xlsx", available on Figshare[29].

Construction of intercity investment network. The aim of this study is to construct an IIN that captures the economic connections between cities. To achieve this, we need to aggregate the inter-enterprise investments from the micro level to the city level. The microdata consist of enterprise–enterprise connections across all industries in China. A single enterprise may invest in multiple other enterprises, and the investee enterprises may be located either in the same city as the investor enterprise or in different cities. Therefore, we aggregated these data based on the city where each enterprise is registered, resulting in "enterprise-city" connections, which were then transformed into "city-city" connections. In this relationship, the strength of connection between cities is represented by the total number of investments between them, defined as investment frequency here. Since our focus is on IIN, we dropped "city-city" connections where both the investor and investee enterprises belong to the same city.

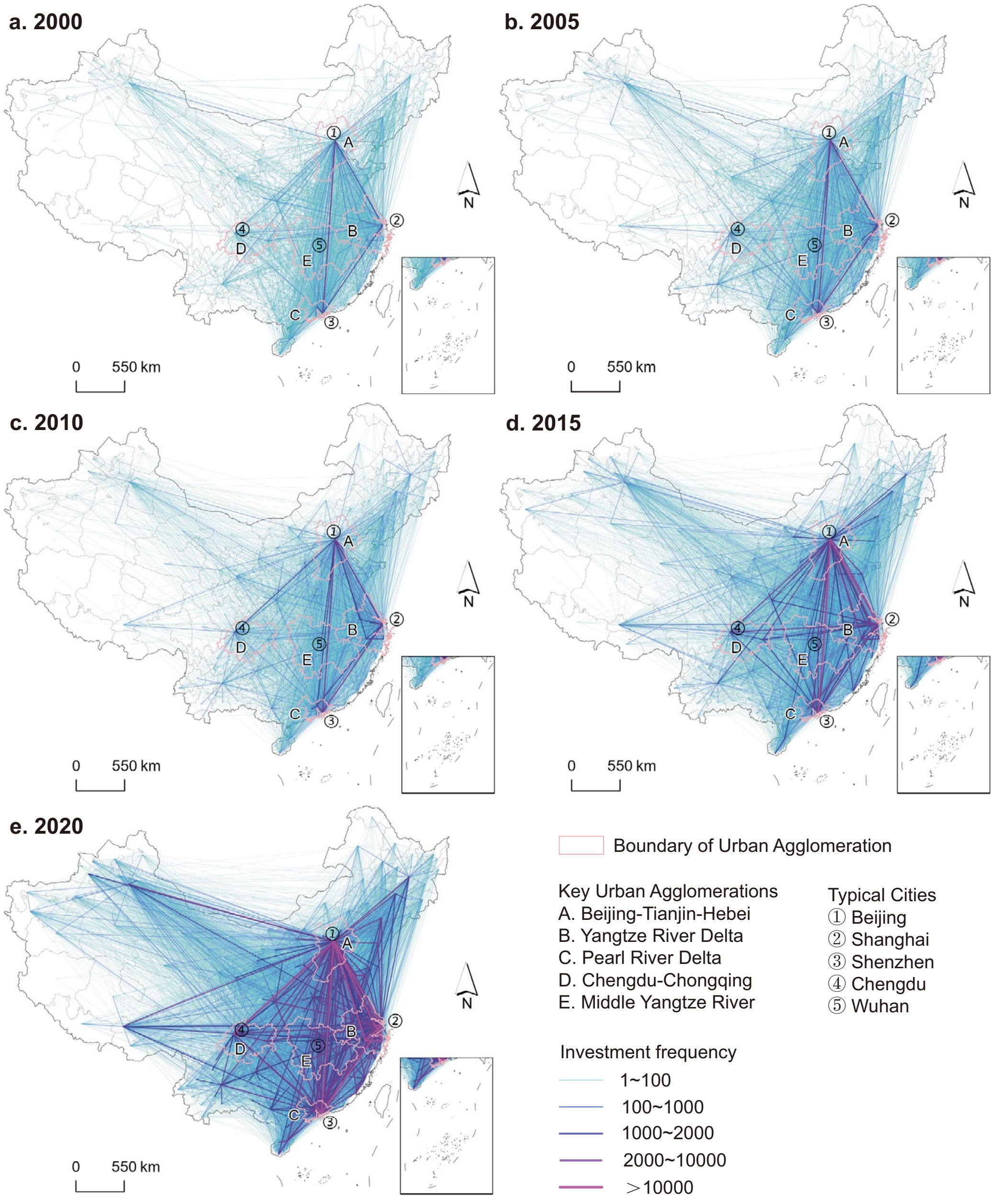

Fig. 2 The undirected intercity investment networks from 2000 to 2020. (**a**,**b**,**c**,**d**,**e**) respectively show the network of 2000, 2005, 2010, 2015 and 2020. Maps are created using ArcGIS Pro software (version 3.0.2).

Formally, we construct the IIN by defining cities as network nodes, the connections between cities as edges, and the investment frequency between cities as the edge weights. Accordingly, we can form two types of networks, namely directed and undirected weighted IINs. In the undirected weighted network, the strength of connection between nodes (cities) is given by:

$$S_{ij} = I_{i\to j} + I_{j\to i}(i \neq j) \tag{1}$$

where S_{ij} represents the connection strength, i.e., the edge weight in the undirected weighted network; $I_{i\to j}$ and $I_{j\to i}$ denote the directed investment frequency between cities i and j, which are the edge weights in the directed IIN and are directly derived from the previously mentioned aggregated "city-city" connections data. Notably, when we include the case where $i = j$, the IIN incorporates intracity investment, which is useful for understanding the localization characteristics of the network. Therefore, in the *Data Records* section, we have also provided datasets that include intracity investment.

Based on the constructed IIN, we can calculate three essential network metrics for each city. The first metric is outdegree, which represents the total number of investments a city makes in other cities. This metric reflects the city's ability to exert capital control over other cities in the network. The second metric is indegree, which

