Lai et al. Computational Urban Science (2025) 5:46
https://doi.org/10.1007/s43762-025-00207-8

Computational Urban Science

ORIGINAL PAPER

Open Access

Characterizing intra-urban population migration networks: a case study of Shenzhen, China

Zhongyu Lai[1†], Yueshan Li[2†], Tao He[2] and Xintao Liu[1*]

Abstract

Intra-urban migration plays a crucial role in shaping urban structure and socio-economic dynamics. Most existing studies rely on small-scale survey data or have a coarse spatial resolution, making it difficult to conduct detailed network analysis at the urban scale to fully understand the complexity and dynamics of migration patterns. To address the gap, this study conducted a subdistrict-level fine-grained network analysis, involving more than 800,000 relocation data with detailed demographic and housing information at subdistrict levels in Shenzhen in 2015, to explore the overall relocation patterns and the relocation differences among different groups. The findings reveal that short-distance relocations dominate, with major hubs serving as central points of population flow in the study area (e.g., Gongming and Shajing areas). The relocation patterns also indicate specific pathways guiding movement between city areas. Moreover, demographic factors such as marital status, education level, and age significantly influence relocation behaviour. For instance, elderly individuals move infrequently, but when they do, they often relocate over longer distances. Men tend to migrate to diverse areas, while women prefer similar ones. Highly educated individuals move longer distances, typically within economic core areas. Overall, our study provides new perspectives for understanding the complex mechanisms of intra-urban population migration.

Keywords Intra-urban migration, Complex networks analysis, Relocation patterns, Demographic factors

1 Introduction

Intra-urban migration, as a crucial aspect of internal migration, involves the relocation of human and social capital within cities, significantly shaping the residential and labour market structures (Cui et al., 2014; Li & Dodson, 2023; Wu, 2006). Migration, viewed as an act of human capital investment, follows similar principles at global, national, and urban levels (Acharya & Leon-Gonzalez, 2014; Di Maria & Lazarova, 2012; Ye et al., 2024). In cities, urban resources such as housing, infrastructure, services, and employment are spatially unequal and constantly shifting (Fan & Xiang, 2020; Huang et al., 2021; Shen et al., 2022; Wu & Wang, 2021). Therefore, intra-urban migration not only reflects the spatial structure of a city but also indicates broader societal changes, including improvements in transportation and individual life trajectories (Brown & Holmes, 1971; Slavko et al., 2020; Sun & Manson, 2012). This article explores that Shenzhen, as a prime example, showcases how rapid urbanization interacts with intra-urban migration patterns with different factors.

Shenzhen's transformation from a small town of 300,000 to a megacity of over 17 million people within just 40 years implies the large scale of population movement within the city (Cheng et al., 2023; Goodburn, 2020;

[†]Zhongyu Lai and Yueshan Li contributed equally to this work.

*Correspondence:
Xintao Liu
xintao.liu@polyu.edu.hk
[1] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, 181 Chatham Road South, Hung Hom, Kowloon, Hong Kong, China
[2] Shenzhen Aerospace Smart City System Technology Co., Ltd, Shenzhen, China

Michael Keith and Rooker, 2013; Lai et al., 2020; Wang et al., 2009). Its population increased nearly 60-fold with an average annual growth rate of 10.3%, driven by government policies that encouraged a vibrant private economy, technological innovation, as well as the city's strategic geographic location, i.e., near Hong Kong (Cheng et al., 2023). As a result, this city shaped by migration has complex internal migration networks (Wang, 2022).

The analysis of this network, using comprehensive 2015 grid community registration dataset, offers a detailed view of migration flows within the city. After preprocessing, this dataset includes key variables such as migration patterns, age distribution, marital status, housing conditions, education levels, and more. This detailed, large-scale dataset provides a unique opportunity to overcome the coarse spatial resolution and small sample sizes that have limited previous studies of intra-city population migration (Lin et al., 2021; Nora, 2023). Therefore, this research uses this fine-grained data to analyze the intra-urban migration network of Shenzhen, aiming to provide a detailed characterization of its patterns and to reveal how these population movements reflect and shape the city's underlying urban structure. To achieve this, the study addresses three key questions:

(a) What are the patterns of population displacement within cities? Are there differences between regions?
(b) Are there critical nodes and specific pathways in relocation?
(c) Are there differences in the migration patterns of different groups?

This study offers insights into the network structure of intra-urban relocation in Shenzhen, providing a theoretical framework for understanding how internal migration, based on different groups of people, shapes urban spaces in rapidly developing cities.

2 Literature review

2.1 Population migration

Since the 1960s, migration has become a key issue in urban studies, seen as a form of human capital investment reflecting individual choices within diverse geographic contexts, including both relocation and the decision to stay (De Haas, 2021). Research has largely focused on national migration, examining factors like income, population density, education, labour structure, and age demographics (Treyz et al., 1993). In China, a rich amount of studies analysis the national migration patterns, highlighting the influence of housing prices, pollution, and institutional reforms on migration patterns (Shi et al., 2020; Zhou & Hui, 2022; Liu et al., 2020). Network analysis is increasingly used in migration research, for example, Shen et al. (2022) examine the population flow patterns in the Sichuan-Chongqing region during the Covid-19, finding out the regional population flow intensity is strong correlated with the population education level. Nonetheless, majority of such kind of research focus on the national level (Lin et al., 2021; Qi et al., 2021).

Despite the significance of intra-city migration for urban development, big data studies on this topic remain limited (You et al., 2023). Most research relies on surveys and interviews, examining how economic conditions, family factors, and job opportunities influence urban mobility (Fan & Xiang, 2020; Wu & Wang, 2021). For example, Liu et al. (2021) found that high accessibility to public and commercial resources drives relocation, while He and Zhang (2023) showed that individuals with low cultural capital tend to move from city centres to suburbs, and those with high cultural capital but low affordability relocate from suburbs to urban cores.

While these studies offer valuable insights into migration factors, they are often constrained by small sample sizes, limiting their ability to capture large-scale population relocations and build comprehensive city-wide migration networks (Nora, 2023). This highlights the need for big data methodologies to construct holistic networks for intra-city migration, addressing the structural and dynamic complexities of urban mobility.

2.2 Migration network

Flow and network analysis has emerged as a prominent methodology in the whole field of geography, especially in migration studies, particularly within national and international contexts (Pei et al., 2024; Zhang et al., 2023). For instance, Zhang et al. (2020) utilized Tencent's location big data to examine the spatial distribution characteristics of population mobility in China, employing the Exponential Random Graph Model to examine the distinctive features including key attractive nodes and their associated influencing factors. Similarly, Charyyev and Gunes (2019) conducted a dynamic analysis of the complex network of domestic migration in the United States, investigating how economic conditions and political factors over 15 years impacted the evolution of the migration network.

Furthermore, Pitoski et al. (2021) constructed a migration network for Austria, exploring the mobility flows between three scales of human settlements: cities, towns, and villages. Schon and Johnson (2021) developed a migration network for refugee flows from 1991 to 2016 and employed the Multiple Regression-Quadratic Assignment Procedure (MR-QAP) model to identify the significant variables associated with refugee mobility on an annual basis.

Although these studies have constructed large-scale migration patterns using big data, providing intuitive insights into network characteristics, they often overlook the internal characteristics of populations, such as gender and educational level (Gong et al., 2024). This gap is essential, as intra-group differences may significantly influence migration decisions and network structures, providing a micro perspective of the migration Therefore, it is necessary for network research to be integrated with potential associations with demographic characteristics (Rogers, 2008). This connection would not only enrich the theoretical framework of network analysis but also provide a more comprehensive perspective for policy formulation, aiding in the understanding of the dynamic changes and influencing factors affecting various populations during the migration process (Bon et al., 2018; Xu et al., 2021). Future research should focus on how to integrate these dimensions to develop a more holistic and nuanced view of migration studies.

2.3 Research objective and significance

This study aims to achieve several objectives: First, exploring migration patterns by examining the distances, directions, and frequencies of movements among different groups, revealing the factors driving migration decisions such as economic opportunities and access to urban services. Second, the research will assess the structural features of Shenzhen's migration network, identifying critical nodes and flow patterns through network analysis techniques. Third, it will examine the formation of migration communities within the urban landscape, elucidating the socio-spatial dynamics and how specific areas serve as migration hubs.

This research is significant for its potential to deepen our understanding of intra-urban relocation's impact on urban environments, particularly in rapidly urbanizing areas like Shenzhen. Consequently, the network of population relocation exhibits complexity and uncertainty (Zhao et al., 2024). By utilizing a big data approach, it addresses the critical gap in existing migration studies regarding intra-urban relocation networks. The nuanced data analysed will enhance migration theory on the basis of network analysis, providing valuable insights for policymakers and urban planners, informing strategies for managing urban growth, optimizing resource allocation, and improving residents' quality of life.

3 Study area and data

3.1 Study area

Shenzhen, located in southern China at N22°32′43.86″, E114°03′10.40″ near Hong Kong, has undergone a remarkable transformation over the past four decades, evolving from a small fishing village into a leading global megacity. Established as a Special Economic Zone (SEZ) in 1980, Shenzhen became a central hub for China's economic reforms, drawing in significant domestic and foreign investments (Cheng et al., 2023; Goodburn, 2020; Michael Keith and Rooker, 2013; Lai et al., 2020; Wang et al., 2009). Shenzhen is one of the world's most dynamic cities in terms of population movement and urban development, for its population of over 17 million, and more than 80% of them are migrants (Cheng et al., 2023).

The city is divided into 9 districts and 74 subdistricts, covering a total area of 1,997.47 square kilometres (Fig. 1). This research will focus on the subdistrict-level relocation networks inside Shenzhen, giving critical insights on intra-urban migration patterns. Shenzhen's unique feature as a migrant-dominated city and its rapid urbanization make it not only an ideal case study for understanding the challenges and dynamics of urbanization in China, but also a valuable lesson for cities facing similar issues globally.

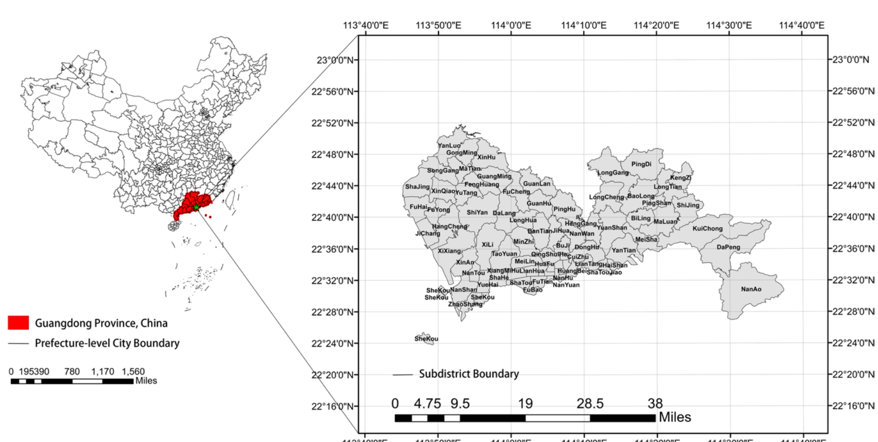


Fig. 1 The study area Shenzhen, China, including 74 subdistrict as spatial units

3.2 Datasets

The data used in this study include the Shenzhen building census dataset and the citywide grid community registration dataset. The Building Census Dataset recorded the locations and IDs of 644,896 buildings across Shenzhen in 2015. These datasets allow for precise tracking of relocation behaviours, identifying both the origin and destination points of each move to build a detailed origin-destination (OD) dataset.

