Cities 132 (2023) 103975

Contents lists available at ScienceDirect

Cities

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

Influential factors of intercity patient mobility and its network structure in China

Jiaqi Ding [a,b,1], Chao Yang [c,d,e,1], Yueyao Wang [a,b], Pengfei Li [e], Fulin Wang [f,g,h], Yuhao Kang [i], Haoyang Wang [j], Ze Liang [a,b], Jiawei Zhang [k], Peien Han [k], Zheng Wang [a,b], Erxuan Chu [l], Shuangcheng Li [a,b,*], Luxia Zhang [c,d,e,g,**]

[a] College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
[b] Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing 100871, China
[c] Renal Division, Department of Medicine, Peking University First Hospital, Peking University Institute of Nephrology, Beijing 100034, China
[d] Research Units of Diagnosis and Treatment of Immune-mediated Kidney Diseases, Chinese Academy of Medical Sciences, Beijing 100034, China
[e] Advanced Institute of Information Technology, Peking University, Hangzhou 311215, China
[f] Institute of Medical Technology, Peking University Health Science Center, Beijing 100191, China
[g] National Institute of Health Data Science at Peking University, Beijing 100191, China
[h] Peking University First Hospital, Beijing 100034, China
[i] GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI 53705, United States
[j] School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia
[k] Department of Health Policy and Management, School of Public Health, Peking University, Beijing 100191, China
[l] Faculty of Geographical Science, School of Geography, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Keywords:
Human mobility
Patient mobility
Public health
Health geography
Complex networks
Community roles

ABSTRACT

Intercity patient mobility reflects the geographic mismatch between healthcare resources and the population, and has rarely been studied with big data at large spatial scales. In this paper, we investigated the patterns of intercity patient mobility and factors influencing this behavior based on >4 million hospitalization records of patients with chronic kidney disease in China. To provide practical policy recommendations, a role identification framework informed by complex network theory was proposed considering the strength and distribution of connections of cities in mobility networks. Such a mobility network features multiscale community structure with “universal administrative constraints and a few boundary breaches”. We discovered that cross-module visits which accounted for only 20 % of total visits, accounted for >50 % of the total travel distance. The explainable machine learning modeling results revealed that distance has a power-law-like effect on flow volume, and high-quality healthcare resources are an important driving factor. This paper provides not only a methodological reference for patient mobility studies but also valuable insights into public health policies.

1. Introduction

Accessibility to healthcare services is a key factor affecting human health. Currently, the uneven distribution of medical resources is still a common phenomenon worldwide (Barber et al., 2017; Weiss et al., 2018). This imbalance in resource allocation, especially for high-quality medical care, has greatly contributed to the intercity mobility of patients.

Human mobility has been extensively studied owing to its importance in real-world applications such as urban planning (Hillier et al., 2010; Kitamura et al., 2000), epidemic modeling (Jia et al., 2020a; Tizzoni et al., 2014), and traffic forecasting (Jiang et al., 2009; Wang et al., 2012). Researchers have made sustained progress in this field with an abundance of mobility data sources such as banknotes (Brockmann et al., 2006), mobile phones (Song et al., 2010; Gonzalez et al., 2008), GPS devices (Bazzani et al., 2010; Liang et al., 2012; Riccardo et al., 2012), and location-based social network (LBSN) services (Noulas et al., 2012). Patient mobility, as an important subset of human mobility,

* Correspondence to: S. Li, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China.
** Correspondence to: L. Zhang, National Institute of Health Data Science at Peking University, Beijing 100191, China.
E-mail addresses: dingjiaqi@stu.pku.edu.cn (J. Ding), scli@urban.pku.edu.cn (S. Li), zhanglx@bjmu.edu.cn (L. Zhang).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.cities.2022.103975
Received 16 April 2022; Received in revised form 20 August 2022; Accepted 2 September 2022
Available online 13 September 2022

reflects the level of accessibility and equity of healthcare resources. In the field of urban studies, several scholars have examined patient mobility patterns within the city using navigation or traffic trajectory data (Gong et al., 2021; Wang et al., 2020a; Xing & Ng, 2022).

However, in addition to intracity patient mobility, there are massive intercity patient flows between cities. Intercity patient flows, either within (Aggarwal et al., 2018; Beal et al., 2019; Borno et al., 2018; Diaz et al., 2019; Diaz et al., 2020; Jia et al., 2015; Jia et al., 2020b; Kornelsen et al., 2021; Pekala et al., 2021; Sharma et al., 2016; Xu et al., 2020a) or across (Andritsos & Tang, 2014; Azzopardi-Muscat et al., 2018; Baeten, 2014; Chikanda & Crush, 2019; Glinos et al., 2010a; Glinos et al., 2010b; Hanefeld et al., 2015; Laugesen & Vargas-Bustamante, 2010; Legido-Quigley et al., 2007; Lunt & Mannion, 2014) countries, have recently drawn the attention of public health scholars. Existing studies have investigated mainly the burden and outcome of patient travel (Beal et al., 2019; Borno et al., 2018; Chikanda & Crush, 2019; Diaz et al., 2019; Diaz et al., 2020; Kornelsen et al., 2021; Pekala et al., 2021; Sharma et al., 2016; Xu et al., 2020a), motivations for such behavior (Aggarwal et al., 2018; Glinos et al., 2010a; Hanefeld et al., 2015; Laugesen & Vargas-Bustamante, 2010; Lunt & Mannion, 2014) and relevant policy issues (Andritsos & Tang, 2014; Azzopardi-Muscat et al., 2018; Baeten, 2014; Glinos et al., 2010b; Jia et al., 2015; Jia et al., 2020b; Legido-Quigley et al., 2007). Recently, a few studies have employed patient mobility big data to analyze intercity patient flows (Wang et al., 2020b; Wang et al., 2021a; Wang et al., 2021b; Wang & Wang, 2021). In China, while scholars have paid attention to intercity mobility networks, there is still a lack of nationwide research supported by real-world data (Fu et al., 2021; Yan et al., 2022).