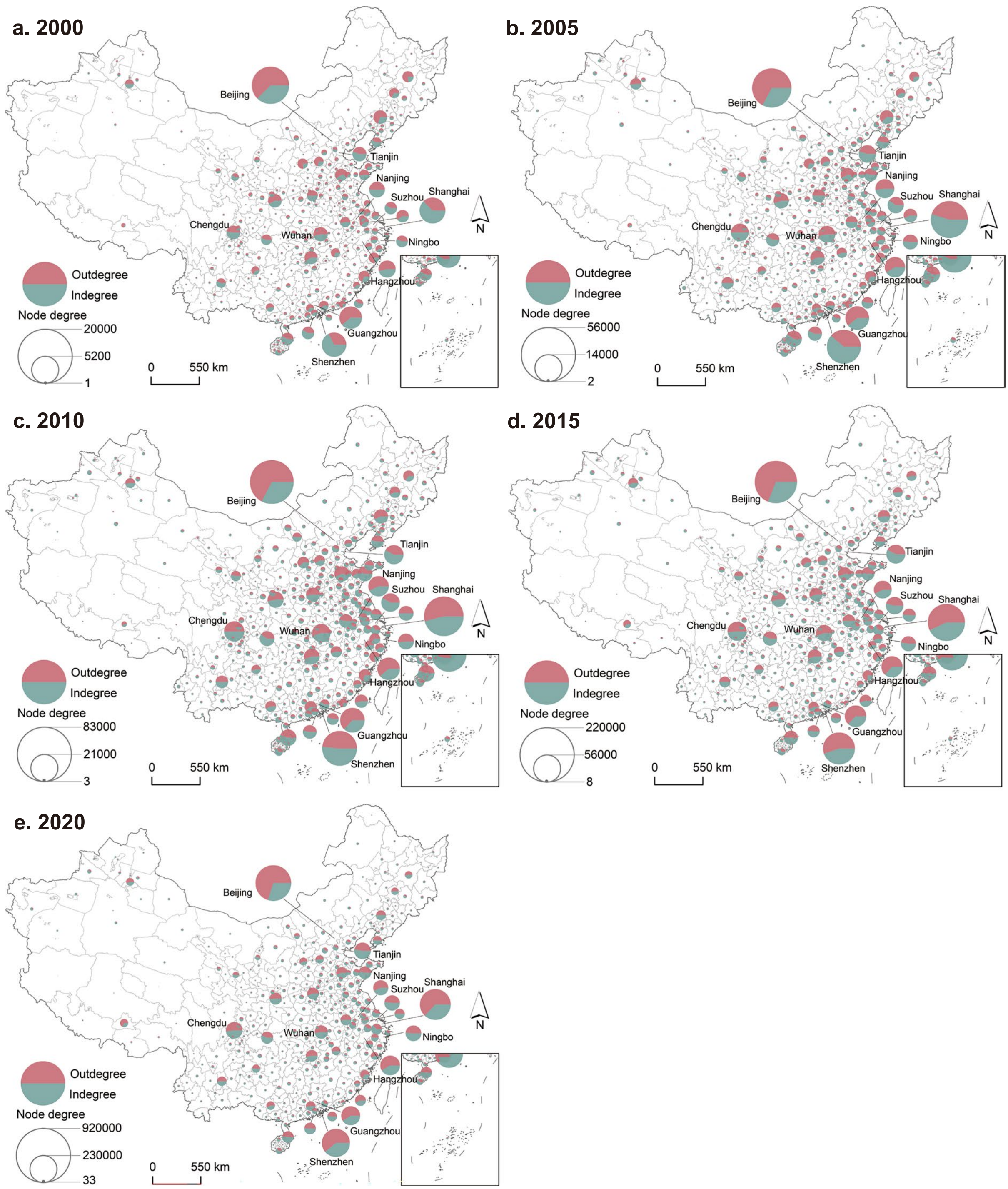

Fig. 3 Spatial distribution of outdegree, indegree and node degree between 2000 and 2020. Maps are created using ArcGIS Pro software (version 3.0.2).

represents the total number of investments received by a city from other cities, providing an indication of the city's ability to attract capital in the network[19]. The third metric is node degree, defined as the sum of outdegree and indegree, which reflects the overall influence of a city within the IIN. These metrics are formally expressed as follows:

$$Indegree_i = \sum\nolimits_j N_{i \to j}(i \neq j) \tag{2}$$

$$Outdegree_i = \sum\nolimits_j N_{j \to i}(i \neq j) \tag{3}$$

Type	Year	Number of city node	Number of city dyad	Sum	Mean	Min	Median	Max	SD
Directed network	2000	362	10671	76186	7	1	2	1540	32
	2005	366	15895	149491	9	1	2	2447	46
	2010	367	22895	271349	12	1	2	3742	63
	2015	367	33621	608031	18	1	2	9655	125
	2020	367	51584	1808636	35	1	3	20318	292
Undirected network	2000	362	7846	76186	10	1	2	2089	49
	2005	366	11474	149071	13	1	2	3326	72
	2010	367	16364	271349	17	1	2	5764	101
	2015	367	23418	608031	26	1	3	15746	203
	2020	367	33892	1808636	53	1	4	36473	490

Table 1. Descriptive statistics of intercity investment network on investment frequency.

$$Node\_degree_i = Indegree_i + Outdegree_i \tag{4}$$

where $N_{i\to j}$ denotes the number of investments made by city i in city j, and $N_{j\to i}$ represents the number of investments city i receives from city j.

As a result, we derived a directed IIN dataset covering 367 cities and 134,666 city dyad edges from 2000 to 2020, and we also constructed an undirected IIN dataset with 92,994 city dyad edges. The descriptive statistics of these networks are summarized in Table 1.

Characteristics of intercity investment network. Using the geographic coordinates (latitude and longitude) of each city's centroid, we visualized the IINs onto the map of China to provide an intuitive network view. Here, we merely present the undirected IIN to observe the dynamic economic interactions between Chinese cities from 2000 to 2020 (Fig. 2). It is evident that, over this period, Chinese cities have become increasingly interconnected through enterprise investments, forming a broad network of economic interactions. The overall structure shows a diamond-shaped pattern with Beijing, Shanghai, Shenzhen, and Chengdu serving as the key vertices. Moreover, the investment network has grown progressively more complex, indicating stronger economic interactions between cities over time. However, we also observe that the most densely connected regions are concentrated in a few urban agglomerations, such as Beijing-Tianjin-Hebei, Yangtze River Delta, and Pearl River Delta.

To better understand the network structure, Fig. 3 visualizes the distribution of network metrics (outdegree, indegree, and node degree) across cities. It is revealed that cities with the highest node degree, such as Beijing, Shanghai, and Shenzhen, have consistently dominated the network, reflecting their leading role in China's economic landscape. An interesting pattern emerges when examining the composition of degree metrics: large cities, especially in eastern regions, typically show higher outdegree than indegree, while smaller cities generally exhibit higher indegree than outdegree. This suggests that large eastern cities tend to function as investment initiators while smaller cities are more likely to be investment recipients.

Further, based on the undirected IIN, we employed the Leiden algorithm[35] for community detection to analyze city clustering patterns. As shown in Fig. 4, the analysis reveals distinct spatial communities that evolved over time. In 2000, the network exhibited 12 communities, with clear regional boundaries largely corresponding to China's traditional economic regions. Over time, the number of communities gradually decreased to 9 by 2020, suggesting increased integration of the investment network. Notably, three major economic regions - the Beijing-Tianjin-Hebei region, Yangtze River Delta, and Pearl River Delta - consistently formed their own distinct communities throughout the study period.

Lastly, utilizing the IIN dataset that includes intracity investment, we examined the frequency of intracity investment and outward investment for each city to assess whether a city's investment is more localized or outward-oriented. As shown in Fig. 5, the investment networks exhibit a significant localization tendency, with intracity investment frequency substantially higher than outward investment frequency. However, we also observed that this localization phenomenon has gradually weakened over time, indicating an evolving trend toward broader geographical investment connections.

Data Records

The dataset we produced, along with its supporting code and additional data, can be accessed on Figshare[29]. The dataset consists of three parts, all stored in Excel format. The first part includes two versions of the directed IIN dataset: one excluding intracity investments ("Directed intercity investment network dataset.xlsx") and another including intracity investments ("Directed intercity investment network dataset (including intracity investment).xlsx"). Both versions contain information such as the investor city, investee city, investment frequency, and the latitude and longitude of the cities' centroids. The data fields and descriptions are as follows (Table 2).

Similarly, the second part also provides two versions of the undirected IIN dataset: the standard version ("Undirected intercity investment network dataset.xlsx") and the comprehensive version ("Undirected intercity investment network dataset (including intracity investment).xlsx"). Both datasets document information about the cities involved in the investment activities and the investment frequency between them. The data fields and descriptions are as follows (Table 3).