Grid community registration data are collected through a variety of ways, including self-registration by residents and door-to-door registration by grassroots workers, and cover more than 19 million pieces of data from 2010 to 2015. This dataset comprises eleven attributes: ID, gender, birth date, household registration address, marital status, industry of employment, education level, residence address, building number, residence type, and living arrangement. It is needed to distinguish between the 'household registration address' (Hukou), which is a person's official place of registration often tied to social services, and the 'residential address', which is their actual physical dwelling. In this study, migration is tracked by changes in the residential address. The user ID is encrypted for privacy reasons, while the other fields have a complete data dictionary (Table 1), providing rich demographic and socio-economic information for the study.

From the available data from 2010 to 2015, this study deliberately selects the 2015 records, justified by their unprecedented completeness. A look at the Table 2 shows that an order-of-magnitude leap in data volume for 2015. We attribute this surge not to a statistical anomaly but to a key policy shift from Guangdong Provincial People's Government (2020): the Shenzhen Special Economic Zone Residence Permit Regulation, effective June 1, 2015. To support this new regulation Linking public

Table 1 Grid community registration data dictionary

Variable Name	Label	Codes/Sample
ID	Unique id that identifies the resident	The encrypted string.
BIRTHDAY	Date of birth	N/A
GEDER	Sex	1 = Male
		2 = Female
REGISTERAD	Household registration location (Hukou)	The official, legal place of registration under China's household registration system.
MARRYID	Marital status	1 = Unmarried
		2 = Married
EDULEVEL	Education level	0 = PhD
		1 = Postgraduate
		2 = Undergraduate
		3 = College
		4 = Secondary School
		5 = High School
		6 = Junior high school
		7 = Below junior high school
DOMICILETY	Domicile type	1 = Renting a house
		2 = Owning a house
BIDEFASHION	Living arrangement	1 = Living alone
		2 = Living with family
		3 = Living with roommates
		4 = Living in collective housing
TRADEID	Industry of employment	1 = Industrial
		2 = Agriculture
		3 = Services
		7 = Others
		8 = Unemployed
HOUSECODE	Building number (foreign key linked to building data)	Sample: '4403030020130300012000056'
HOMEADDRES	Actual Residential Address	The physical address where the resident currently lives.
INSERTTIME	Date of Data Addition	Range: 2010 - 2015

Table 2 Annual volume of Shenzhen grid community registration records (2010-2015)

Year	Number of Records
2010	73,579
2011	143,031
2012	206,906
2013	412,461
2014	1,836,623
2015	16,357,068

services to residency, the city completed the construction and trial operation of a team of grid information officers and a community information collection system in late 2014, and fully implemented a standardised, full-coverage grid-based information collection in 2015 (General Office of the CPC Shenzhen Municipal Committee and General Office of Shenzhen Municipal People's Government, 2014). This policy-driven effort produced a near-census-level population snapshot, with completeness and coverage far exceeding other Years that relied on non-systematic, non-compulsory registration. Therefore, using the 2015 data provides the most accurate depiction of the intra-urban migration network while minimizing sampling bias. Given the central role of this dataset in our analysis, a thorough investigation of its representativeness and potential biases is conducted and presented in Section 3.4.

3.3 Data cleaning

To ensure the accuracy and reliability of the data, this study filtered the 2015 records from the Grid Community Registration Dataset and removed entries with missing or incorrect information. Individuals were identified by a unique combination of gender, birth date, and household registration address to track each resident's registration history. Only individuals with two or more different residential registrations within 2015 were defined as having relocated. While this conservative definition intentionally excludes potential relocations where the origin was registered prior to 2015, it is a necessary methodological choice. It ensures that both the origin and destination for every analyzed move are drawn from the same dataset, which is characterized by its uniform, high-completeness collection method. This approach guarantees the internal consistency and reliability of the relocation events identified for our network analysis. A relocation event was defined as the movement from one registered address to another. Additionally, building IDs from the relocation records were matched with the Building Census Dataset to determine the exact locations of the origin and destination for each relocation. These preprocessing steps ensured the accuracy and completeness of the relocation data. After this cleaning process, our final dataset contains 814459 relocation records involving 770701 unique individuals, which form the basis of our subsequent analysis.

3.4 Data representativeness and bias analysis

To critically evaluate the validity of our findings, we conducted a two-layer comparative analysis of our dataset's representativeness against the official city-wide demographics reported in the *Shenzhen 2015 National 1% Population Sample Survey Communiqué*. We first assessed the source grid registration dataset and then the final relocation sample used in our analysis.

Table 3 compares the source dataset (approx. 15.5 million records) with the official statistics. The source data already shows some bias: it over-represents the working-age population (15-64 years) and slightly over-represents

Table 3 Comparison of the source registration dataset with Shenzhen official statistics (2015)

Attribute	Source dataset (N=15,499,783)	Official statistics (N=11,378,700)	Bias interpretation
Gender			
Male	54.47%	53.61%	Minor over-representation
Female	45.52%	46.40%	Minor under-representation
Age Structure[a]			
0-14 years	11.04%	13.40%	Under-represented
15-64 years	86.23%	83.23%	Over-represented
65+ years	2.07%	3.37%	Under-represented
Education Level[b]			
University or above	16.87%	22.67%	Under-represented
High school	19.82%	25.29%	Under-represented

[a] Official age brackets are 0-14, 15-64, and 65+ years

[b]Official education statistics are for the population aged 6 and over. Our data is for the entire sample

Table 4 Comparison of the relocation sample with Shenzhen official statistics (2015)

Attribute	Relocation sample (N=770,701)	Official statistics (N=11,378,700)	Bias Interpretation
Gender			
Male	57.93%	53.61%	Significant over-representation
Female	42.06%	46.40%	Significant under-representation
Age Structure[a]			
0-14 years	5.58%	13.40%	Strongly under-represented
15-64 years	93.33%	83.23%	Strongly over-represented
65+ years	0.68%	3.37%	Strongly under-represented
Education Level[b]			
University or above	18.20%	22.67%	Under-represented
High school	20.95%	25.29%	Under-represented

[a] Official age brackets are 0-14, 15-64, and 65+ years

[b] Official statistics are for the population aged 6 and over. Our data is for the entire sample

males, while under-representing children, the elderly, and individuals with higher education. This indicates that the grid registration system itself was more effective at capturing the city's active labor force.

More importantly, Table 4 examines the final relocation sample (N=770,701) that forms the basis of this study. The biases observed in the source data are substantially amplified in this analytical sample. The relocating population is significantly more male-dominated (57.93% vs. 53.61% city-wide). Most strikingly, the working-age population constitutes an overwhelming 93.33% of the sample, while children and the elderly are severely under-represented (5.58% and 0.68% respectively). This pivotal finding clarifies that our study's conclusions are primarily a detailed characterization of the intra-urban mobility of Shenzhen's core labor force, rather than its entire residential population.

In addition to demographic bias, we assessed the spatial representativeness of our data. Due to the unavailability of official population statistics at the subdistrict level for 2015, this analysis was conducted at the district level. We calculated a sampling rate for each district by comparing the number of records in our initial registration dataset against the official 2015 district-level population data, with the results visualized in Fig. 2. The analysis reveals a clear spatial bias. The suburban, industrial districts in the north and west exhibit significant over-sampling, such as Guangming (sampling rate of 2.05), Longgang (1.75), Longhua (1.60), and Bao'an. Conversely, the core urban districts of Nanshan (0.74) and Futian (0.86) are under-sampled. This indicates that our data source was most effective at capturing the population in high-density, migrant-heavy areas, which are also the epicenters of manufacturing and labor-intensive industries. While this district-level analysis may mask intra-district variations, it provides crucial context for interpreting our findings. The implications of these demographic and spatial biases are further addressed in the Sections 6.1 and 6.5.

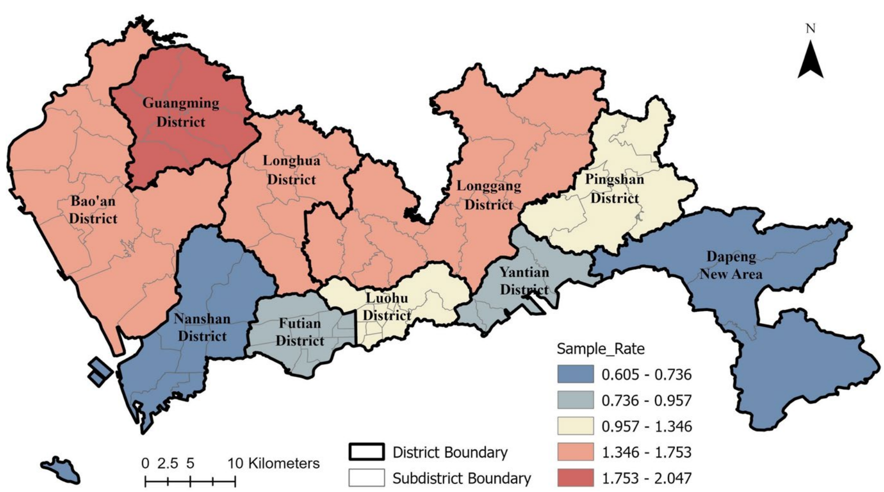


Fig. 2 Spatial Distribution of Sampling Rate at the District Level (2015)

4 Methodology

4.1 Research framework

The research framework and process are described in Fig. 3. The study begins by processing and cleaning the Grid Community Registration and Building Census datasets, followed by constructing a relocation network based on the derived migration records. Next, spatial analysis of overall relocation is performed, including both intra-subdistrict and inter-subdistrict migration, migration distance, and spatial differentiation in relocation behaviour. Network centrality metrics such as weighted degree, degree centrality, betweenness, and closeness centrality are calculated to identify critical nodes and paths of population flow. In parallel, community detection is applied to discover small migration communities. The final part of the framework focuses on group-specific relocation analysis. Sixteen demographic groups are considered, and for each group, hierarchical backbone extraction and network metrics are analysed to understand the differences in migration networks and spatial behaviour across various population groups.

4.2 Network construction

Based on graph theory, a migration origin-destination (OD) network is constructed and illustrates spatial relationships and enables efficient analysis. Subdistricts are nodes, with migration behaviours as directed edges weighted by relocation counts. Data consistency was ensured by matching building-scale relocation records and cleaning grid community data to the subdistrict scale. Only active nodes (non-zero in-degree or out-degree) and edges were retained, forming a weighted directed graph.