According to government statistics, China has a large number of patients who seek healthcare across cities. From 2014 to 2017, the percentage of nonlocal patients in tertiary hospitals nationwide was approximately 7.9 %, with the highest percentage in the surgery department reaching 24.2% (National Health Commission of the People's Republic of China, 2019). The intercity health-seeking behavior of Chinese patients is influenced by both push and pull factors. On the one hand, the lack of a rigid referral system has allowed patients to choose hospitals freely. On the other hand, the procedures for nonlocal medical insurance reimbursement are more complicated, and the reimbursement rate is also usually lower (Yan et al., 2022).

Improving the accessibility and quality of healthcare is a top priority in many countries (Barber et al., 2017). Over the past decade, China has been undergoing a series of healthcare reforms to achieve universal health coverage (Tao et al., 2020). To achieve "healthy" patient mobility and reduce unnecessary and long-distance patient mobility, the Chinese government has applied a series of new policy practices, such as establishing healthcare alliances (HCAs) and regional medical centers, providing counterpart support of high-quality hospitals, developing telemedicine, and eliminating barriers to reimbursement of intercity medical expenditures in pilot areas (Yan et al., 2022; National Health Commission of the People's Republic of China, 2019; Tao et al., 2020).

Most existing papers on intercity mobility from the perspective of urban research focus on the whole population (Hu et al., 2020; Pan & Lai, 2019; Zhang et al., 2020a). Less attention has been paid to patient mobility, but related studies also have significant theoretical and practical implications. On the one hand, public health policymakers need to have a quantitative understanding of the current state of patient mobility, such as complex network structure indicators like key nodes, key edges and city clusters. These findings will help complement the series of ongoing healthcare reforms mentioned above. For instance, the identification of supply and demand hubs can inform the development of regional medical centers. For researchers, it is also an interesting topic whether there are some universal patterns of patient mobility like the results revealed by human mobility studies, such as the scaling laws and geographic heterogeneity. On the other hand, examining the relationship between patient mobility and cities' various characteristics is also an important research topic. This knowledge will not only improve our understanding of how human behavior is influenced by city features but also help us understand and predict the impact of macro planning policy. The results have implications for optimizing the distribution of various resources, and will ultimately contribute to the goal of reducing uneven development and improving public service inequality.

Electronic medical record data (Huang et al., 2019; Xiong & Luo, 2020; Yang et al., 2020) contain patient location information and offer high-volume, high-accuracy and rich information on patient attributes. This data source has rarely been used in urban studies. With improved data availability, it can be expected that its academic and practical values will be fully explored, providing new opportunities for scholars to quantitatively study patient mobility at large geographic scales.

In this paper, we conducted a series of quantitative analyses of patient mobility data and discussed the implications of these results for ongoing public health policies and subsequent research. We first explored the mobility patterns of millions of patients by constructing an intercity patient mobility network (IPMN) based on a national database and compared the results with those obtained from human mobility studies. Next, we examined the basic properties and multiscale community structures of the constructed IPMN. We then designed a framework to identify cities' roles in the IPMN so that we could provide different policy advice. Finally, we utilized traditional spatial interaction and machine learning methods to model patient mobility and identify influential factors.

2. Data

2.1. Mobility data

Patient mobility data were extracted from the medical record front page (MRFP) data of the hospital quality monitoring system (HQMS), which records patients' geographic information, including their current residential addresses.

The HQMS is a mandatory patient-level national database for hospital accreditation by the National Health Commission (NHC) of the People's Republic of China. As of December 2018, HQMS covered >75 % of tertiary hospitals in 31 provinces, autonomous regions and municipalities directly under the central government (excluding Hong Kong, Macao and Taiwan). Tertiary hospitals refer to those hospitals that have at least 500 beds and are ranked as top-level hospitals in the Chinese medical system.

The MRFP provides a standardized discharge summary that contains 346 patient-level variables, including demographic characteristics, discharge diagnoses, procedures, and medical expenses. As part of stringent standard practice in China, MRFP has legal validity and must be filed by doctors who have the most accurate and comprehensive understanding of the patient's medical condition. Records of patients with chronic kidney disease (CKD, as both the primary diagnosis and the first two secondary diagnoses based on ICD-10 coding (Zhang et al., 2020b)) were extracted from the HQMS. Data from other diseases were not used due to limited data availability. However, CKD is one of the major emerging chronic diseases that have received much attention in recent years, and its disease burden and consumption of healthcare resources are very high compared to those of other chronic diseases (Yang et al., 2020; Yang et al., 2022). Second, healthcare resources for CKD are relatively scarce and unevenly distributed; thus, patients are more likely to seek nonlocal healthcare services. This study was approved by the Ethics Committee of Peking University First Hospital (2020-018). Acquisition of informed consent was exempted. More detailed descriptions of these data sources can be found in existing literature (Huang et al., 2019; Yang et al., 2020).