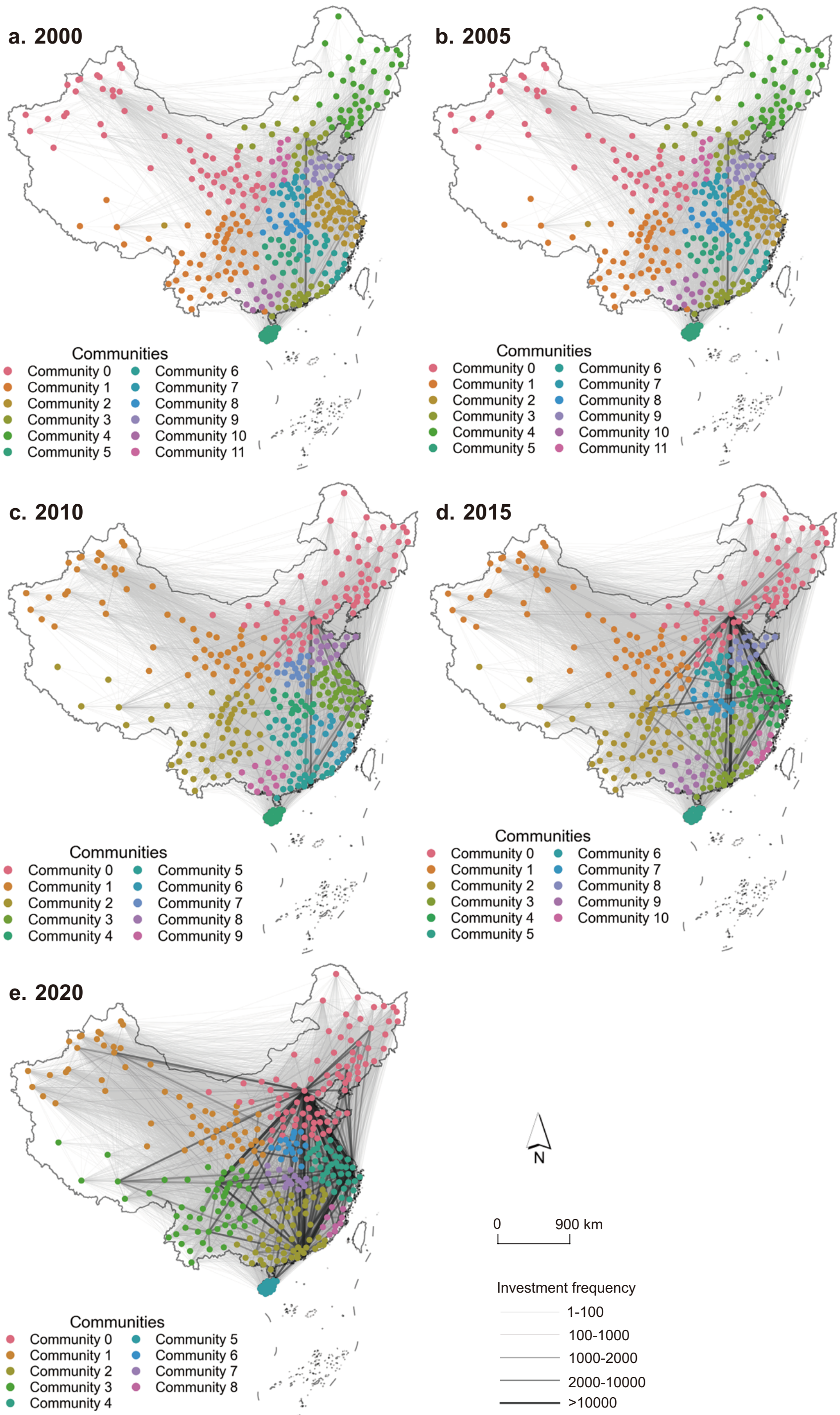

Fig. 4 Community detection results using the Leiden algorithm from 2000 to 2020.

The third part is the dataset of three network metrics at the city node level, titled "Three network metric dataset.xlsx". It includes three important network metrics: outdegree, indegree, and node degree, as well as the city rankings for these metrics across different years. The data fields and descriptions are as follows (Table 4).

In addition, we provide several datasets used to support the production and validation (see *Technical Validation* section) of the main dataset. These include "CityInfo.xlsx" (Information of 367 cities), "Inter-enterprise investment records (sample).xlsx", "Population mobility in 2020.xlsx", "Urban GDP (2000–2020).xlsx" and "Investment records of specialized investment institutions in 2020.xlsx". Detailed descriptions for each dataset can be found in the notes of the individual Excel files.

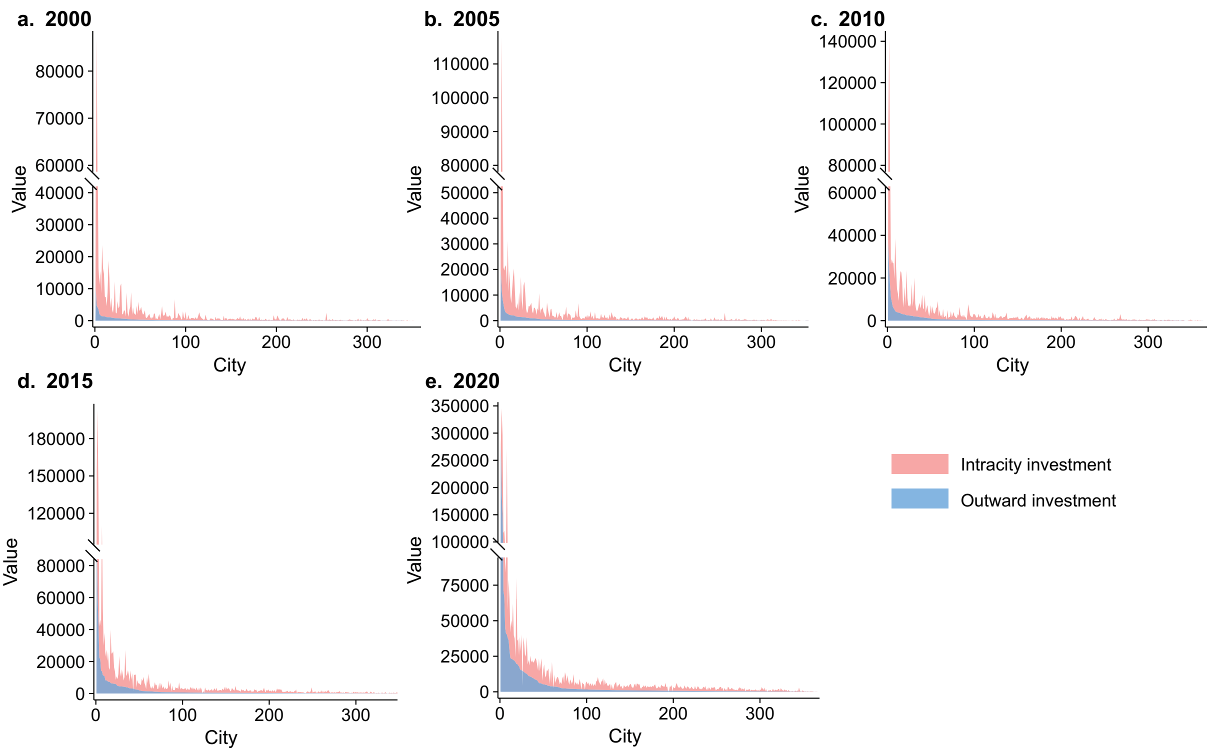

Fig. 5 Comparison of intracity and outward investment frequency of IINs (2000–2020).

Technical Validation

To verify the reliability and validity of our dataset, we conducted validation through four distinct ways. First, we performed a consistency check between the collected inter-enterprise investment records and data queried from the official NECIPS. Second, we compared the produced IIN data with the urban networks derived from China's population mobility data and economic gravity model. Third, we tested the distance decay effect of intercity investments. Fourth, we compared our results behind the data with findings from extant related studies.

Consistency check with official query from national enterprise credit information publicity system. Our initial data were obtained from the Qichacha platform. Although Qichacha's data are sourced from NECIPS enterprise registration records, we conducted an additional verification by cross-checking our inter-enterprise investment records with data queried directly from NECIPS. Since NECIPS does not support bulk queries, we performed a manual, one-by-one verification process. We randomly sampled 100 records each for 2000, 2005, 2010, 2015, and 2020, resulting in a total of 500 samples. Five co-authors of this paper manually verified these records over the course of a week by entering either the investor or investee enterprise into the NECIPS query and checking the "Shareholder and Investment Information" subsection to confirm the accuracy of each investment record. Our checks show that 100% of the records that could be found in the NECIPS system matched our intercity investment data. However, we also noticed that some investment records obtained from Qichacha could not be found in the NECIPS system, suggesting that the official system may not be updated as frequently. This indicates that our data collection is more comprehensive, while still ensuring consistency with official records.

Comparison with population mobility networks and economic gravidity model. Since the dataset we produced is unique, it is challenging to find direct references for detailed comparison and validation. Therefore, we opted for an indirect validation of the dataset. IIN not only reflects economic interactions between cities but may also be related to the flow of human resources[20]. Thus, we validated the dataset indirectly by examining the population mobility network and the economic gravity model (albeit imperfectly). This choice is based on the following theoretical assumptions. For population mobility, the investment activity is often accompanied by the movement of labor, and it may create more job opportunities, attracting population flows[36]. Conversely, population mobility may also bring new investment opportunities, further promoting capital flows. Therefore, if our dataset is valid, we would expect a significant correlation and similar distribution between the IIN and the population mobility network. Regarding the economic gravity model, it assumes that the intensity of interactions between cities is proportional to their economic size (e.g., GDP)[37]. This suggests that cities with larger economies may interact more strongly. As a form of economic interaction, intercity investments should theoretically align with the gravity model's expectations. If our dataset is reliable, the predicted economic interaction levels from the gravity model should exhibit similar characteristics to the IIN in terms of interaction intensity and distribution.