4.3 Migration intensity

In-Relocation Rate (R_{in}) and Out-Relocation Rate (R_{out}) are derived from the network's weighted degree. The raw

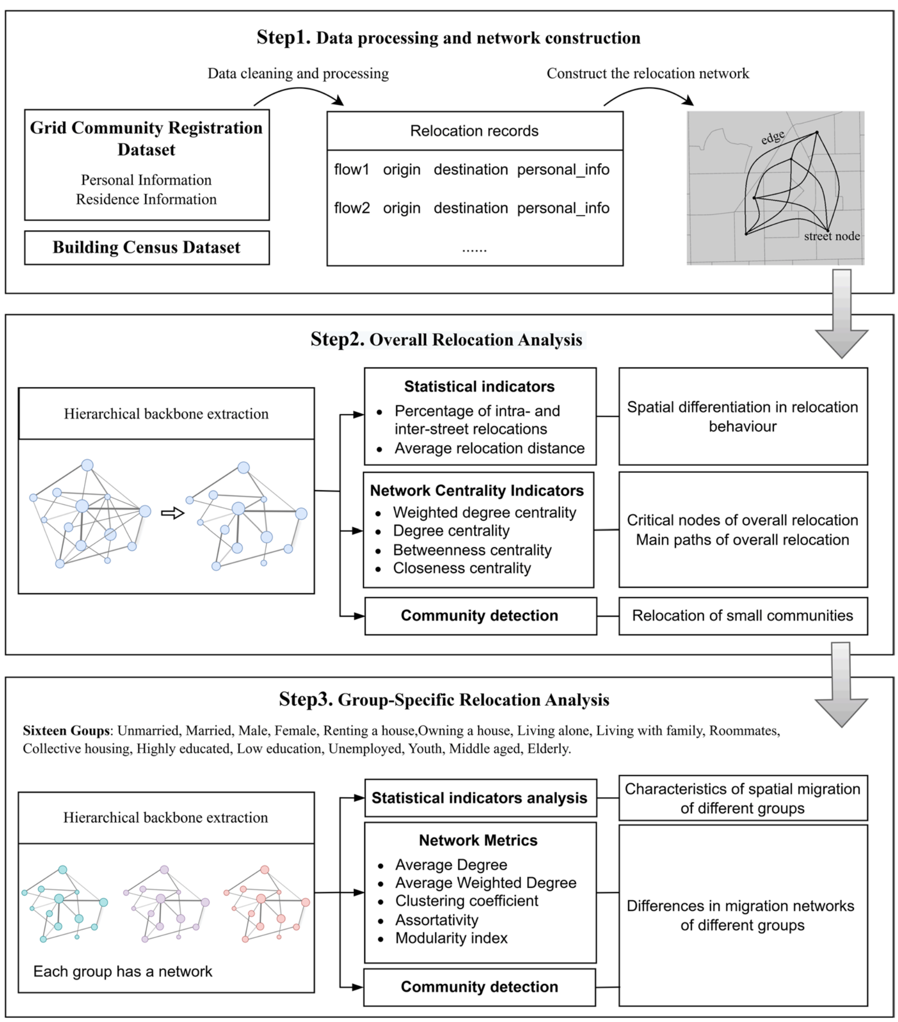


Fig. 3 Research framework

weighted in-degree ($\sum w_{ji}$) and out-degree ($\sum w_{ij}$) represent the total volume of people moving into or out of a subdistrict, respectively. However, these raw volumes are heavily influenced by a subdistrict's population size, masking the actual intensity of migration. To create a standardized and comparable measure of migration intensity, we normalize these volumes by the population of each subdistrict. This approach correctly measures the per-capita intensity of migration, allowing for meaningful comparisons between subdistricts of different sizes.

$$R_{\text{in}}(i) = \frac{\sum_{j \in N_{\text{in}}(i), i \neq j} w_{ji}}{p_i} \times 1000 \tag{1}$$

$$R_{\text{out}}(i) = \frac{\sum_{j \in N_{\text{out}}(i), i \neq j} w_{ij}}{p_i} \times 1000 \tag{2}$$

$$R_{\text{total}}(i) = R_{\text{in}}(i) + R_{\text{out}}(i) \tag{3}$$

where, the term $i \neq j$ clarifies that we are considering inter-subdistrict flows. p_i is the registered population of subdistrict i. Rates are expressed per 1,000 persons.

The Net Relocation Rate (R_{net}) is the difference between the in-migration and out-migration rates. A positive value indicates the subdistrict is a "population magnet", attracting more people than it loses on a per-capita basis. A negative value signifies it is a "population source", experiencing a net outflow. This rate is crucial for identifying the primary directions of population redistribution across the city.The Internal Relocation Rate (R_{Internal}) measures the intensity of residential moves that occur within the boundaries of a single subdistrict. This metric reveals the level of internal residential churn. A high rate suggests a transient, high-turnover community, whereas a low rate points to a stable, established residential area.

$$R_{\text{net}}(i) = R_{\text{in}}(i) - R_{\text{out}}(i) \tag{4}$$

$$R_{\text{Internal}}(i) = \frac{w_{ii}}{p_i} \times 1000 \tag{5}$$

4.4 Network metrics

Weighted degree Weighted degree centrality considers the edge weights in the network, providing a measure of the total strength of a node's connections. Weighted in-degree centrality measures the sum of the weights of edges directed toward a node, quantifying a region's ability to attract incoming migrants. Weighted out-degree centrality measures the sum of the weights of edges originating from a node, representing the intensity of population outflow from an area (Stamos, 2023). To provide a standardized measure of a subdistrict's migration intensity that is not dependent on the total number of subdistricts, we normalize the raw strength. The normalisation process is done by dividing the raw intensity by $N-1$, the maximum number of other nodes to which a given node can be connected. This standardization controls for network size, ensuring that a high centrality value reflects significant migration activity relative to the opportunities for connection, not just an artefact of a large network.

$$C_{\text{in}}(i) = \frac{\sum_{j \in N_{\text{in}}(i)} w_{ji}}{N-1} \tag{6}$$

$$C_{\text{out}}(i) = \frac{\sum_{j \in N_{\text{out}}(i)} w_{ij}}{N-1} \tag{7}$$

$$C_{\text{total}}(i) = C_{\text{in}}(i) + C_{\text{out}}(i) \tag{8}$$

where, $N(i)$ refers to the set of all nodes directly connected to node i, $N_{\text{in}}(i)$ refers to the set of all nodes pointing to node i, and $N_{\text{out}}(i)$ refers to the set of all nodes that node i points to. w_{ij} is the weight of the edge from node i to node j. N is the total number of nodes in the network.

Degree, Betweenness, and Closeness Degree centrality (C_D) measures the number of direct connections each subdistrict has, indicating the relative importance and central position of each area within the migration flow (Cai et al., 2019).Betweenness centrality (C_B) identifies nodes that frequently lie on the shortest paths between other nodes, highlighting areas that serve as "bridges" or connectors within the network, and facilitating movement between different regions (Brandes, 2001). Closeness centrality (C_C) evaluates how quickly a node can reach other parts of the network, representing the strategic positioning of nodes that can efficiently access all other areas (Okamoto et al., 2008). Together, these three measures allow for a nuanced understanding of each area's influence and connectivity within Shenzhen's migration patterns.

$$C_D(i) = \frac{k_i}{N-1} \tag{9}$$

$$C_B(v) = \sum_{i \neq j \neq t} \frac{\sigma_{ij}(v)}{\sigma_{ij}} \tag{10}$$

$$C_C(i) = \frac{N-1}{\sum_{i \neq j} d\left(i,j\right)} \tag{11}$$

where, k_i represents the degree of node i (for a directed graph, it is the sum of in-degree and out-degree). σ_{ij} represents the number of shortest paths from node i to node j, $\sigma_{ij}(v)$ denotes the number of those shortest paths that pass through node v. $d\left(i,j\right)$ represents the shortest path length from node i to j.

Modularity index Modularity is a measure that evaluates the quality of the division of a network into modules (or communities). A network with high modularity indicates that the nodes within each module are densely connected compared to what would be expected by random chance, while connections between different modules are relatively sparse (Fang et al., 2022; Poisot, 2013).

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m}\right] \delta(c_i, c_j) \tag{12}$$

where, A_{ij} represents the adjacency matrix of the network, where $A_{ij} = 1$ if there is an edge between nodes i and j, and $A_{ij} = 0$ otherwise. m is the total number of edges in the network.$\delta(c_i, c_j$ is an indicator function that equals 1 if nodes i and j belong to the same community, and 0 otherwise.

Clustering coefficient The clustering coefficient measures the degree of connectivity among a node's neighbours, with a higher clustering coefficient typically indicating a tight-knit community or group (Saramäki et al., 2007). The local clustering coefficient is calculated for individual nodes, while the network's overall clustering coefficient is defined as the average of the local clustering coefficients of all nodes.

$$C_i = \frac{2T_i}{k_i(k_i - 1)} \tag{13}$$

$$C = \frac{1}{N} \sum_{i=1}^{n} C_i \tag{14}$$

where T_i represents the number of actual triangular relationships (fully connected subgraphs of three nodes) between the neighbors of node i. The factor of 2 accounts for each triangle being counted twice (once for each direction).

Assortativity Assortativity describes the tendency of nodes to connect with similar or dissimilar types of nodes (Noldus & Van Mieghem, 2015). A positive value indicates a tendency for similar nodes to connect, while a negative value suggests a preference for connections between nodes of different types.

$$r = \frac{\sum_{xy} xy(e_{xy} - q_x q_y)}{\sigma_q^2} \tag{15}$$

where e_{xy} represents the proportion of node pairs in the graph with degrees x and y. q_x denotes the expected proportion of nodes with degree x, which is calculated as $q_x = \sum_y e_y$. σ_q^2 is the variance, calculated as $\sigma_q^2 = \sum_x x^2 q_x - (\sum_x x q_x)^2$.

4.5 Hierarchical backbone extraction

Migration networks are complex, with numerous nodes and chaotic edges. Identifying backbone structures is crucial for optimizing functionality and visualizing architecture. The Disparity Filter algorithm extracts multi-scale weighted network backbones, revealing core structures (Serrano et al., 2009). To enable effective analysis of network structure and performance, the methodology involves five key steps:

(1) Calculate node strength. For each node in the network, compute its strength s_i. Node strength refers to the sum of the weights of all connections associated with that node, expressed as:

$$s_i = \sum_{j \in \mathcal{N}(i)} w_{ij} \tag{16}$$

(2) Calculate local significance of weights. For each edge (i, j) of a node i, calculate the local significance of the edge's weight relative to the weights of all other edges. This is measured by a probability value defined as:

$$p_{ij} = 1 - (k_i - 1) \int_0^{w_{ij}/s_i} (1-x)^{k_i - 2} dx \tag{17}$$

(3) Select significance threshold. Establish a significance level α. This indicates the probability that the normalized edge weight is greater than or equal to a specified value p_{ij} within the framework of the null model. Typically, this value is determined based on the network being studied and the specific application area.

$$\alpha_{ij} = \left(1 - p_{ij}\right)^{k-1} \tag{18}$$

(4) Filter edges. For each edge (i, j), by setting a significance level α (between 0 and 1), any links α_{ij} with normalized weights p_{ij} greater than the specified threshold α will be filtered out.

(5) Construct network backbone. Repeat the above process until all edges have been examined. The edges that remain after this filtering process constitute the backbone of the network.

4.6 Community detection

In network research, community detection is a powerful analytical technique that moves beyond examining individual locations (nodes) and their direct connections (edges) to reveal the network's meso-scale structure (Yang et al., 2016). Specifically, this technique identifies clusters of subdistricts (i.e., communities in the network) that are more densely interconnected by migration flows internally than they are with the rest of the city. In the context of intra-urban migration, these communities represent "functional regions" or self-sufficient "residential communities" of city, which may not align with predefined administrative boundaries (Danchev & Porter, 2021). This analysis contributes to a better understanding of the sub-regional dynamics of population turnover, thereby enabling the examination of urban fragmentation and integration. It highlights which areas function as cohesive residential zones and where the boundaries between them lie.

For this purpose, we employed the Louvain algorithm, a hierarchical community detection algorithm based on a greedy optimization strategy, designed to uncover community structure in large networks. Its core principle is the optimization of the modularity metric (Zhang et al., 2021; Blondel et al., 2008).

(1) Initialization. Initially, each node is treated as an independent community.
(2) Local optimization. The algorithm tries to move each node into a neighbouring community, calculating the modularity gain to see if the move would increase overall modularity. If it does, the move is executed.
(3) Network aggregation. Once local moves no longer increase modularity, the algorithm"aggregates" the communities found into single nodes and constructs a new"reduced" network.
(4) Repeat steps. Steps (1) and (2) are repeated on the reduced network until modularity index is maximized, meaning no further community merging can increase modularity.
(5) Termination. The algorithm ends when no further modularity increases can be achieved through merging. The resulting community partition is considered optimal.