2.2. Urban socioeconomic data

To model intercity mobility and identify important factors, we collected multisource data to characterize cities' socioeconomic conditions, medical resources and transportation capability, creating 45 metrics. Detailed metric names, meanings and data sources are listed in Supplementary Table 1. Most statistical metrics were obtained from the *China City Statistical Yearbook 2015* (National Bureau of Statistics of China, 2015) (a city in this study refers to the administration area of a municipality).

The population data were extracted from *LandScan*, a widely used resident population dataset combining the fine-resolution demographic and geographic data and state-of-the-art remote sensing techniques (Bright et al., 2016).

Intercity flight and train information was obtained from the official platform of the Railway Customer Service Center of China and the Civil Aviation Administration of China. The shortest traffic time between cities was calculated as edge weights to construct networks with different transportation modes. A series of network-based metrics, such as degree centrality and closeness centrality, were then inferred (Freeman, 1978).

To measure cities' ability to provide high-quality healthcare services, we obtained the total number of tertiary grade A hospitals and high-level nephrologist doctors. Tertiary grade A hospitals refer to the best level of Chinese tertiary hospitals. To calculate the total number of tertiary grade A hospitals, we first retrieved point-of-interest (POI) data for the type "tertiary grade A hospital" from the web map service provider *Amap* (lbs.amap.com) and then aggregated the data based on geographical distance to eliminate duplicate records, i.e., each POI represents a unique hospital.

We consider doctors who registered on online paid medical consultation platforms as proxies for high-level and well-recognized doctors. Our data come from two leading companies, *Haodaifu* ("Good Doctor" in Chinese, www.haodf.com) and *WeDoctor* (www.guahao.com). Amid the booming growth of Chinese online healthcare platforms, the two leading companies have already served tens of millions of patients, thus indicating the representativeness of the dataset (Li et al., 2019; Liu, 2021; Yang et al., 2019).

3. Methods

The framework of this study is shown in the Fig. 1.

3.1. Community detection of mobility network

Modules (or communities, clusters) in complex networks refer to a group of nodes (in this study, cities) that have strong connections. In this paper, we first adopted community detection algorithms to infer community structure, as it is one of the most important properties of complex networks. Two scale-fixed algorithms, modularity optimization and Infomap, were employed to detect how vertices were organized into modules. To obtain a finer community partition, we applied another resolution-limit-free method (constant Potts model optimization, CPM) (Fortunato & Barthelemy, 2007). Modules of different sizes can be obtained by adjusting the resolution parameter of this method.

3.1.1. Infomap algorithm

Infomap is a dynamics-based approach that turns community detection into a coding problem (Rosvall et al., 2010; Rosvall & Bergstrom, 2008). When performing an infinite random walk on the network,