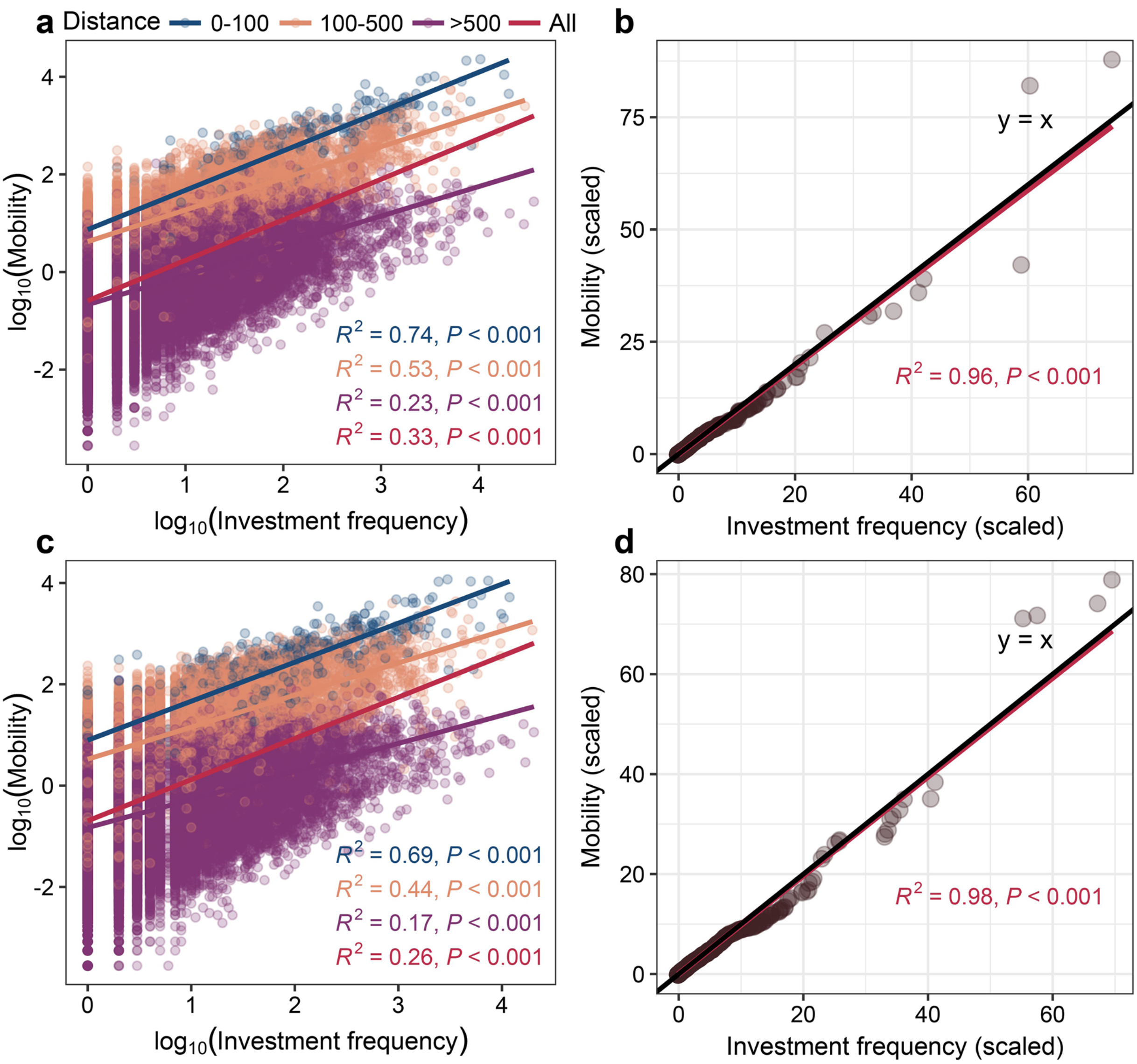

Fig. 6 Comparison between intercity investment network and population mobility network in 2020. (**a**) scatter plot for undirected network. (**b**) Q-Q plot for undirected network. (**c**) scatter plot for directed network. (**d**) Q-Q plot for directed network.

For the population mobility data, we obtained it from the Amap Population Migration Data Platform[38] in the year of 2020, corresponding to our dataset's timeframe (as earlier years do not have available data). We aggregated the daily migration flow index to annual values, resulting in a population mobility network between 367 cities.

To compare the IIN with the population mobility network, we employed two methods. First, we can create scatter plots to directly compare the two datasets and observe the fit between them. A positive slope in the fitted curve will indicate a correlation between the two networks. Second, following the method of Ref. [39], we can use a Q-Q (quantile-quantile) plot to assess the similarity in distribution of the two datasets. If the distributions of the two network metrics tend to be the same one, their data points should align along a line defined by $y = kx$, where x and y represent the percentiles of the two network metrics, and k is the coefficient. Due to differences in data sources and scales, we first scaled both the intercity investment data and the population mobility data using the following formula:

$$flow\_scaled = \frac{flow - \mu}{\sigma} \tag{5}$$

where *flow* denotes the investment frequency or population mobility volume between cities, and *flow_scaled* is the standardized value; μ and σ are the sample mean and standard deviation, respectively. In this study, we compared both the directed and undirected networks in 2020. To account for the potential bias introduced by city distance, we also examined the correlation between investment frequency and population mobility volume

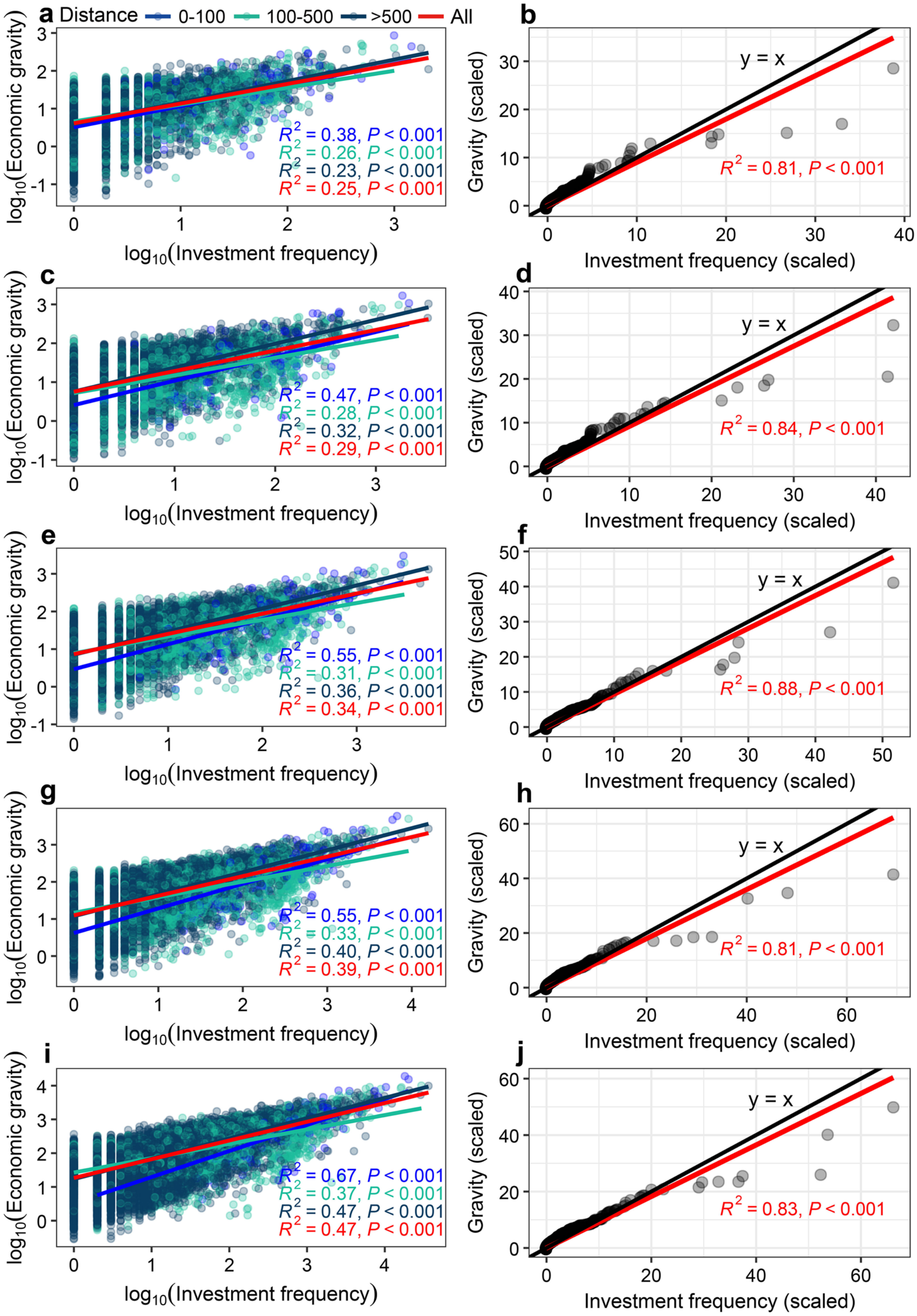

Fig. 7 Comparison between intercity investment network and economic gravity model. (**a**,**c**,**e**,**g**,**i**) are scatter plots for the years of 2000, 2005, 2010, 2015, 2020. (**b**,**d**,**f**,**h**,**j**) are Q-Q plots for the years of 2000, 2005, 2010, 2015, 2020.

across three distance intervals: 0–100 km, 100–500 km, and greater than 500 km. The 0–100 km range approximates the radius of metropolitan regions in China, while the 100–500 km range corresponds to the distance between cities within urban agglomerations. Typically, cities within a metropolitan region exhibit the strongest linkage[40], followed by those within urban agglomerations. Therefore, we expect a stronger correlation between investment frequency and mobility flow within the metropolitan regions.