5 Analysis results

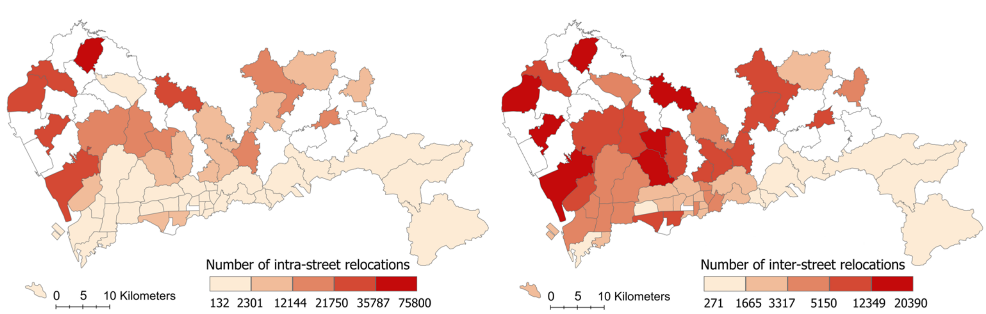
This study analysed population migration data in Shenzhen, focusing on relocation records from 2015. According to the statistics, a total of 770,701 residents in Shenzhen engaged in at least one relocation during this period, resulting in 814,459 independent relocation flows. These figures reflect the high level of population mobility in Shenzhen, a rapidly developing metropolis.

5.1 Overall migration dynamics

5.1.1 Spatial patterns of intra-urban migration

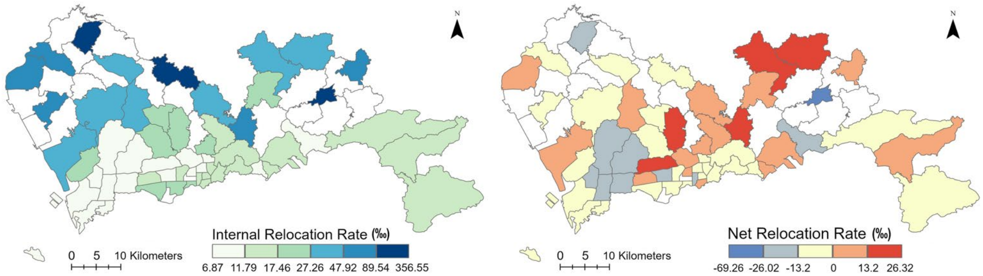
In terms of relocation types, intra-subdistrict moves (58.8%) are more frequent than inter-subdistrict ones. The absolute volume of relocations (Fig. 4) shows that peripheral industrial zones around Shenzhen are at the centre of the absolute amount of relocation. The Gongming and Xixiang subdistricts in Bao'an and Longhua dominate both internal and external relocation. However, this volume-based view is skewed by large population sizes and masks the underlying behaviors.

A more revealing picture emerges when flows are normalized by population to assess migration intensity (Fig. 5). The internal relocation rate (Fig. 5a) exposes the internal residential dynamics of the subdistrict. This rate is extraordinarily high in peripheral hubs like Gongming (356.5‰) Guanlan (189.3‰) and Pingshan (168.3‰). This phenomenon is attributed to their function as manufacturing bases, which attract a large,



(a) Number of intra-subdistrict relocations (b) Number of inter-subdistrict relocations

Fig. 4 Spatial distribution of intra-subdistrict and inter-subdistrict relocations



(a) Internal relocation rate (per 1000 persons) (b) Net relocation rate (per 1000 persons)

Fig. 5 Spatial distribution of normalized migration rates. **a** Internal mobility rate, showing the intensity of residential churn within subdistricts. **b** Net migration rate, revealing population magnets (positive values) and sources (negative values) on a per-capita basis. Rates are expressed per 1,000 registered persons

transient workforce housed in low-cost, high-turnover urban villages and factory dormitories. Residents frequently move for small differences in rent, or to get a new place closer to a new factory, but their lives and work circles are "locked" within the larger area. This environment facilitates a "hyper-local" mobility pattern of frequent, short-distance moves. In contrast, mature urban cores such as Lianhua Subdistrict in Futian (7.5 per 1,000) have very low internal relocation. This indicates that these areas have stable residential communities with high relocation costs, and that moving is a major decision that does not happen easily.

Furthermore, the net migration rate (Fig. 5b) identifies the city's true population magnets and sources. Emerging suburban areas in Longgang, such as Henggang (26.3‰) and Longgang (18.0‰), are the primary beneficiaries of urban expansion, attracting residents seeking lower costs and better living conditions. Conversely, significant net outflows occur in two distinct areas. First, mature, high-cost urban cores like Yuehai (−26.0‰) act as "population sources" due to a classic centrifugal push from high living expenses. Second, industrial hubs like Pingshan (−69.3‰) and Gongming also show strong net outflows. Their high churn and negative net migration suggest they function as "stepping-stones" or transient processing hubs for migrants, rather than as final destinations.

This interpretation is strongly reinforced by residents' relocation preferences (Fig. 6). The high degree of self-containment in peripheral hubs (over 70% of moves in Gongming are internal) confirms their nature as distinct, churning local economies. Conversely, the outward orientation of residents in the urban core, where over 80% of moves are to other subdistricts, solidifies their role as source communities whose populations are actively dispersing across the metropolis.

5.1.2 Distribution of relocation distances

Relocation distance was calculated as the Euclidean distance between the geographical centroids of the origin and destination buildings. For relocations within the same building, the distance was considered zero. In terms of relocation distances, 0.6% of relocations occurred

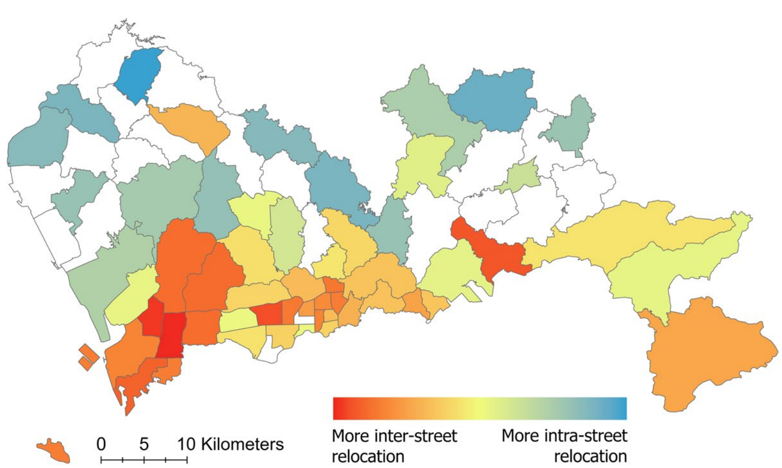


Fig. 6 Distribution of intra-subdistrict and inter-subdistrict relocation preferences

within the same building (56,928 cases), and approximately 12% of relocations were within 5 kilometres. The curve in Fig. 7 shows a sharp increase near the 10-kilometer mark and a cumulative distribution function value close to 0.75, indicating that about 75% of the relocations occurred within 10 kilometres or less.

Figures 8 and 9 visualize the average in-relocation distance and average out-relocation distance for each subdistrict in Shenzhen. It can be noticed that most of the subdistricts have similar move-in distances and move-out distances, with small differences in the distribution of distances between the source of the population moving in and the destination of the population moving out. Subdistricts located in the eastern suburbs of Shenzhen, such as Nan'ao, Dapeng, and Kui Chong, are farther away from other subdistricts, and have larger average in- relocation and out-relocation distances. Donghu Subdistrict in Luohu District, as well as most subdistricts in Bao'an, Longhua and Guangming Districts have smaller average move-in and move-out distances, and population movement is more localized. At the same time, the average in-movement distance in Pingdi, Longgang and Longcheng subdistrict is slightly greater than the average out-movement distance, characterized by "long-distance attraction and short-distance mobility".

5.1.3 Relocation network structure

Based on the relocation data of Shenzhen residents, this study generated a subdistrict-level relocation network for 2015 and visualized its structure. The Fig. 10 below shows the backbone of the relocation network. High-volume relocations were concentrated in the northeastern and northwestern outskirts of Shenzhen, particularly between several subdistricts in Bao'an District, where relocation flows were significantly higher than in other areas. The inter-subdistrict relocation network within Longhua and Longgang districts was also robust, though its structure was relatively simpler. In the core areas of Shenzhen—Nanshan, Futian, and Yantian—the network was less intense in terms of volume, but the relocation flows were densely concentrated, indicating frequent resident movements between these areas. This could be due to their geographical proximity or specific economic and social ties linking these regions.

Additionally, several key relocation pathways were identified. The primary relocation route between the eastern and central areas of the city runs from Gongming to Songgang, Shajing, Fuyong, Xixiang, and Xin'an, accounting for 5% of all relocations. A secondary route connecting the east and the city center extends from Gongming to Shiyan, Xixiang, and Xin'an. In the areas of Guanlan, Longhua, Dalang, Meilin, and

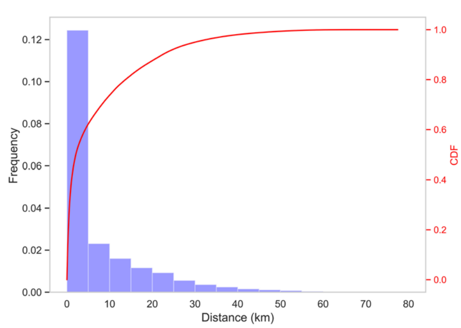
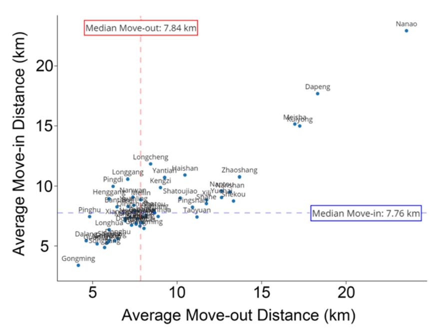
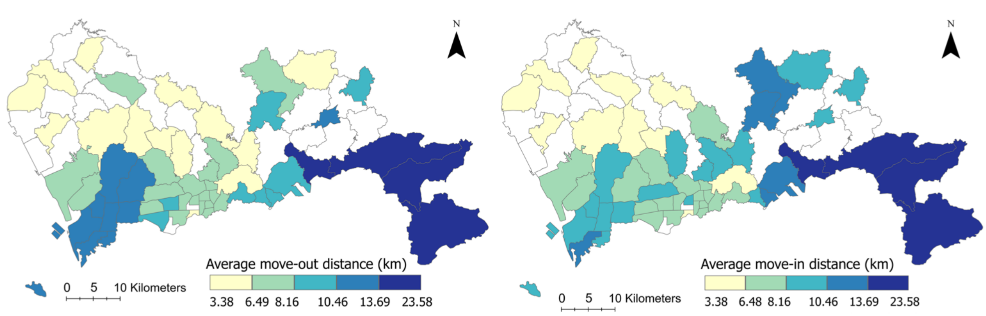


Fig. 7 Migration distance distribution and CDF curve





(a) Average move-in distance (km)

(b) Average move-out distance (km)

Fig. 8 Average move-in and move-out distances across subdistricts

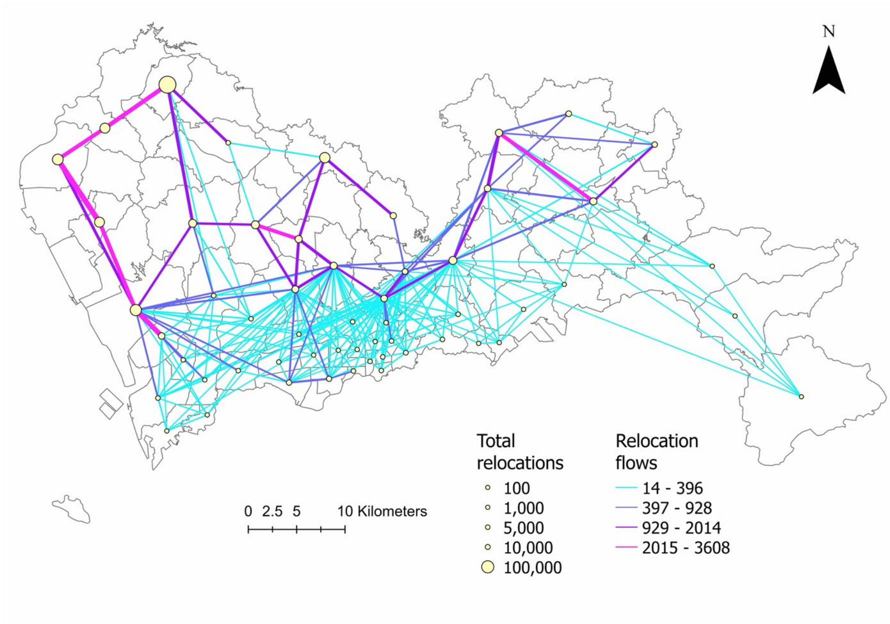


Fig. 9 Spatial distribution of average distance of intra-subdistrict and inter-subdistrict relocations