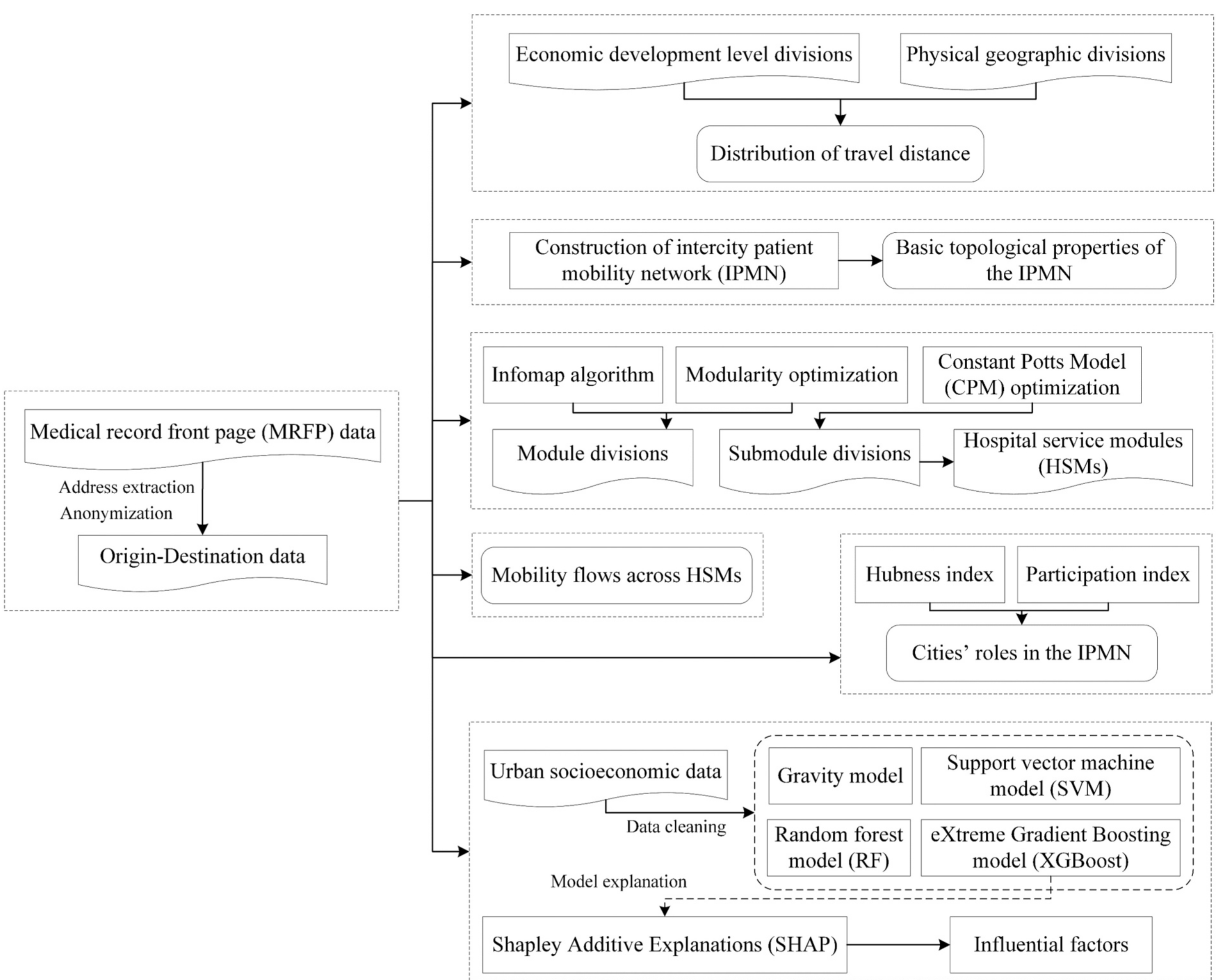

Fig. 1. Study framework.

the paths can be efficiently described by Huffman coding. The average description length of a single step is given by an information-theoretic map equation. The goal of the Infomap method is to find a partition **M** that minimizes the expected description coding length $L(\mathbf{M})$. That is,

$$L(\mathbf{M}) = q_{\curvearrowright} H(\mathcal{L}) + \sum_{i=1}^{m} p_{\circlearrowright}^{i} H\left(\mathcal{P}^{i}\right) \tag{1}$$

This equation consists of two terms, expressing the Shannon entropy of the movement within and between communities weighted by their frequency of occurrence. Here, $q_{\curvearrowright}$ is the probability that the random walk enters communities at any given step. $H(\mathscr{L})$ is the entropy of the codes of communities. The weight $p_{\circlearrowright}^{i}$ is the fraction of inside-module movements in module i plus the probability of exiting module i. $H(\mathscr{P}^i)$ is the entropy of the inside-module movements, including the exit code for module i.

3.1.2. Modularity optimization

Modularity is a structure-based quality function that measures the goodness of a partition (Newman, 2006). A high modularity score Q indicates that the nodes are densely connected internally within each community but sparsely connected across different communities. For directed networks, it is defined as (Leicht & Newman, 2008):

$$Q = \frac{1}{m} \sum_{i,j} \left(A_{i,j} - \frac{k_i^{out} k_j^{in}}{m} \right) \delta(i,j), \tag{2}$$

where m is the total edge weight, $A_{i,\,j}$ represents the weight of the edge from i to j, k_i^{out} and k_j^{in} refer to the out-degree of i and in-degree of j, respectively, and $\delta(i,j)$ equals 1 if i and j belong to the same community and 0 otherwise.

One of the most commonly used methods to optimize modularity is the Louvain algorithm introduced by Blondel (Blondel et al., 2008). It is a heuristic method that enables rapid unfolding of communities with good partition quality. However, the Louvain algorithm may yield arbitrarily badly connected communities (Traag et al., 2019). To address this problem, Traag (Traag et al., 2019) recently proposed the Leiden algorithm, which improves on the Louvain algorithm and guarantees that communities are well connected. In the present study, we used the Leiden algorithm to optimize modularity and the CPM, as discussed below.

3.1.3. Constant Potts model optimization

Community detection methods may suffer from the so-called resolution limit that prevents the identification of communities smaller than a certain scale (Fortunato & Barthelemy, 2007). Modularity depends on the total size of the network and on the degree of interconnectedness of the modules (Fortunato & Barthelemy, 2007; Good et al., 2010). For the map equation in the Infomap method, the resolution limit is set by the total number of links between modules (Kawamoto & Rosvall, 2015). In practice, it is suggested that the resolution limit of the map equation is orders of magnitude smaller than the modularity, which makes it less restrictive (Kawamoto & Rosvall, 2015).

CPM is an alternative to modularity that was proven to be resolution-limit-free (Traag et al., 2011). This quality function employs a resolution parameter that can be used to explore the community structure at various scales. A higher resolution leads to more communities and thus provides finer information on the community structure of networks. It is formulated as a sum over communities:

$$Q = \sum_{c} \left[m_c - \gamma \binom{n_c}{2} \right], \tag{3}$$

where m_c is the number of internal edges of community c, n_c is the number of nodes in community c, and γ is the resolution parameter.

3.2. Identification of the functional roles of cities in mobility network

Identifying the roles of nodes is vital to representing complex networks and understanding their structure at the mesoscale. Guimerà and Nunes Amaral (Guimerà & Nunes Amaral, 2005) proposed a framework to divide nodes into different roles according to their intra- and inter-community connection patterns. This classic framework is based on two parameters: an indicator of within-module hubness and an index describing the distribution of cross-module links of a node. Despite its wide application, some scholars have pointed out its limitations and developed new parameters accordingly (Dugué et al., 2015; Klimm et al., 2014; Pedersen et al., 2020; Xu et al., 2020b).