Figure 6a,c demonstrate a significant positive correlation between investment frequency and population mobility volume (both log-transformed) in both the undirected and directed networks. The undirected network

Field	Description
CityDyad	Identity of network edge
InvestmentFrequency	Number of investments from investor city to investee city
InvestorCityCode	Administrative division code of investor city
InvesteeCityCode	Administrative division code of investee city
InvestorCity_LNG	Centroid longitude of investor city
InvestorCity_LAT	Centroid latitude of investor city
InvesteeCity_LNG	Centroid longitude of investee city
InvesteeCity_LAT	Centroid latitude of investee city
InvestorCityName_CN	Chinese name of investor city
InvestorCityName_EN	English name of investor city
InvesteeCityName_CN	Chinese name of investee city
InvesteeCityName_EN	English name of investee city
Distance_km	Distance between investor city and investee city
InvestmentYear	Year of investment

Table 2. Field description of the directed intercity investment network dataset.

Field	Description
CityDyad	Identity of network edge
InvestmentFrequency	Number of mutual investments between two cities
CityACode	Administrative division code of city A
CityBCode	Administrative division code of city B
CityA_LNG	Centroid longitude of city A
CityA_LAT	Centroid latitude of city A
CityB_LNG	Centroid longitude of city B
CityB_LAT	Centroid latitude of city B
CityAname_CN	Chinese name of city A
CityAname_EN	English name of city A
CityBname_CN	Chinese name of city B
CityBname_EN	English name of city B
Distance_km	Distance between city A and city B
InvestmentYear	Year of investment

Table 3. Field description of the undirected intercity investment network dataset. City A represents one city in the city dyad, while city B represents the other, with no directional orientation between the two.

Field	Description
CityCode	City administrative division code
CityName_CN	Chinese city name
CityName_EN	English city name
LNG	City centroid longitude
LAT	City centroid latitude
Outdegree	Number of investments sent by the city
Indegree	Number of investments received by the city
NodeDegree	Sum of outdegree and indegree
Rank_out	City ranking of outdegree
Rank_in	City ranking of indegree
Rank_node	City ranking of node degree
InvestmentYear	Year of investment

Table 4. Field description of three network metrics dataset.

shows a better fit, with an R^2 of 0.33, compared to the directed network, which has an R^2 of 0.26. This indicates that the investment frequency between cities is closely related to population mobility, and the undirected network better captures this relationship. When considering different distance intervals, cities within shorter distances exhibit a stronger model fit between investment frequency and mobility volume. For cities within the 0–100 km range (roughly corresponding to the radius of metropolitan regions in China), the undirected

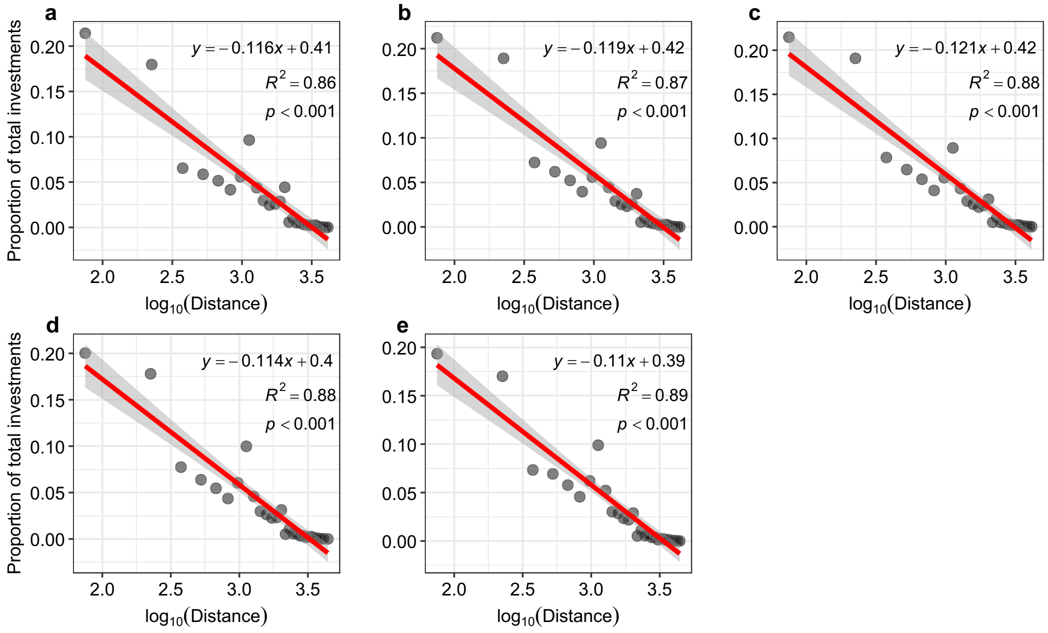

Fig. 8 The relationship between geographical distance and intercity investments. (**a**,**b**,**c**,**d**,**e**) respectively show the scatter plots and fitted lines of 2000, 2005, 2010, 2015 and 2020.

	Gravity model parameters		Pearson's correlation			
Year	***k***	***β***	**Overall**	**0–100**	**100–500**	**>500**
2000	0.0010000	0.4780	0.608	0.763	0.730	0.496
2005	0.0002511	0.3897	0.708	0.850	0.769	0.636
2010	0.0001548	0.4784	0.708	0.869	0.783	0.610
2015	0.0001345	0.4804	0.691	0.899	0.797	0.588
2020	0.0001984	0.4775	0.740	0.911	0.763	0.668

Table 5. Estimation results of gravity model parameters and Pearson's correlation between economic gravity and investment frequency.

network achieves an R^2 as high as 0.74, while the directed network shows an R^2 of 0.69. As the distance increases, the model fit weakens, suggesting that investment flows and population mobility are more closely linked within the scale of metropolitan regions.

Figure 6b,d further illustrate this relationship through Q-Q plots, where the investment frequency and population mobility volume form a nearly perfect line $y = x$. The R^2 values are 0.96 for the undirected network and 0.98 for the directed network, indicating a high degree of alignment between the distributions of two datasets. Although a few data points deviate from the fitted line in the high-investment range, these outliers are minimal and do not significantly affect the overall data distribution. In conclusion, the strong correlation and consistent distribution between the IIN and the population mobility network confirm the reliability of our dataset. The data suggest that intercity investment flows are closely linked to population mobility, particularly within metropolitan regions, further validating the robustness of the IIN dataset we constructed.