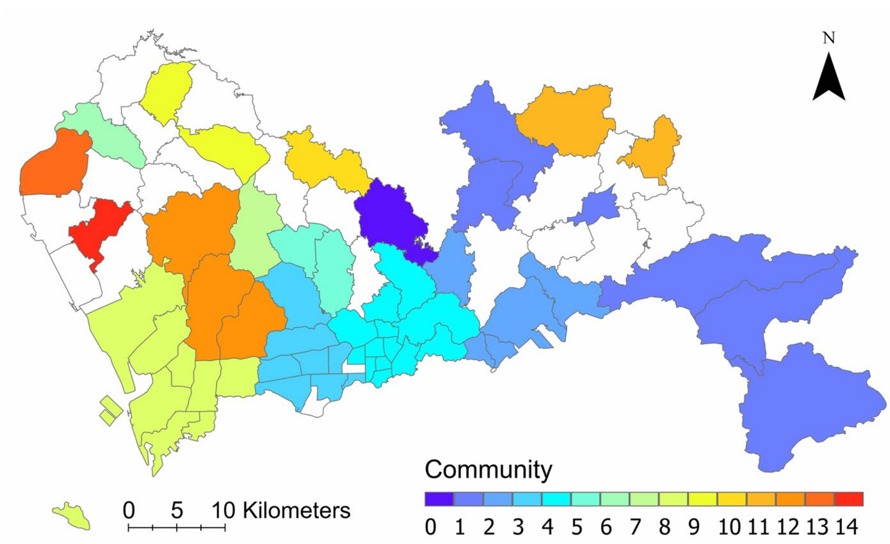


Fig. 10 Structure of the relocation network for all inhabitants

Bantian, relocation flows were similarly strong, with a variety of routes available for residents. The main relocation pathway between the western areas and the city centre follows the route of Kengzi, Pingshan, Longgang, Henggang, and Buji.

5.1.4 Key nodes and communities in the relocation network

Analysis of network centrality (Table 5) reveals that a subdistrict's importance is multi-faceted, distinguishing between nodes that are central due to their structural position versus their flow dynamics. Bantian emerges as the primary structural hub, ranking first in both Degree and Betweenness Centrality. This establishes it as the network's most critical bridge, connecting the most diverse set of subdistricts. Its role is a direct consequence of its geographic position, linking the major residential and industrial zones of Longhua and Longgang with the core districts of Futian and Luohu, and

Table 5 Centrality measures for different subdistricts

Rank	Degree	Relocation rate	Weighted degree	Betweenness	Closeness
1	Bantian	Gongming	Gongming	Bantian	Bujie
2	Bujie	Pingshan	Shajing	Xixiang	Henggang
3	Henggang	Guanlan	Xixiang	Bujie	Bantian
4	Minzhi	Fuyong	Guanlan	Henggang	Xixiang
5	Xixiang	Henggang	Songgang	Minzhi	Minzhi
			⋮		
51	Fuyong	Meisha	Kuichong	Yantian	Lianhua
52	Nanyuan	Dapeng	Shatoujiao	Shatoujiao	Huafu
53	Pinghu	Shekou	Dapeng	Dapeng	Sungang
54	Guangming	Shatoujiao	Meisha	Meisha	Meisha
55	Dapeng	Yantian	Nanao	Nanao	Nanao

its status as a high-tech hub (e.g., Huawei's headquarters) that fosters a mix of residential and employment functions. Buji complements this role with the highest Closeness Centrality, indicating its optimal accessibility to the entire network, which is logical given its function as a massive population hub adjacent to the city core.

In stark contrast, Gongming is the undisputed flow epicenter, ranking first in both absolute volume (Weighted Degree) and per-capita intensity (Relocation Rate). This powerful combination confirms its function as a massive "population processing hub" characterized by extreme churn. This is driven by its economic base in manufacturing, which attracts a large, transient workforce, and the availability of low-cost housing, which facilitates high turnover as residents seek marginal improvements in rent or proximity to factory jobs. This analysis, therefore, identifies distinct functional roles: Bantian and Buji act as the network's structural integrators ensuring system-wide connectivity, while Gongming serves as the high-volume, high-intensity engine of mobility. At the other extreme, subdistricts like Dapeng and Nanao consistently rank at the bottom of all centrality measures, confirming their peripheral status due to geographical isolation and a non-industrial, tourism-focused economy.

The community detection algorithm divided Shenzhen's subdistricts into several distinct communities based on the characteristics of the relocation network (Fig. 11). For instance, the eastern areas, such as Nanao and Longgang, form one community, while the western areas, including Nanshan and Xixiang, constitute another. The central urban districts, such as Futian and Buji, are divided into multiple communities, reflecting the strong internal connections in terms of population flow. Overall, the community divisions highlight the migration patterns between different regions of

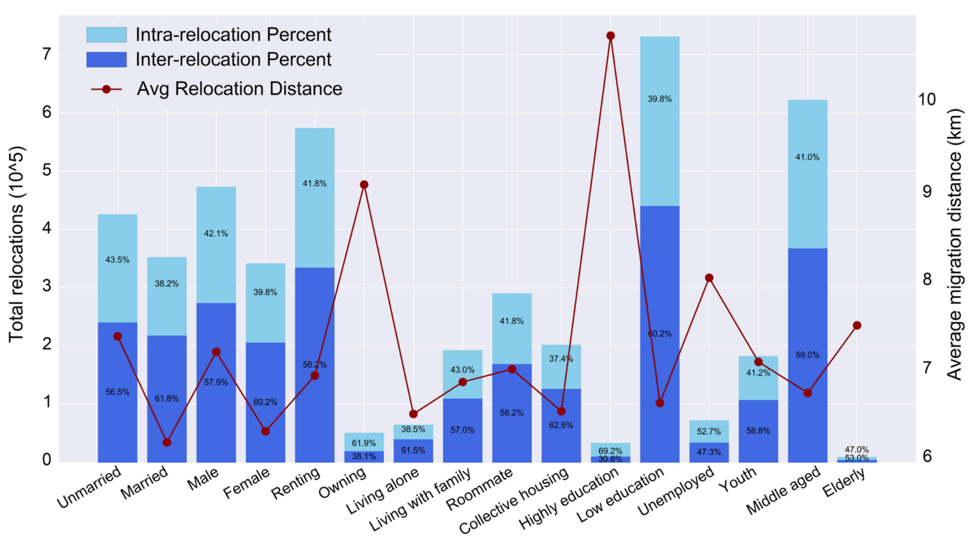


Fig. 11 Resident relocation network community detection results

Shenzhen, revealing how geographic proximity and economic ties influence population movements, transcending administrative boundaries.

5.2 Exploration of differences in relocation behaviour across groups

5.2.1 Relocation patterns and distance variations

Having established the demographic composition of our relocation sample and its inherent biases in Section 3.4, we now explore the differences in relocation behaviour among various internal groups. Table 6 provides a detailed breakdown of the demographic characteristics of the 770,701 individuals in our relocation sample, which serves as the basis for the following analysis. The data visualized in Fig. 12, reveals significant differences in relocation patterns and distances across groups. For instance, individuals with lower education levels, Younger age, and those in rental Housing constitute the majority of relocation activities. Within our sample, married people show a higher tendency to move within the subdistrict than unmarried people, accounting for 61.8% of their moves and covering shorter distances, suggesting a preference for residential stability. Women tended to move less frequently and over shorter distances than men. The higher the education level, the greater the proportion of cross-subdistrict moves and the longer the distance, indicating a pursuit of work and living conditions. Those with low educational attainment, on the other hand, preferred to relocate within the subdistrict over short distances. Young and middle-aged people are active in relocating, but at moderate distances. Notably, the small cohort of elderly individuals captured in our data showed a distinct pattern: they relocated less frequently, but a slightly higher proportion of their moves were long-distance and across subdistricts. A possible explanation, which requires further study, is that these moves are driven by the need to be close to family members or specialized medical facilities. Overall, our data suggests that factors such as marriage, housing, education and age combine to influence the relocation decisions of different groups within Shenzhen's core labor force.

5.2.2 Analysis of relocation network structure: differences in migration patterns and flow across groups

The structure of the relocation networks varies significantly across different population groups (Fig. 13). For married individuals, the relocation network is concentrated in Bao'an District, primarily following the relocation route from Gongming to Songgang, Shajing, Fuyong, Xixiang, and Xin'an. In contrast, the migration patterns of unmarried individuals are more evenly distributed among nodes such as Guanlan, Shajing, and Xixiang, with Longhua being particularly active.The relocation networks of men and women are structurally similar, but women tend to relocate more frequently from the northern parts of the city through Longhua and Minzhi toward the central districts of Shahe and Futian. Homeowners' relocation networks are simpler, with the relocation network of renters is more complex and fluid, with higher overall mobility.

Of the four groups with different living arrangements, the relocation networks of individuals living alone and living with roommates are similar, with large relocation streams spread across the periphery of the city. In

Table 6 Demographic composition of the relocation sample

Group	Description	Count	Percent %
Unmarried	-	401341	52.07
Married	-	335013	43.47
Male	-	446515	57.94
Female	-	324186	42.06
Renting a house	-	545590	70.79
Owning a house	-	48038	6.23
Living alone	-	61390	7.97
Living with family	-	183395	23.8
Living with roommates	Living with roommates	275951	35.81
Living in collective housing	Living in a collective housing	190228	24.68
Highly educated	Bachelor's degree or higher	30565	3.97
Low education	Below Bachelor's Degree	693127	89.93
Unemployed	-	68515	8.89
Youth	22 years and below	172340	22.36
Middle-aged	23 to 59 years of age	588913	76.41
Elderly	60 years and older	9448	1.23