In this paper, we introduce a framework to assess directed networks like the intercity patient mobility network (IPMN). Our framework is based on two kinds of parameters: hubness indices measure how "well-connected" a node is to other nodes inside or outside its module, and the participation index measures how "well-distributed" the links of a node are to different destinations (Guimerà & Nunes Amaral, 2005; Xu et al., 2020b). We examine cities' roles at both the regional and national scales.

3.2.1. Hubness index

We use two hubness indices Z_i and B_i to measure the strength of a node's connections to other nodes inside and outside its own module, respectively.

The inside-module hubness index Z_i is defined as

$$Z_i = \frac{k_{iT} - \overline{k_T}}{\sigma_{kT}}, \tag{4}$$

where k_{iT} is the total weights of links between node i and other nodes in its own module T; and $\overline{k_T}$ and σ_{kT} are the mean and the standard deviation of intramodular link weights over all the nodes in module T, respectively.

The outside-module hubness index B_i is defined as

$$B_i = \frac{m_{iT} - \overline{m_T}}{\sigma_{mT}}, \tag{5}$$

where m_{iT} is the total weights of links between node i and other nodes outside of its own module T; and $\overline{m_T}$ and σ_{mT} are the mean and the standard deviation of out-module link weights over all the nodes in module T, respectively.

3.2.2. Participation index

We apply two participation indices to characterize the distribution of a node's connections inside and outside the modules, denoted as PZ_i and PB_i, respectively.

To calculate the participation index P_i, we utilize the concept of participation vector P_i (Klimm et al., 2014). From the regional perspective, we propose an inside-module participation vector PZ_i, whose elements represent the probabilities of links between node i and other nodes in its own module T. From the national perspective, we construct an outside-module participation vector PB_i to capture the probabilities of links between node i and modules other than T. For instance, suppose a network in which node i has links equally distributed to 3 nodes within its module and 3 other modules; then both PZ_i and PB_i will be represented as $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

To obtain participation indices PZ_i and PB_i, we calculate the standard deviation of the participation vector $\sigma(\mathbf{P}_i)$ and compare it with the most extreme case $\sigma_{\max}(m)$, i.e., all links of node i are connected to a single node or module. Here, m refers to the number of other nodes inside module T or the number of other modules. Therefore, the participation index is defined as

$$P_i = 1 - \frac{\sigma(\mathbf{P}_i)}{\sigma_{\max}(m)} = 1 - \frac{m}{\sqrt{m-1}} \sigma(\mathbf{P}_i). \tag{6}$$

For a node having links equally distributed to other nodes inside its module (or other modules), we have $PZ_i = 0$ (or $PB_i = 0$). For a node that connects to only a single node inside its module (or only one other module), we have $PZ_i = 1$ (or $PB_i = 1$).

3.2.3. Role identification framework

By calculating the hubness indices (Z_i and B_i) and participation indices (PZ_i and PB_i), we can obtain two parameter spaces, $Z - PZ$ and $B - PB$, describing the position each node takes in the network from regional and national perspectives. Considering directions, we end up with four parameter spaces, $Z_{supply} - PZ_{supply}$ and $B_{supply} - PB_{supply}$ for incoming flows and $Z_{demand} - PZ_{demand}$ and $B_{demand} - PB_{demand}$ for outgoing flows.

We classify cities into four types of roles according to their positions in the parameter spaces. Cities with a hubness index >2 are defined as hubs. Hubs are further divided into three categories according to the extent of their connections with other cities. Those with a participation index of 0–0.3 are defined as exclusive hubs, and they are connected to only a few cities. Those with an index of 0.3–0.6 are defined as inclusive hubs, and they have relatively extensive connections. Those with an index of 0.6–1 are defined as extensive hubs, and they have the most extensive connections.

3.3. Patient mobility modeling and interpretation

To model patient mobility, a traditional gravity model and three machine learning methods were employed. We used multisource data to construct a dataset to characterize cities from three aspects: social economy, medical resources, and traffic convenience (refer to the Supplementary Information for detailed features). Data were randomly divided into a training set and a testing set at a ratio of 7:3. The effectiveness of the models was evaluated by calculating the root mean squared error (RMSE) and the goodness-of-fit (R^2) for the testing set.

3.3.1. Gravity model

The gravity model is one of the most common spatial interaction models. In Zipf's equation proposed in 1946, the intensity of the mobility flow from i to j can be approximated by $T_{ij} \propto \frac{P_i P_j}{r_{ij}}$, where P_i and P_j are the respective populations and r_{ij} is the distance between i and j (Zipf, 1946). To predict the intensity of intercity patient flow, we developed a gravity model that takes into account the expected medical benefits and the distance decay effect (Liu et al., 2014). It is defined as follows:

$$W_{ij} = aP_i^b P_j^c \left(\frac{M_j}{M_i}\right)^d f\left(d_{ij}\right)^e, \tag{7}$$

where W_{ij} is the number of patients going to city j from city i, P_i and P_j are the population of origin and destination, respectively, $\frac{M_j}{M_i}$ is the ratio of influential CKD doctors, and $f(d_{ij})$ is the complementary cumulative distribution function of patients' travel distance derived from empirical data. To extract feature importance, we normalized the input data so that the coefficients obtained were comparable.