On the other hand, the gravity model has been widely used to estimate economic interactions between cities, namely economic gravity model[37,41–43]. Specifically, the theory suggests that the economic interaction between two cities is more likely to be stronger if their economic sizes are larger and their geographical distance is shorter. This can be expressed by the following formula:

$$E_{ij} = k\frac{GDP_i \times GDP_j}{D_{ij}^{\beta}} \tag{6}$$

where E_{ij} refers to economic gravity between cities i and j, mirroring the strength of economic interaction; k is a gravitational constant; GDP_i and GDP_j are the economic sizes of cities i and j, respectively; D_{ij} is the geographical distance between the centroids of cities i and j; and β is the distance decay coefficient. To estimate parameters k and β, following existing lierature[39,44], we employed PSO (Particle Swarm Optimization) algorithm to minimize the RMSE (Root Mean Square Error) between estimated economic gravity and actual investment frequency. Specifically, we collected GDP data for the years 2000, 2005, 2010, 2015, and 2020 from the China City Statistical Yearbook. The distance between cities was calculated as the straight-line distance between their geographic

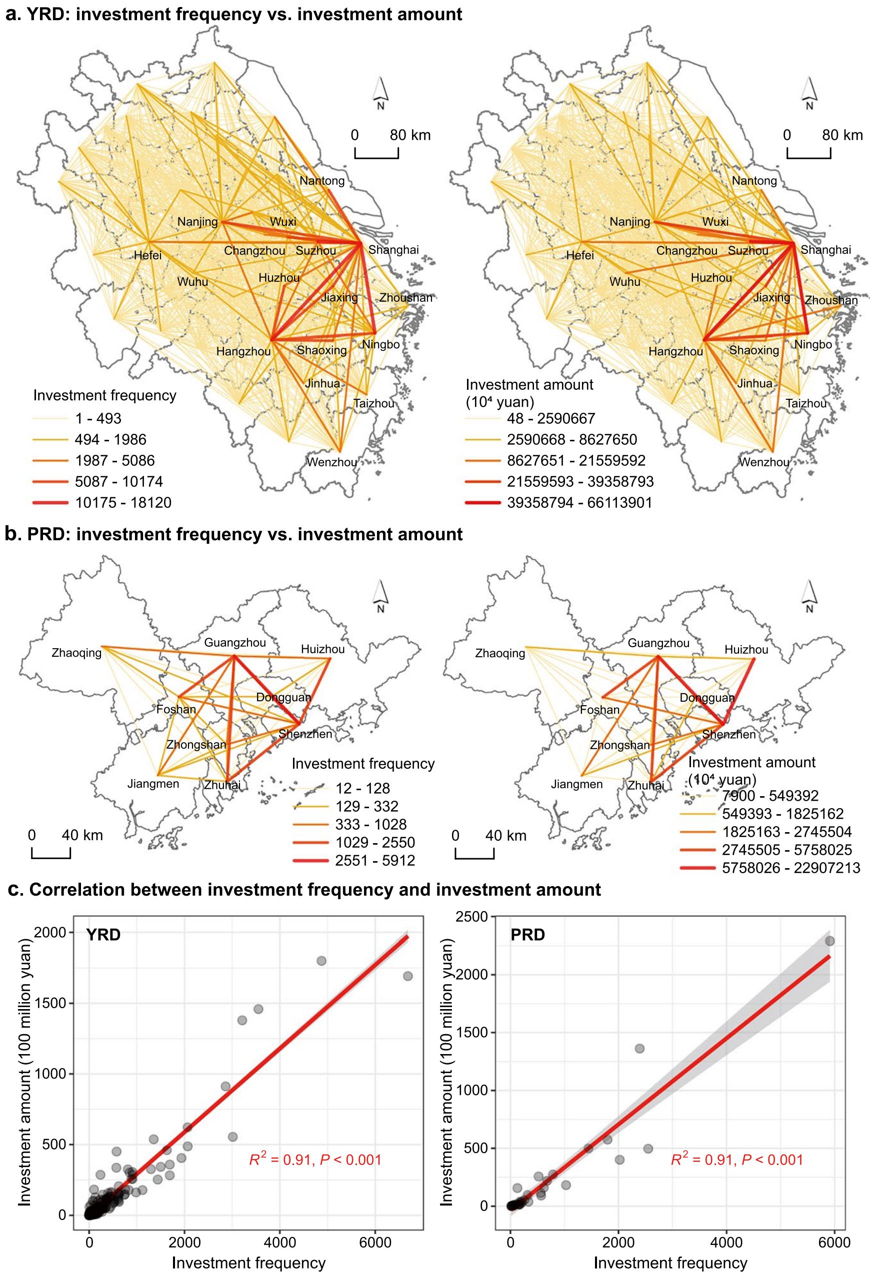

Fig. 9 Comparison of intercity investment networks based on investment frequency and investment amount in the Yangtze River Delta and Pearl River Delta regions (2020).

centroids. Using the "pso" package in R[45], we set the search ranges for k and β to 0–1 and 0–2, respectively, and obtained the parameter estimates for each year (Table 5). Based on these parameters, we calculated the economic gravity values using Eq. (6) and then computed their Pearson's correlation with the actual investment frequency. We also calculated the correlation coefficients for different distance intervals (Table 5).

The results show strong correlations (overall above 0.6) that increase over time (despite a slight fluctuation in 2015). Notably, the strongest correlations were observed in the 0–100 km distance range, reaching above 0.9 in 2020, providing preliminary validation of our dataset's effectiveness. Following the analysis procedure for the IIN and population mobility networks, we examined the model fit and distributional similarity between economic gravity and investment frequency.

Figure 7a,c,e,g,i show a significant positive correlation between economic gravity and investment frequency (both log-transformed), with the R^2 value for the fit increasing over time, reaching 0.47 in 2020. Similar to the population mobility network, the best fit between economic gravity and investment frequency is observed in the 0–100 km distance range. In 2020, the R^2 for this distance range was 0.67, indicating that the IIN better mirrors economic connections between cities at shorter distances. From the Q-Q plots (Fig. 7b,d,f,h,j), we can see that the quantiles of investment frequency and economic gravity are almost perfectly aligned along the $y = x$ line across all years, with only a very few outliers at high values. The R^2 for all years exceeds 0.80, demonstrating

2000		2005		2010		
Our study	***Wu & Yao*[24]**	**Our study**	***Zhang et al.*[20]**	**Our study**	***Wu & Yao*[24]**	***Zhang et al.*[20]**
Beijing	Beijing	Beijing	Beijing	Beijing	Beijing	Beijing
Shenzhen	Daqing	Shanghai	Shanghai	Shanghai	Shanghai	Shanghai
Shanghai	Shanghai	Shenzhen	Shenzhen	Shenzhen	Shenzhen	Shenzhen
Guangzhou	Shenzhen	Guangzhou	Haikou	Guangzhou	Tianjin	Haikou
Hangzhou	Guangzhou	Hangzhou	Guangzhou	Hangzhou	Chengdu	Chengdu
Nanjing	Dongying	Nanjing	Tianjin	Nanjing	Hangzhou	Guangzhou
Tianjin	Jinan	Chengdu	Nanjing	Chengdu	Taiyuan	Tianjin
Chengdu	Tianjin	Tianjin	Chengdu	Tianjin	Guangzhou	Nanjing
Wuhan	Chengdu	Wuhan	Zhuhai	Suzhou	Jinan	Wuhan
Suzhou	Hangzhou	Suzhou	Wuhan	Wuhan	Nanjing	Chongqing
Zhuhai	Baoding	Haikou	Xi'an	Haikou	Chongqing	Xi'an
Haikou	Changsha	Xi'an	Lanzhou	Xi'an	Xi'an	Hangzhou
Xi'an	Shijiazhuang	Ningbo	Suzhou	Ningbo	Suzhou	Suzhou
Ningbo	Nanjing	Changsha	Dalian	Changsha	Zhuhai	Zhuhai
Changsha	Zhengzhou	Zhuhai	Chongqing	Jinan	Dalian	Urumqi
Shenyang	Shenyang	Jinan	Hangzhou	Qingdao	Wuhan	Shenyang
Wuxi	Nanning	Shenyang	Shenyang	Chongqing	Kunming	Dalian
Jinan	Xi'an	Wuxi	Urumqi	Wuxi	Zhengzhou	Kunming
Chongqing	Suzhou	Chongqing	Qingdao	Shenyang	Qingdao	Lanzhou
Zhengzhou	Baoji	Qingdao	Kunming	Zhengzhou	Shenyang	Wuxi
Number/Proportion of occurrences with the same city	14/70%	/	15/75%	/	16/80%	15/75%

Table 6. Comparison of the top 20 cities in node influence of intercity investment network in our study with other similar works. .

	Intracity		Intercity	
Group	**Count**	**Percentage**	**Count**	**Percentage**
A	5188	53.6%	4500	46.4%
B	2777	28.7%	6911	71.3%
C	2542	26.2%	7146	73.8%

Table 7. Distribution of intracity and intercity investment relationships across different layers of specialized investment institutions in 2020. Group A: City relationship between investment institutions and their established entities; Group B: City relationship between investment entities and investees; Group C: Indirect city relationship between investment institutions and investees (mediated by the investment entities).

a strong fit. This indicates a high degree of similarity in the data distribution between intercity investment frequency and economic interaction from economic gravity model across different years. To conclude, these findings provide further evidence of the scientific validity of the dataset we produced. The strong correlation between the IIN and the economic gravity supports the reliability of our data.