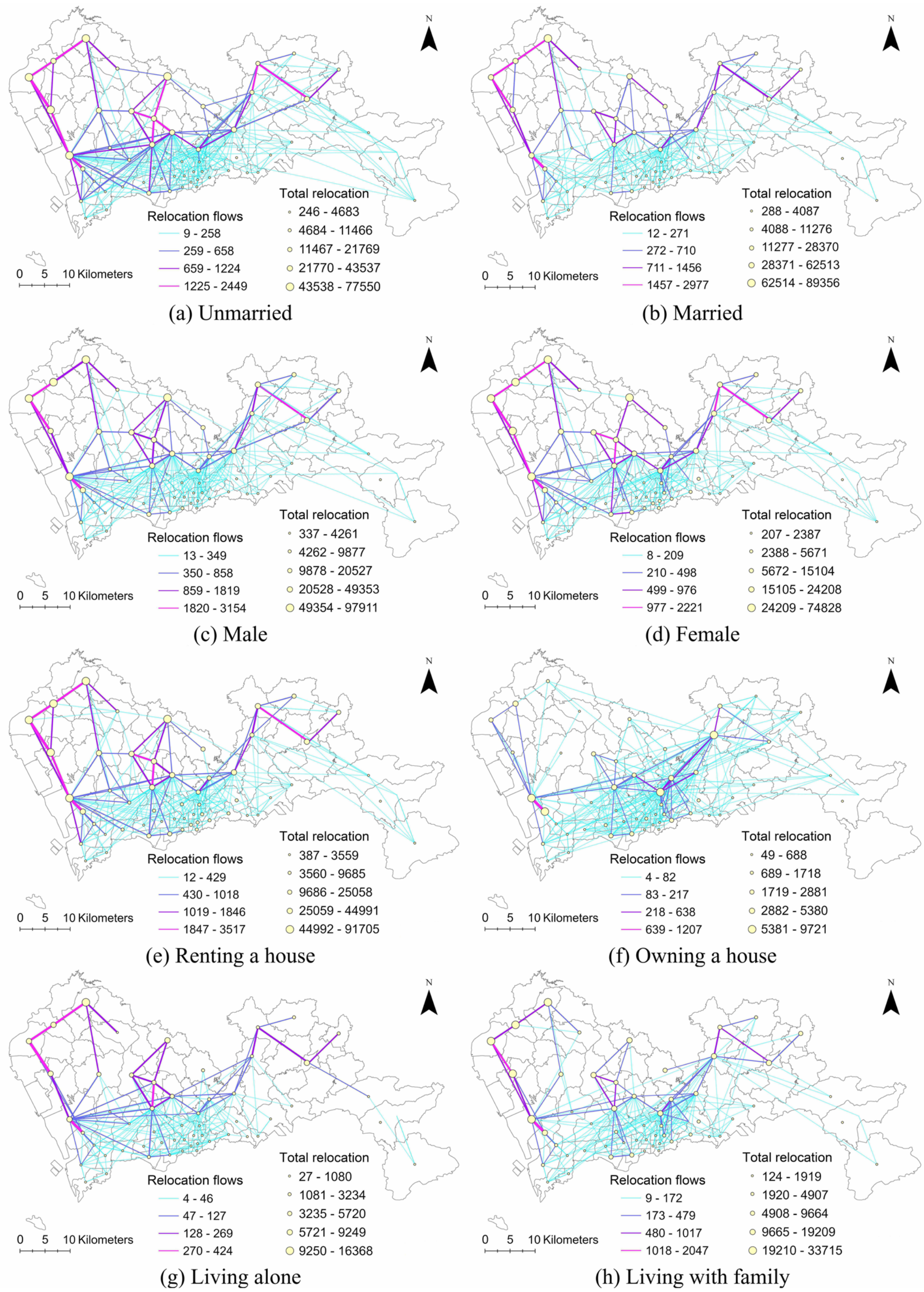


Fig. 12 Migration volume and distance across different population groups

contrast, families show a more balanced relocation flow across the network, although paths between Shajing, Fuyong, Xixiang, and Xin'an are notably more active. The relocation network for group-living individuals is far more complex, with dense connections, especially around Gongming, extending to other nodes in the western part of the city, indicating tight community dynamics.

The relocation networks of individuals with different educational backgrounds also show stark differences. Highly educated individuals are concentrated in the core urban districts, such as Futian and Nanshan, where

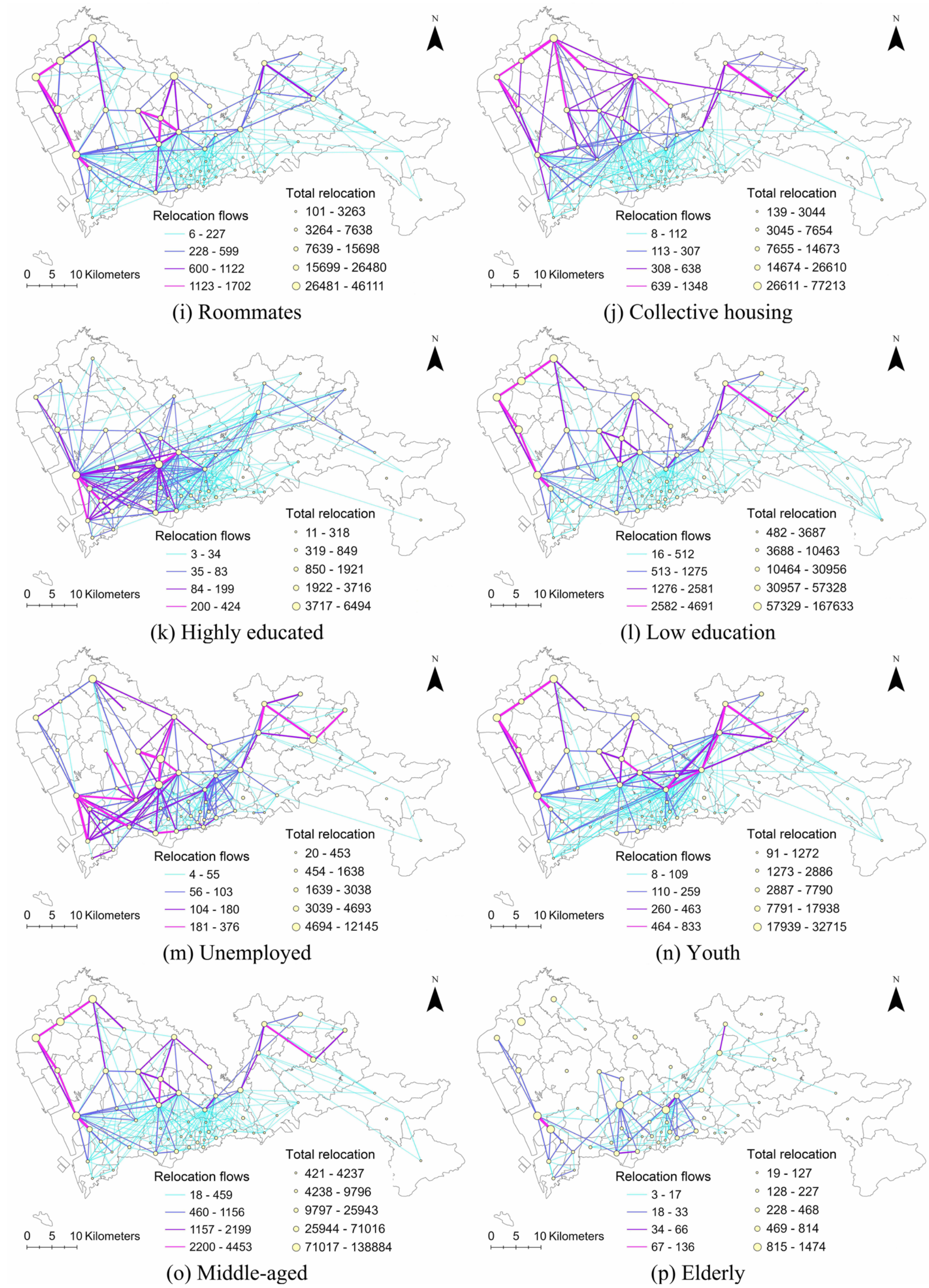


Fig. 13 Network structures for different demographic groups

relocation flow is very dense, and nodes are closely interconnected. Xixiang and Minzhi emerge as key nodes, handling a significant portion of the relocation flows. In contrast, the relocation networks of less educated individuals are more prevalent in the suburban areas of the city.

Age also plays a crucial role in shaping relocation networks. Younger individuals are primarily concentrated in the northeastern part of the city, with major migration routes including Pingshan, Longgang, Longcheng, Henggang, and Buji. The relocation network for older adults is more dispersed, with frequent relocations between

Xixiang and Xin'an, while other activities are centred in Futian and Luohu. Unemployed individuals exhibit a relatively complex network structure with no clear migration paths.

5.2.3 Relocation network metrics analysis: social structures and behavioural differences

By analysing key network metrics in the inter-subdistrict relocation networks of different groups (Table 7 and Fig. 14), insights can be gained into their social structures and relocation behaviours. For example, unmarried individuals are active in broader networks, with higher degree centrality, suggesting fewer constraints and a more exploratory approach to job and housing searches across the city. In contrast, married individuals move within tightly knit communities, shown by higher modularity, which reflects a decision-making process constrained by factors like schooling, family ties, and a stronger preference for residential stability, leading to more localized moves. Gender differences are also notable. Men connect to a wider variety of nodes, showing higher negative assortativity, indicating connections to dissimilar nodes. This may reflect broader and more diverse job-seeking patterns across different economic sectors and geographic areas.On the other hand, women exhibit higher positive assortativity, meaning they are more likely to connect to similar nodes, potentially influenced by the importance of specific social support networks, safety considerations, or industries concentrated in particular areas.

Housing type also matters. Homeowners form denser local connections, a logical outcome of the high transaction costs and long-term investment associated with homeownership, which discourages frequent, long-distance moves. In contrast, renters relocate more frequently across the city, as they can respond more flexibly to labor market changes and rental price differentials. The effect of education level is demonstrated by the fact that the relocation networks of the highly educated have lower clustering coefficients and modularity than those of the less educated, reflecting their broader social networks and fewer geographic constraints. This might because specialized, high-skilled jobs are unevenly distributed, often requiring national or city-wide searches, thus breaking out of tight, local community structures. In addition, younger people tend to be more connected to their peers, whereas older people have relatively stable community structures, although they have fewer network connections.

6 Discussion

6.1 Relocation patterns and regional differences

This study reveals that intra-urban relocations in Shenzhen are predominantly short-distance, with most moves occurring within a 10km radius and often within the same subdistrict. This finding aligns with established literature suggesting residents prefer to move within familiar environments (Li & Mao, 2019). This pattern is particularly pronounced in the city's peripheral districts—such as Bao'an, Longgang, and Longhua, which host active subdistricts like Gongming and Shajing.

Table 7 Grid community registration data dictionary

Group	Average degree	Average weighted degree	Clustering coefficient	Modularity	Assortativity
Unmarried	12.182	7088.60	0.572	0.808	−0.032
Married	10.509	6440.55	0.557	0.840	−0.041
Male	11.382	8087.05	0.580	0.819	−0.111
Female	10.836	5947.91	0.558	0.834	0.127
Renting a house	11.309	10035.87	0.561	0.820	0.067
Owning a house	12.889	836.22	0.596	0.672	0.175
Living alone	9.673	1112.00	0.504	0.828	0.146
Living with family	10.473	3454.20	0.562	0.835	0.073
Roommates	11.164	4958.69	0.608	0.814	0.154
Collective housing	11.164	3504.07	0.536	0.793	−0.063
Highly educated	13.200	500.42	0.609	0.586	0.009
Low education	11.055	12731.95	0.591	0.830	−0.136
Unemployed	10.691	1131.07	0.566	0.776	0.142
Youth	10.727	2968.31	0.575	0.815	0.189
Middle-aged	11.345	10973.53	0.586	0.825	0.036
Elderly	6.815	167.98	0.372	0.819	0.083

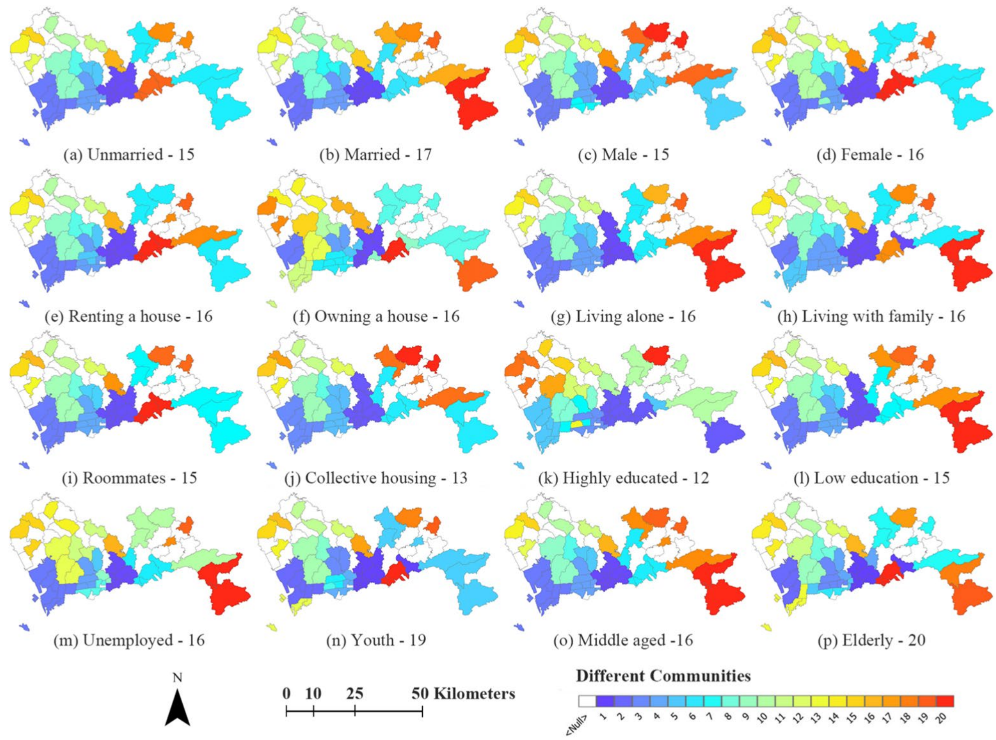


Fig. 14 Relocation communities across different demographic groups in Shenzhen. Each subplot displays the geographic distribution of migration communities detected for a specific demographic group at the subdistrict level. Subdistricts rendered in the same color belong to the same community, indicating they form a cohesive subsystem with dense internal migration flows. The title of each subplot specifies the demographic category and the total number of communities identified for that group

These areas, characterized by lower living costs and a high concentration of manufacturing jobs, logically foster localized mobility.