3.3.2. Machine learning models

Three machine learning models were also employed, including support vector machine (SVM), random forest (RF), and eXtreme gradient boosting (XGBoost) (Breiman, 2001; Chen & Guestrin, 2016; Cortes & Vapnik, 1995). A fivefold cross-validation was performed on the training set for hyperparameter tuning using a grid search strategy. Despite their impressive performance in real-world applications, a major drawback of machine learning algorithms is the lack of interpretability. They are often referred to as "black boxes" because it is difficult to infer how feature variables affect model prediction. To address this issue, we adopted a novel technique in the field of explainable artificial intelligence (XAI), namely, the Shapley additive explanation (SHAP) method (Adadi & Berrada, 2018). It is a unified framework that allocates the impact of each feature on a particular prediction based on Shapley values originating from game theory (Lundberg et al., 2018; Lundberg et al., 2020; Lundberg & Lee, 2017).

SHAP provides a local perspective to understand the impact of features. For each prediction, we calculated the SHAP value of each feature, which holds an additivity property that can be expressed by:

$$f(x) = \phi_0(f, x) + \sum_{i=1}^{m} \phi_i(f, x), \tag{8}$$

where $f(x)$ is the prediction, $\phi_0(f, x)$ is the expected baseline value of the model over the training data ($E[f(x)]$), m is the number of features, and $\phi_i(f, x)$ is the SHAP value of feature i.

4. Results

4.1. Distribution of travel distance

We extracted a total of 4,353,885 inpatient records of patients with CKD over a five-year period from 2014 to 2018 (see the Methods section). This dataset has been preanonymized, and patients' geo-location information has been aggregated to the city level. Over years one through five, the proportion of cross-city inpatients to the total number was 22.2 %, 22.1 %, 22.4 %, 22.9 % and 22.3 %, respectively. The patient mobility flow map is shown in Fig. 2. The figure shows that dominant intercity mobility flows are confined to provincial-level administrative regions.

Using the obtained nationwide mobility flow data, we found that the truncated power law distribution significantly outperformed other candidate distributions (p *value* < 0.001) for the tail of the displacement distribution (150 km above) (Clauset et al., 2009), where β=1.81 and k=1758.5:

$$P(r) \sim r^{-\beta} exp(-r/k). \tag{9}$$

In addition, we examined differences among diverse groups of patients. We first classified cities into high-, middle- and low-income classes according to their contributions to the gross domestic product (GDP). The result is shown in Fig. 3b. The displacements of patients in high-income cities are longer than those of patients in other cities. Such a difference becomes largest when $P = 0.2$, which means that approximately 80 % of displacements in low- and middle-income cities are <300 km, and for high-income cities, this upper limit reaches 800 km. This finding suggests that most patient mobility flows in low- and middle-income cities are confined to the provincial level, while patients in high-income cities have more opportunities to seek healthcare across provinces.

We also examined behavioral differences among patients in different physical geographic regions (refer to the Supplementary Information for division maps). First, we used the well-known *Hu Line*, which divides China into densely populated east and sparsely populated west. The results illustrate that patients from west of the *Hu Line* have longer displacements, and the difference between the two is reflected mainly in travel exceeding 200 km (Fig. 3c). Regional differences are also reflected in Fig. 3d, which further divides the country into four regions: the south, the north, the northwest, and the Qinghai-Tibet Plateau. For patient flows exceeding 100 km, the distances traveled by patients from the southern region are often less than those from the other three regions. Among them, patients from the northwest have the longest travel distances, reflecting the poor accessibility of medical resources in this region.

4.2. Basic topological properties of the IPMN

We calculated the average weight of each intercity edge over 5 years to construct a weighted directed network, since the proportion of cross-city visits was relatively stable during this time. The basic topological

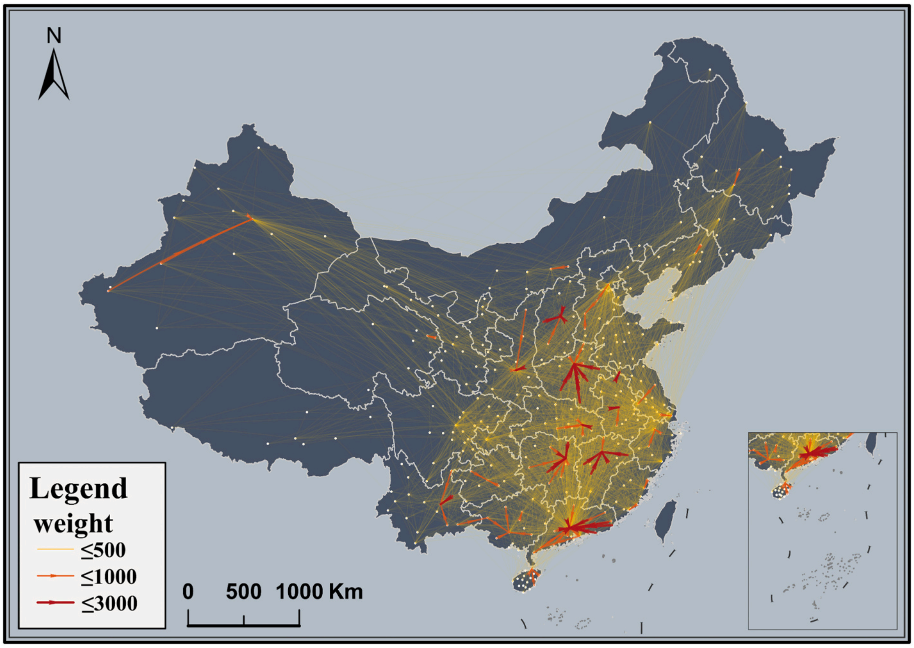

Fig. 2. Patient mobility flow map. White dots represent cities, and gray lines are the boundaries of Chinese provincial-level administrative regions. The edges are set to three gradient colors according to the hierarchy of weights.

properties of the IPMN are reported in Fig. 4. The IPMN is a network with 358 nodes (i.e., cities) and 6824 edges (i.e., patient flows). The average degree $<K>$ is 587.288, and the average shortest path length $<L>$ is 1.562.