Test on the distance decay effect of intercity investments. Distance decay effect indicates that as distance between two cities increases, the intensity of economic activities such as investments tends to decrease[46,47]. This is due to higher transaction costs, including transportation and communication, as well as the increasing difficulty in maintaining efficient information flow. As a result, cities that are geographically closer to each other tend to engage in more frequent investment activities, while distant cities experience a reduction in such interactions. To verify if our data satisfy such principle, we calculate the proportion of total investment frequency within different distance ranges relative to the total number of intercity investments for each year, and then plot a scatter graph (Fig. 8). Clearly, investment intensity decreases as distance increases. Over time, the absolute value of the slope of the fitted line has decreased, suggesting that the constraint of distance on investment is weakening. This demonstrates that our data aligns with theoretical principles, thereby indirectly validating the reliability of the dataset to some extent.

Comparison with existing research findings. Here, we compare the findings of our dataset with results from existing related studies. First, the overall "diamond" structure of the IIN, as visualized on the map of China (Fig. 2), is similar to that observed in a few other nationwide studies using intercity investment data, such as *Lu & Sun*[26], *Zhang & Tang*[48], *Zhang et al.*[20], and *Wu & Yao*[24], although there may be differences in data sources. Additionally, we examined the nodal influence of cities nodes in our network and compared our findings with results reported in comparable studies. Since different studies use varying metrics to measure the importance of cities within a network, we focused on the overlap in the top 20 most influential cities, instead of comparing the specific influencing values. In our study, city influence is measured based on node degree. Due to the limited availability of specific city information in other studies, we were able to obtain comparable data only from the studies of *Wu & Yao*[24] and *Zhang et al.*[20], which primarily cover the years 2000, 2005, and 2010 (see Table 6). As shown in Table 6, about 15 cities consistently overlap with those in our study, accounting for the majority of the top cities in both studies. This high level of overlap suggests that our dataset is reasonable and aligns well with previous studies, reinforcing the validity of our IIN dataset.

Usage Notes

Our dataset provides investment frequency between city dyads as well as their geographic coordinates (latitude and longitude), making it easy for potential users to visualize the network using GIS software. For example, in ArcGIS Pro, the "XY To Line" tool can be used to generate the investment network. Additionally, advanced network visualization tools like Gephi can be employed to create visually appealing and detailed network graphs. Beyond visualization, users can apply social network analysis methods[49] to further explore the dataset. In the context of this study, several aspects of social network analysis hold potential. For example, users can examine centrality measures (such as betweenness and closeness) to identify key cities that act as critical hubs in the IIN. Another avenue is analyzing network modularity to investigate how cohesive sub-networks (city clusters) evolve over time, potentially offering insights into how urban regions are economically structured. Additionally, users can combine our network metrics (such as node degree, outdegree, and indegree) with other variables to explore interaction mechanisms. For instance, as *Zhang et al.*[20] demonstrated, institutions and markets can influence the evolution of China's IIN.

While we have invested considerable effort in producing and validating the IIN dataset, like any dataset, it has certain limitations that require acknowledgment. First, our choice to use investment frequency rather than investment amount as network weights warrants specific discussion. While investment amount could indeed better capture capital flows, several important factors influenced our methodological choice. First, there are significant issues with data completeness and quality. In our dataset, the paid-in capital data from investor enterprise to investee enterprise is largely missing, especially for the years 2000, 2005, and 2010, which could lead to a lack of representativeness. Second, we observed several anomalous capital flow data between certain city dyads. For instance, in 2020, while the investment frequency between Shenzhen and Sansha was relatively low (70 occurrences, ranking 2,639th among all city dyads), their total capital flow reached 240.5 billion RMB (ranking third overall) - a clear outlier that could distort network analysis. Third, by using frequency data, we ensure robust comparability across our long time series (2000–2020) as our dataset is not affected by such anomalous investment amount. Nevertheless, our IIN dataset is sufficient to capture the economic connectivity between cities, with strong comparability, and the visualized dataset aligns well with the actual situation in China (see Section *Characteristics of intercity investment network*).

To address this concern empirically, we conducted a detailed comparison in a small sample using 2020 data from the Yangtze River Delta (YRD) and Pearl River Delta (PRD) regions, where data quality is higher and market mechanisms are more mature[50]. As shown in our new analysis (Fig. 9), we compared networks constructed using investment frequency versus investment amount. The results show that the structures of the two networks are highly similar, with high-frequency investments concentrated among a few large cities, and this concentration is slightly more pronounced in the network constructed using investment amount. However, the overall differences are minimal. The scatter plots (Fig. 9c) also demonstrate a high correlation, with an R^2 value of 0.91 for both regions. This confirms the validity of our IIN dataset constructed based on investment frequency, despite its imperfections. It is important to note, however, that if researchers are interested in specific investment amount among enterprises rather than the macro-level patterns of investment connectivity, our dataset may not fully meet their needs.

Another limitation concerns the investment by specialized investment institutions (SII), such as venture capital (VC) and private equity (PE) institutions. In China's investment landscape, SIIs often establish separate investment entities that may be located in different cities from their operational controls, typically in areas offering tax incentives or favorable policies. To examine this potential bias, we collected data from PEDATA of Zero2IPO Group (https://max.pedata.cn/), a leading integrated service provider in China's private equity industry. Using their comprehensive database of SII investment deals from 2020, supplemented with city location information from the Qichacha platform (see detailed records in "Investment records of specialized investment institutions in 2020.xlsx" of the data repository[29]), we identified a complex three-layer structure of investment relationships. As shown in Table 7, while 53.6% of investment institutions establish their investment entities in the same city (Group A), the proportion of intracity investments drops to 28.7% when examining the relationship between investment entities and investees (Group B). When considering the indirect relationship between investment institutions and their ultimate investees (Group C), only 26.2% occur within the same city. However, while 46.4% (4,500) of SII investment events in our analysis are intercity investments, these potentially biased cases only account for 0.25% of our dataset's 1,808,636 intercity investment relationships in 2020. Therefore, unless these SII investments are highly concentrated in specific cities, this limitation is unlikely to significantly affect the overall network structure. Nevertheless, researchers should exercise caution when using this dataset for micro-level analysis of enterprise-to-enterprise investment patterns.

Despite the limitations, our IIN dataset focuses on intercity economic connections, combined with the scientific validation, ensures the dataset is robust and reliable. For users investigating economic connectivity between cities, this database should sufficiently meet their needs.

Code availability

The pre-processing, dataset production, and validation of the dataset in this study are primarily implemented using R programming and ArcGIS Pro software, while the Leiden algorithm for community detection is executed using Python language. The R and Python codes are available in the file titled "Code.R" and "Leiden.py", respectively, which have been uploaded alongside the dataset. Users can easily replicate our results using the provided materials.

Received: 23 October 2024; Accepted: 17 February 2025;
Published online: 03 March 2025