However, a crucial question arises from this observation: does this high volume of localized mobility reflect a stable community where residents' needs are met locally (a 'non-mover' effect), or does it signify a preference for proximity among a highly mobile population? The calculation result of migration rate points decisively to the latter. The subdistricts with the highest internal churn, Gongming and Pingshan, are also the largest population sources, exhibiting the city's highest out-relocation rates and most negative net migration rates. This reveals these peripheral districts are not stable communities but high-turnover 'population processing hubs'. Their extreme internal mobility is a symptom of instability, driven by a transient workforce frequently relocating within a precarious, low-cost housing market. While residents' lives are economically 'locked' into the area, their residency is in constant flux.

This dynamic also explains the spatial sampling bias in our data. The high data density in these areas is not a methodological error but a direct reflection of the hyper-mobility that defines them. Ultimately, the high sampling rate, extreme internal churn, and massive population outflow are interconnected facets of the same phenomenon: the socio-economic dynamism of Shenzhen's transient periphery. Conversely, the economically developed core districts (Futian, Nanshan) exhibit a higher proportion of inter-subdistrict relocations. The fact that these patterns are clearly visible despite the significant under-sampling in these core areas reinforces the significance of the core-periphery divide in shaping residents' relocation behaviour.

6.2 Key migration pathways and the hub role of core nodes

The relocation network in Shenzhen exhibits different hub nodes and major migration paths, reflecting not only the features of population mobility, but also the city's layout and economic dynamism. Our analysis identifies subdistricts such as Gongming, Shajing, and Xixiang as dominating the city's migration flows. While these hubs are located in districts with high data sampling rates, their prominence is fundamentally rooted in their economic function. These areas, particularly in the city's western and northern periphery, are clustered with manufacturing and small-to-medium enterprises, providing a vast number of employment opportunities. The observed dense and stable transport network, for example, connecting Gongming with Songgang, Shajing, and Fuyong, is therefore a direct representation of a highly active, localized labor market dynamic, primarily serving the city's core industrial workforce. Furthermore, the network reveals a clear hierarchical structure in these migration pathways. As one moves closer to the urban core, nodes like Xixiang and Xin'an serve as crucial transit points, offering access to more high-value job opportunities than the farther peripheries. This geographic and economic gradient creates a visible migration trajectory, where migrant groups are drawn to the peripheral nodes and then gradually move towards the city center.

Crucially, the role of the core districts, Futian and Nanshan, as primary attractors is powerfully confirmed, especially when considering the data's spatial bias. Despite being in under-sampled districts, they remain the unequivocal destination for the highly educated population and for a large proportion of cross-subdistrict moves. Their magnetic pull is so strong that it clearly stands out even with less data coverage, reinforcing their status as the city's economic and innovation engines. It exhibits a dual structure: localized, high-volume mobility in the manufacturing-heavy periphery and long-distance, talent-oriented flows toward the urban core. This structure illustrates the "live in the periphery, work in the centre" pattern that defines Shenzhen's multi-level urbanization. Future urban planning should therefore pay special attention to the distinct functions of these different hubs and nodes to enhance overall mobility and urban efficiency by optimizing transport and resource allocation.

6.3 Differences in relocation behaviour across groups and social significance

Profound influence of social structure is found on relocation behaviour. For example, married people are more likely to relocate within the subdistrict and have shorter average relocation distances. This reflects their need for family stability, which may be related to family responsibilities and children's educational needs. The highly educated group, on the other hand, exhibit longer relocation distances and are concentrated in economically developed core areas. This phenomenon suggests higher mobility and more opportunities for choice in their career development.

It is particularly noteworthy that the small cohort of elderly individuals in our sample exhibited a distinct relocation pattern: a higher propensity for long-distance, cross-subdistrict moves. This observation appears to diverge from some established literature that suggests seniors tend to prefer short-distance moves to improve their living environment (Li et al., 2022; Atkins, 2018; Hou et al., 2024). While the under-representation of this group in our data prevents us from making a city-wide generalization, this finding raises a compelling hypothesis for future investigation. It is possible that for Shenzhen's elderly, intra-urban relocation is not a frequent event for minor residential improvement, but rather a rare, high-impact decision, potentially driven by major life events such as retirement, proximity to family caregivers, or the need for specialized healthcare. This pattern, if validated by future targeted studies, would suggest that planners need to consider not just local amenities but also cross-regional connectivity and services when addressing the needs of an aging population in a dynamic megacity.

6.4 Urban planning and policy implications

The fine-grained analysis of Shenzhen's intra-urban migration network offers more than just a characterization of population movement, it provides a data-driven foundation for evidence-based urban planning and policy formulation, addressing critical challenges in a rapidly urbanizing megacity (Yeh & Chen, 2020). Specifically, the identification of major migration hubs, such as the Gongming and Shajing areas, alongside key "bridge" subdistricts with high betweenness centrality, presents a clear roadmap for strategic infrastructure investment. This insight allows planners to move beyond a static, city-wide allocation model and instead adopt a dynamic, targeted approach consistent with a polycentric urban development strategy (Sun & Lv, 2020; Wang, 2021). By prioritizing the allocation of high-quality public services and enhancing public transport capacity in these high-flow zones, particularly through Transit-Oriented Development (TOD) principles (Wang & Xia, 2024), these areas can be transformed into functional, self-sufficient urban sub-centers. This approach can mitigate the pressure on the traditional city core and improve the city's overall job-housing balance (Wang et al., 2019; Zhang & Luo, 2024). Furthermore, the distinct migration patterns among different demographic groups call for nuanced, people-centric policies that transcend one-size-fits-all solutions, aligning with principles of equity planning. The high rate of short-distance, internal churn observed in certain subdistricts, for instance, underscores the urgent need for robust affordable rental housing programs and improved regulation of the leasing market to enhance residential stability for the city's vital labor force (Rui et al., 2024; Tong et al., 2020). Concurrently, the long-distance mobility of the highly educated population within economic core areas validates the strategic placement of talent apartments and innovation-conducive amenities, a strategy aimed at attracting and retaining high-skilled human capital (Qiu et al., 2024). Crucially, even the inherent biases within our dataset serve as a profound policy insight: the over-representation of the working-age population in suburban, industrial districts reveals precisely where the city's governance and service systems for its vast migrant population are most stressed. This challenges the conventional planning focus on established urban centers like Futian and Nanshan, making a compelling case for a more equitable distribution of resources to burgeoning migrant-heavy areas like Guangming and Longgang. Such a shift is essential for fostering inclusive development and ensuring that the benefits of urbanization are shared more broadly across the entire municipal landscape.

6.5 Limitations and perspectives

This study has several Limitations that warrant discussion and offer avenues for future research. The most significant is the sampling bias detailed in Section 2.4. Our dataset is skewed towards the working-age (15-64) and male population, so our findings characterize the migration of the city's core workforce rather than the entire population. And spatially the dataset is biased towards suburban industrial areas, while the core urban area is undersampled. While we have accounted for spatial biases in our discussion, migration volumes from peripheral areas may be inflated. To address these demographic and spatial biases, future research could employ statistical correction methods. For instance, post-stratification weighting, where weights are assigned to individual records to align the sample's distributions with official statistics, could be used to construct a more representative city-wide migration model.

Secondly, the cross-sectional design of this study and its definition of migration resulted in an inability to capture the long-term dynamics of the migration network and may have underestimated the true total number of migrations. While the 2015 dataset provides an exceptionally robust snapshot, data from prior years were collected under a different and less comprehensive framework, merging different datasets could introduce significant bias. Consequently, to ensure the internal reliability of the network, we adopted a conservative definition requiring a move's origin and destination to be registered within 2015. This choice, while ensuring data integrity, likely results in a conservative estimate of the total relocation volume and means the long-term evolution of the network cannot be captured. Longitudinal analysis based on consistently collected multi-year data is a crucial next step.

Finally, this study analysed relocation patterns mainly based on administrative divisions, which may lead to biased interpretations of relocation due to the Modifiable Areal Unit Problem (MAUP). Future research could incorporate finer spatial analyses to explore the relationship between relocation behaviour and geographic features. Finally, this study used the Disparity Filter algorithm to simplify the network, which may have overlooked some minor but potentially significant relocation paths. Future work could also explore machine learning models to build predictive models for migration trends, which can help urban planners to respond to possible population movement peaks and bottlenecks in advance and optimise resource allocation.

7 Conclusion

Shenzhen's migration patterns reveal a multi-layered and regionally differentiated system, where core districts exhibit more outward migration and peripheral districts maintain internal stability. Key migration hubs connect different regions, while demographic factors such as marital status, education level, and gender further shape individual migration behaviours. Shenzhen's unique migration characteristics and development trajectory have made it a one-of-a-kind city in China, and a representative of rapidly urbanizing cities worldwide, offering significant reference value. Studying its internal urban migration and community dynamics not only aids in understanding the spatial patterns of this exemplary migrant city, but also serves as a model for migration research, providing opportunities to validate or construct migration network theories. This exploration offers theoretical support for optimizing urban planning and policies, while providing scientific foundations for improving resource allocation efficiency.

Acknowledgements
Not applicable.

Authors' contributions
Zhongyu Lai: writing-original draft, methodology, writing-review & editing, formal analysis, data curation, conceptualization. Yueshan Li: writing-original draft, validation, investigation, data curation. Tao He: writing-review & editing, resources, methodology, Conceptualization. Xintao Liu: writing-review & editing, validation, supervision, funding acquisition, conceptualization.

Funding
The work is financially supported by a project (CDL1) from Research Institute for Land and Space (RILS), The Hong Kong Polytechnic University.

Data availability
The raw data supporting the conclusions of this article are not available for public release due to ethical restrictions and privacy concerns, as the dataset contains sensitive, personally identifiable information, including detailed residential addresses. The data processing and analysis methods are described in detail within the manuscript. Further inquiries regarding the data may be directed to the corresponding author.