The assortativity coefficient is the Pearson correlation coefficient of the degree between pairs of linked nodes (Newman, 2003). It measures the tendency that nodes are preferentially connected with other nodes with similar degree values. We investigate the relationship of in-degrees between two ends of the connected edges, expressed by the in-assortativity coefficient *As_in*. For out-degrees, we calculated the out-assortativity coefficient *As_out*. We discovered that the IPMN shows a weak disassortativity with respect to in-degrees ($As\_in = -0.087$); in other words, cities with high in-degrees are more likely to be linked to cities with low in-degrees. Moreover, the out-assortativity coefficient suggests a slight tendency of assortative mixing ($As\_out = 0.135$), i.e., cities connected to high out-degree destinations usually have large outflows as well.

This pattern is also verified in maps of in-degrees and out-degrees. As seen in Fig. 4b, cities in the central region of China have a relatively high out-degree, and these cities generally have a large population with relatively good economic conditions but weak medical resources. In contrast, Fig. 4c shows a clear polarization, indicating that high-quality medical resources are aggregated in a few cities.

We also examined the distributions of in-degrees and out-degrees using the statistical method proposed by Clauset (Clauset et al., 2009). The log-normal distribution fits the data well and significantly outperforms other candidate distributions, including the power law, exponential and truncated power law ($p\ value < 0.001$) (Alstott et al., 2014). The complementary cumulative distributions of in-degrees and out-degrees are shown in Fig. 4d and e, respectively.

4.3. *Multiscale community structure of the IPMN*

As shown in Fig. 5a and b, the two scale-fixed community detection algorithms, modularity optimization and Infomap, lead to similar partitions. We summarize the results as "universal administrative constraints and a few boundary breaches". Cities tend to cluster in the same provincial region because they have similar policies and cultural practices. However, several cities have tighter links with other provinces than their own, breaking this administrative boundary. We classify our findings in community detection into three types. The type A module is made up of individual provincial regions, the type B module is composed of multiple provincial regions, and the type C module contains the "boundary breakers" mentioned above.

There are two major differences between the two algorithms, i.e., the northwest C2 and southeast C3 modules obtained by modularity optimization were refined into multiple modules with the Infomap algorithm. Both methods grouped northeastern China into one module, C1, which indicates Beijing's strong single core position in this region. In addition, both methods grouped Shanghai, Zhejiang, Jiangsu and Anhui into the same module B1, which reflects the influence of the Yangtze River Delta urban agglomeration. The influence of urban agglomerations can also be seen in many other places. For instance, the Beijing-Tianjin-Hebei, Pearl River Delta, and Chengdu-Chongqing urban agglomerations (for further literature on Chinese urban agglomerations, see reference (Fang, 2015)).

Boundary breach is also an interesting phenomenon in regional development. Inner Mongolia, the province with the longest latitudinal geographical distance, was divided into two modules, C1 and C2, by both algorithms. Module C4 as obtained by modularity optimization takes in two cities on the edge of Yunnan and Guizhou provinces; module C4 as identified by Infomap absorbs several cities in Gansu, and module C6 absorbs the southernmost city in Jiangxi. These cities are attracted by the richer medical resources of nearby provinces, thus alleviating the constraining effect of provincial administrative regions.

CPM optimization, a resolution-limit-free algorithm, was applied to explore the network structure of the IPMN at a finer scale. By adjusting the resolution parameter γ, we obtained submodules of the IPMN (Fig. 5c and d).

This partition further divided the large modules into small ones based on the results of previous methods. For example, the original module C1 was divided into C1, C2, C3 and C4, i.e., four provincial dominant submodules, while cities with closer connections in the Yangtze River Delta urban agglomeration were identified as module C8; the Chengdu-Chongqing urban agglomeration was separated from previous large modules. In addition, some cities are identified as individual