References

1. Derudder, B. & Taylor, P. J. Three Globalizations Shaping the Twenty-first Century: Understanding the New World Geography through Its Cities. *Annals of the American Association of Geographers* **110**, 1831–1854 (2020).
2. Li, Z., Zhen, F. & Liu, W. Exploring the spatial pattern and influencing factors of intercity capital flows from 2005 to 2019: A case study of Yangtze River Delta region, China. *Growth and Change* **55**, e12694 (2024).
3. Neal, Z. P. *The Connected City: How Networks Are Shaping the Modern Metropolis*. https://doi.org/10.4324/9780203101728 (Routledge, New York, 2012).
4. Castells, M. *The Rise of the Network Society*. (John Wiley & Sons, 2011).
5. Ducruet, C. & Beauguitte, L. Spatial Science and Network Science: Review and Outcomes of a Complex Relationship. *Netw Spat Econ* **14**, 297–316 (2014).
6. Taylor, P. & Derudder, B. *World City Network: A Global Urban Analysis*. https://doi.org/10.4324/9781315730950 (Routledge, London, 2015).
7. Li, L., Derudder, B., Shen, W. & Kong, X. Exploring the dynamics of the disaggregated intercity corporate network in the Yangtze River Delta, China: a relational event approach. *J Geogr Syst* **24**, 115–140 (2022).
8. Zhu, B., Pain, K., Taylor, P. J. & Derudder, B. Exploring external urban relational processes: inter-city financial flows complementing global city-regions. *Regional Studies* (2022).
9. Lan, X., Hu, Z. & Wen, C. Does the opening of high-speed rail enhance urban entrepreneurial activity? Evidence from China. *Socio-Economic Planning Sciences* **88**, 101604 (2023).
10. Pan, F. & Yang, B. Financial development and the geographies of startup cities: evidence from China. *Small Bus Econ* **52**, 743–758 (2019).
11. Belderbos, R., Benoit, F. & Derudder, B. World City Innovation and Service Networks and Economic Growth. *Papers in Regional Science* **101**, 979–1000 (2022).
12. Du, Z. *et al*. Inter-urban mobility via cellular position tracking in the southeast Songliao Basin, Northeast China. *Sci Data* **6**, 71 (2019).
13. Liang, X., Hidalgo, C. A., Balland, P.-A., Zheng, S. & Wang, J. Intercity connectivity and urban innovation. *Computers, Environment and Urban Systems* **109**, 102092 (2024).
14. Ma, H., Wei, Y. D., Huang, X. & Zhang, W. The innovation networks shaped by large innovative enterprises in urban China. *J. Geogr. Sci.* **33**, 599–617 (2023).
15. Pan, F., Bi, W., Lenzer, J. & Zhao, S. Mapping urban networks through inter-firm service relationships: The case of China. *Urban Studies* **54**, 3639–3654 (2017).
16. Wang, J., Liu, F., Li, J., Du, D. & Wu, J. Geographical distance, investment strategies and performance of venture capital: evidence from Chinese enterprises. *Spatial Economic Analysis* 1–25, https://doi.org/10.1080/17421772.2024.2342336 (2024).
17. Zhao, P., Wang, H., Liu, Q., Yan, X.-Y. & Li, J. Unravelling the spatial directionality of urban mobility. *Nat Commun* **15**, 4507 (2024).
18. Taylor, P. J. Specification of the World City Network. *Geographical Analysis* **33**, 181–194 (2001).
19. Liu, X. & Derudder, B. Analyzing urban networks through the lens of corporate networks: A critical review. *Cities* **31**, 430–437 (2013).
20. Zhang, R., Lin, J. & Sun, D. The role of institutions and markets in shaping intercity investment networks in China. *Cities* **153**, 105221 (2024).
21. Derudder, Taylor, Witlox, & Catalano. Hierarchical Tendencies and Regional Patterns in the World City Network: A Global Urban Analysis of 234 Cities. *Regional Studies* **37**, 875–886 (2003).
22. Taylor, P. J. & Csomós, G. Cities as control and command centres: Analysis and interpretation. *Cities* **29**, 408–411 (2012).
23. Wall, R. S. & Van Der Knaap, G. A. Sectoral Differentiation and Network Structure Within Contemporary Worldwide Corporate Networks: ECONOMIC GEOGRAPHY. *Economic Geography* **87**, 267–308 (2011).
24. Wu, K. & Yao, C. Exploring the association between shrinking cities and the loss of external investment: An intercity network analysis. *Cities* **119**, 103351 (2021).
25. Guo, J., Qin, Y. & Guo, S. The Influence of Enterprise Investment Behavior on the Cyberspace Connection of Coastal Cities: Empirical Evidence from China. *Journal of Urban Planning and Development* **149**, 05022046 (2023).
26. Lu, J. & Sun, D. Spatiotemporal Evolution and Complexity of Urban Networks in China, 1978–2019: An Enterprise Linkages Perspective. *Complexity* **2021**, e9931985 (2021).
27. Du, D., Wang, J., Li, J. & Huang, J. Evolution of China's intercity venture capital network: Preferential attachment vs. path dependence. *Cities* **148**, 104874 (2024).
28. Zhang, W., Derudder, B., Wang, J. & Witlox, F. An Analysis of the Determinants of the Multiplex Urban Networks in the Yangtze River Delta. *Tijdschrift voor Economische en Sociale Geografie* **111**, 117–133 (2020).
29. Shu, T. *et al*. An intercity investment network dataset of China based on the enterprise registration records (2000–2020). *figshare* https://doi.org/10.6084/m9.figshare.28248503 (2025).
30. Dong, L., Yuan, X., Li, M., Ratti, C. & Liu, Y. A gridded establishment dataset as a proxy for economic activity in China. *Sci Data* **8**, 5 (2021).
31. Zhao, J., Zhang, X. & Song, J. The Changing Geography of Domestic Financial City Network in China, 1995–2015. *Growth and Change* **49**, 490–511 (2018).
32. Ministry of Civil Affairs of the People's Republic of China. Code of the Administrative Divisions of the People's Republic of China in 2020. https://www.mca.gov.cn/mzsj/xzqh/2020/20201201.html (2021).
33. DQinYuan. Chinese province city area mapper. https://github.com/DQinYuan/chinese_province_city_area_mapper (2024).
34. Amap. Geododing-Amap API. https://lbs.amap.com/api/webservice/guide/api/georegeo (2024).
35. Traag, V. A., Waltman, L. & Van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233 (2019).
36. Javorcik, B. S., Özden, Ç., Spatareanu, M. & Neagu, C. Migrant networks and foreign direct investment. *Journal of Development Economics* **94**, 231–241 (2011).
37. Anderson, J. E. The Gravity Model. *Annual Review of Economics* **3**, 133–160 (2011).
38. Amap. Ranking of population migration willingness of major Chinese cities. https://report.amap.com/migrate/page.do (2024).

39. Kang, Y. *et al*. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. *Sci Data* **7**, 390 (2020).
40. Ma, L., Xu, D., Liang, R. & Song, J. Exploring Spatial Connection Networks in Metropolitan Areas Led by Megacities: A Case Study of the Shanghai Metropolitan Area. *J. Urban Plann. Dev.* **149**, 04023029 (2023).
41. Anderson, J. E. *The Gravity Model of Economic Interaction*. 1–9 (2016).
42. Gan, C., Voda, M., Wang, K., Chen, L. & Ye, J. Spatial network structure of the tourism economy in urban agglomeration: A social network analysis. *Journal of Hospitality and Tourism Management* **47**, 124–133 (2021).
43. Zhao, Y., Zhang, G. & Zhao, H. Spatial Network Structures of Urban Agglomeration Based on the Improved Gravity Model: A Case Study in China's Two Urban Agglomerations. *Complexity* **2021**, 6651444 (2021).
44. Liang, Y., Gao, S., Cai, Y., Foutz, N. Z. & Wu, L. Calibrating the dynamic Huff model for business analysis using location big data. *Transactions in GIS* **24**, 681–703 (2020).
45. Bendtsen, C. pso: Particle Swarm Optimization. *R package version 1.0.4*. https://CRAN.R-project.org/package=pso (2022).
46. Elhorst, J. P., Tziolas, I., Tan, C. & Milionis, P. The distance decay effect and spatial reach of spillovers. *J Geogr Syst* **26**, 265–289 (2024).
47. Duranton, G. Distance, Land, and Proximity: Economic Analysis and the Evolution of Cities. *Environ Plan A* **31**, 2169–2188 (1999).
48. Zhang, Z. & Tang, Z. Examination and Interpretation of the Quantitative Validity in China's Corporate-based Urban Network Analysis. *Chin. Geogr. Sci.* **31**, 41–53 (2021).
49. Wasserman, S. & Faust, K. *Social Network Analysis: Methods and Applications*. (Cambridge University Press, Cambridge, 1994).
50. Yu, T., Shu, T. & Xu, J. Spatial pattern, and evolution of China's urban agglomerations. *Frontiers of Urban and Rural Planning* **2**, 7 (2024).

Acknowledgements

This work was supported by National Key Research and Development Program of China (2022YFC3800101) and by Polish National Science Centre (NCN, 2020/38/A/HS4/00312) and KU Leuven (C14/21/021).

Author contributions

T.S., T.Y. and X.L. conceived and designed the research; T.S., S.Y., X.L. and F.S. collected and processed the raw data; T.S., X.L., S.Y., G.C. and Y.R. analyzed and visualized the data; T.S., S.Y., T.Y., B.D. and X.L. validated the data; T.Y. and X.L. supervised the research; T.S., S.Y. and X.L. wrote the paper. All authors reviewed the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.Y. or X.L.