Declarations

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

Received: 8 May 2025 Revised: 2 September 2025 Accepted: 3 September 2025
Published online: 18 September 2025

References
Acharya, C. P., & Leon-Gonzalez, R. (2014). How do migration and remittances affect human capital investment? The effects of relaxing information and liquidity constraints. *Journal of Development Studies, 50*(3), 444–460.
Atkins, M. T. (2018). "On the move, or staying put?" an analysis of intrametropolitan residential mobility and ageing in place. *Population, Space and Place, 24*(3), e2096.
Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: Theory and experiment, 2008*(10), P10008.
Bon, N. G., Josipovič, D., & Kanjir, U. (2018). Linking geomorphological and demographic movements: The case of Southern Albania. *Applied Geography, 100*, 55–67.
Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology, 25*(2), 163–177.
Brown, L. A., & Holmes, J. (1971). The delimitation of functional regions, nodal regions, and hiberarchies by functional distance approaches. *Journal of regional science, 11*(1), 57–72.
Cai, B., Zeng, L., Wang, Y., Li, H., & Hu, Y. (2019). Community detection method based on node density, degree centrality, and k-means clustering in complex network. *Entropy, 21*(12), 1145.
Charyyev, B., & Gunes, M. H. (2019). Complex network of United States migration. *Computational Social Networks, 6*, 1–28.
Cheng, J., Chen, M., & Tang, S. (2023). Shenzhen-a typical benchmark of Chinese rapid urbanization miracle. *Cities, 140*, 104421.
Cui, C., Geertman, S., & Hooimeijer, P. (2014). The intra-urban distribution of skilled migrants: Case studies of Shanghai and Nanjing. *Habitat International, 44*, 1–10.
Danchev, V., & Porter, M. A. (2021). Migration networks: applications of network analysis to macroscale migration patterns. In *Research handbook on international migration and digital technology* (pp. 70–90). Edward Elgar Publishing.
De Haas, H. (2021). A theory of migration: The aspirations-capabilities framework. *Comparative Migration Studies, 9*(1), Article 8.
Di Maria, C., & Lazarova, E. A. (2012). Migration, human capital formation, and growth: An empirical investigation. *World Development, 40*(5), 938–955.
Fan, X., & Xiang, H. (2020). The migration destination of rural labor and nearby urbanization of small cities: A case study of Henan province, China. *Papers in Applied Geography, 6*(2), 85–100.
Fang, B., Tu, W., Li, M., Cao, J., Gao, W., Yue, Y., & Li, Q. (2022). Uncovering spatial synergy of the megacity region: A flow perspective. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43*, 521–528.
General Office of the CPC Shenzhen Municipal Committee and General Office of Shenzhen Municipal People's Government. (2014). *Notice on the issuance of the implementation plan for comprehensively promoting the social construction "network project"*. Retrieved from https://www.sz.gov.cn/zfgb/2014/gb870/content/post_4993071.html.
Gong, Y., Cao, Z., & Tong, D. (2024). Social ties and talent migration: Considering the intentions of migrants to permanently settle in Chinese cities. *Applied Geography, 165*, 103227.
Goodburn, C. (2020). Growing up in (and out of) Shenzhen: The longer-term impacts of rural-urban migration on education and labor market entry. *The China Journal, 83*(1), 129–147.
Guangdong Provincial People's Government. (2020). *Notice on several policy measures to further stabilize and promote employment in Guangdong province*. Retrieved from https://www.sz.gov.cn/zfgb/2014/gb870/content/post_4993071.html.
He, Q., & Zhang, Y. (2023). Residential locations and residential moves between the city centre and suburb in Beijing, China. *Habitat International, 131*, 102711.
Hou, Z., Li, X., & Li, X. (2024). Determinants of intra-city residential migration patterns of older adults: A GIS and decision tree analysis of Yancheng city, China. *ISPRS International Journal of Geo-Information, 13*(9), 324.
Huang, X., Van Weesep, J., & Tang, S. (2021). To move or not to move? Residential mobility of rural migrants in a medium-sized Chinese city: The case of Yangzhou. *Housing Studies, 36*(2), 278–301.
Lai, Y., Chen, K., Zhang, J., & Liu, F. (2020). Transformation of industrial land in urban renewal in Shenzhen. *China. Land, 9*(10), 371.
Li, T., & Dodson, J. (2023). Worker intra-urban residential migration and spatial labour market change in Melbourne's functional economic regions. *Population, Space and Place, 29*(3), Article e2618.
Li, S., Hu, W., & Guo, F. (2022). Recent relocation patterns among older adults in the United States: Who, why, and where. *Journal of the American Planning Association, 88*(1), 15–29.
Li, S.-M., & Mao, S. (2019). The spatial pattern of residential mobility in Guangzhou, China. *International Journal of Urban and Regional Research, 43*(5), 963–982.
Lin, J., Wu, K., Yang, S., & Liu, Q. (2021). The asymmetric pattern of population mobility during the spring festival in the Yangtze river delta based on complex network analysis: an empirical analysis of "tencent migration" big data. *ISPRS International Journal of Geo-Information, 10*(9), 582.
Liu, R., Li, T., & Greene, R. (2020). Migration and inequality in rental housing: Affordability stress in the Chinese cities. *Applied Geography, 115*, 102138.
Liu, Y., Yu, S., & Sun, T. (2021). Heterogeneous housing choice and residential mobility under housing reform in China: Evidence from Tianjin. *Applied Geography, 129*, 102417.
Michael Keith, Scott Lash, J. A., & Rooker, T. (2013). *China Constructing Capitalism: Economic Life and Urban Change*. Taylor & Francis Group.
Noldus, R., & Van Mieghem, P. (2015). Assortativity in complex networks. *Journal of Complex Networks, 3*(4), 507–542.
Nora, C. L. H. (2023). Qualitative and quantitative methods as applied to international migration. In *Methodological approaches in integrated geography* (pp. 99–114). Springer.
Okamoto, K., Chen, W., & Li, X.-Y. (2008). Ranking of closeness centrality for large-scale social networks. In *International workshop on frontiers in algorithmics* (pp. 186–195). Springer.
Pei, T., Yan, X., Shu, H., Wang, X., Chen, X., Guo, S., Jiang, J., Song, C., & Chen, J. (2024). Geographical flows: A fresh perspective on quantifying urban function. *Urban Informatics, 3*(1), 1–9.
Pitoski, D., Lampoltshammer, T. J., & Parycek, P. (2021). Network analysis of internal migration in Austria. *Digital Government: Research And Practice, 2*(3), 1–24.
Poisot, T. (2013). An a posteriori measure of network modularity. *F1000 Research, 2*, 130.
Qi, W., Abel, G. J., & Liu, S. (2021). Geographic transformation of China's internal population migration from 1995 to 2015: Insights from the migration centerline. *Applied Geography, 135*, Article 102564.
Qiu, J., Lyu, P., & Tian, M. (2024). Do talent housing policies foster regional innovation? An analysis based on labor force heterogeneity. *Economic Analysis and Policy, 83*, 150–164.
Rogers, A. (2008). Demographic modeling of the geography of migration and population: A multiregional perspective. *Geographical Analysis, 40*(3), 276–296.
Rui, J., Xu, Y., & Li, X. (2024). Destigmatizing urban villages by examining their attractiveness: Quantification evidence from Shenzhen. *Habitat International, 150*, 103120.
Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., & Kertesz, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E, 75*(2), Article 027105.
Schon, J., & Johnson, J. C. (2021). How inter-state amity and animosity complement migration networks to drive refugee flows: A multi-layer network analysis, 1991–2016. *PLoS One, 16*(1), e0245712.
Serrano, M. Á., Boguná, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences, 106*(16), 6483–6488.

Shen, Y.-S., Lung, S.-C.C., & Cui, S. (2022). Exploring multiple pathways and mediation effects of urban environmental factors for suicide prevention. *Environmental Pollution, 294*, Article 118642.
Shi, G., Lu, X., Deng, Y., Urpelainen, J., Liu, L.-C., Zhang, Z., Wei, W., & Wang, H. (2020). Air pollutant emissions induced by population migration in China. *Environmental Science & Technology, 54*(10), 6308–6318.
Slavko, B., Glavatskiy, K., & Prokopenko, M. (2020). City structure shapes directional resettlement flows in Australia. *Scientific Reports, 10*(1), 8235.
Stamos, I. (2023). Transportation networks in the face of climate change adaptation: A review of centrality measures. *Future Transportation, 3*(3), 878–900.
Sun, T., & Lv, Y. (2020). Employment centers and polycentric spatial development in Chinese cities: A multi-scale analysis. *Cities, 99*, 102617.
Sun, S., & Manson, S. M. (2012). Intraurban migration, neighborhoods, and city structure. *Urban Geography, 33*(7), 1008–1029.
Tong, D., Zhang, Y., MacLachlan, I., & Li, G. (2020). Migrant housing choices from a social capital perspective: The case of Shenzhen. *China. Habitat International, 96*, 102082.
Treyz, G. I., Rickman, D. S., Hunt, G. L., & Greenwood, M. J. (1993). The Dynamics of U.S. Internal Migration. *The Review of Economics and Statistics, 75*(2), 209–214.
Wang, M. (2021). Polycentric urban development and urban amenities: Evidence from Chinese cities. *Environment and Planning B: Urban Analytics and City Science, 48*(3), 400–416.
Wang, G. Y. (2022). Talent migration in knowledge economy: The case of China's silicon valley, Shenzhen. *Journal of International Migration and Integration, 23*(3), 1175–1196.
Wang, M., Derudder, B., & Liu, X. (2019). Polycentric urban development and economic productivity in China: A multiscalar analysis. *Environment and Planning A: Economy and Space, 51*(8), 1622–1643.
Wang, Y. P., Wang, Y., & Wu, J. (2009). Urbanization and informal development in China: Urban villages in Shenzhen. *International Journal of Urban and Regional Research, 33*(4), 957–973.
Wang, L., & Xia, H. (2024). A comprehensive review of the development characteristics and future trends of TOD in Chinese urban rail transit. *Urban Rail Transit, 10*(4), 335–349.
Wu, W. (2006). Migrant intra-urban residential mobility in urban China. *Housing Studies, 21*(5), 745–765.
Wu, W., & Wang, G. (2021). Shifting residential and employment geography: Shanghai's bifurcated trajectory of spatial restructuring. *Cities, 113*, Article 103142.
Xu, F., Wang, Z., Chi, G., Wang, D., Zhang, Z., & Zuo, D. (2021). Differentiation and progress of urban regionalization in China: Perspectives of land use and geography. *Applied Geography, 137*, 102600.
Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports, 6*(1), 30750.
Ye, J., Cai, L., Shi, X., & Cheng, M. (2024). Pursuing a brighter future: Impact of the hukou reform on human capital investment in migrant children in China. *China Economic Review, 85*, 102160.
Yeh, A.G.-O., & Chen, Z. (2020). From cities to super mega city regions in China in a new wave of urbanisation and economic transition: Issues and challenges. *Urban Studies, 57*(3), 636–654.
You, S., Feng, Z., You, Z., Shi, H., & Zhao, G. (2023). Identification and structural characteristics of urban agglomerations in China based on Baidu migration data. *Applied Geography, 156*, 102999.
Zhang, W., Chong, Z., Li, X., & Nie, G. (2020). Spatial patterns and determinant factors of population flow networks in China: Analysis on tencent location big data. *Cities, 99*, 102640.
Zhang, J., Fei, J., Song, X., & Feng, J. (2021). An improved louvain algorithm for community detection. *Mathematical Problems in Engineering, 2021*(1), 1485592.
Zhang, Y., & Luo, F. (2024). Exploring the relationship between jobs and housing in China's megacities: the case of Guangzhou and Shenzhen. *Journal of Asian Architecture and Building Engineering, 24*(4), 3312–3328.
Zhang, J., Senousi, A. M., Zhao, P., Law, S., & Liu, X. (2023). Exploring 3D spatial morphology using multilayered space syntax, network science and wi-fi log data. *Urban Informatics, 2*(1), 1.
Zhao, Z., Jin, M., Jin, J., Liu, L., Gong, Y., & Liu, Y. (2024). Discovering the spatial heterogeneous constraints of distance on migration from counties to Shenzhen in China. *Applied Geography, 171*, 103384.
Zhou, J., & Hui, E.C.-M. (2022). Housing prices, migration, and self-selection of migrants in China. *Habitat International, 119*, 102479.

Publisher's Note

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