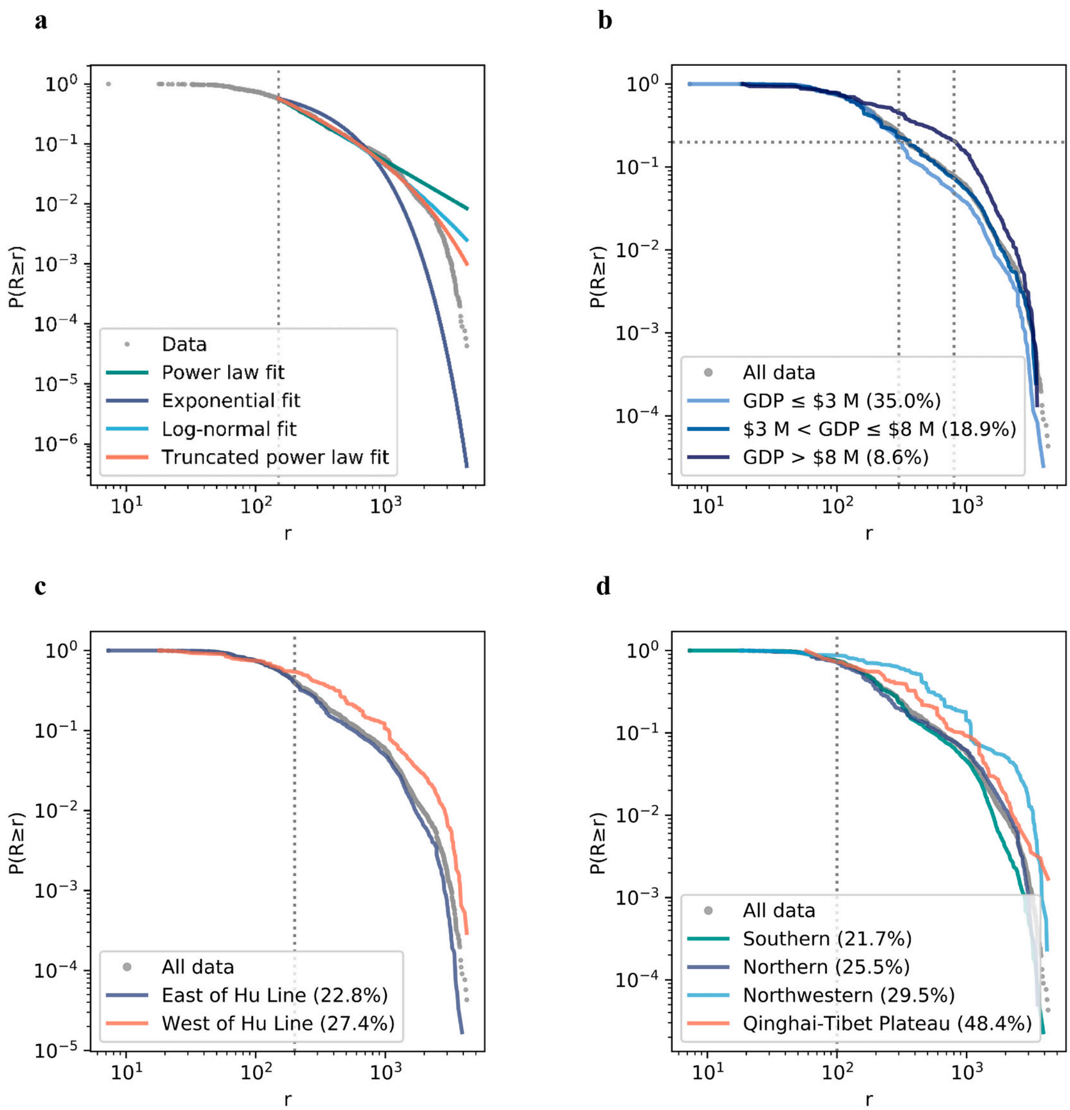

Fig. 3. Complementary cumulative distribution of patient's travel distance. a. Fitted by different functions. b. Patients divided by economic status. c. Divided by *Hu Line*. d. Divided by four physical geographic regions. Percentages in parentheses refer to the ratio of nonlocal visits.

clusters, and we merged them into adjacent submodules for the convenience of analysis (Fig. 5d).

With a relatively high quality of partition (modularity score $Q = 0.736$), CPM optimization yielded submodules that are relatively uniform in geographical area, making them potentially appropriate units for providing practical policy advice. Therefore, we term such divisions "hospital service modules" (HSMs) to study the regional characteristics of patients' healthcare-seeking behavior (Fig. 5d).

4.4. *Mobility flow across HSMs*

To validate the rationality of using the HSM as the basic zoning unit, we calculated the proportion of intramodular flows. The result shows that the proportion of flows within the HSM reaches 80.0 %, while the proportion of flows within the provincial administrative districts is 74.9 %, which indicates that HSMs can better describe the community structure of the IPMN.

We further calculated the cost of the medical visit flows inside and outside the HSMs. Here, we consider geographic distance as a proxy for medical cost, as a longer distance indicates a higher payment burden. In addition, we calculated the scaled distance considering the difference in affordability between the origin and the destination. In particular, we use the ratio of the mean wage of employees and workers between the two cities as a scaling factor. This dataset originates from the *China City Statistical Yearbook* (National Bureau of Statistics of China, 2015), which covers 86 % of the patient flow data. For each edge in the IPMN, its cost is defined as the product of the great circle distance between two cities and the intensity (number of patients) of that edge.

The distribution of link weight, distance cost and scaled distance cost of visits inside and outside HSMs are shown in Fig. 6. Notably, cross-module healthcare seeking flows accounted for 52.5 % of the distance with 20 % of the total link strength, indicating that a large percentage of patients seek medical treatment over long distances, which also makes the entire system quite uneconomic. If the income gap between origin and destination is taken into consideration, this proportion rises to 54.8 %. This result shows that most patients travel from cities with poorer economic conditions. Such an income gap will further amplify the burden of seeking nonlocal medical care (Table 1).

4.5. *Cities' roles in the IPMN*

Cities' positions in 4 parameter spaces are shown in Fig. 7 (for map visualizations, see the Supplementary Information), and the functional roles of the cities in the IPMN are well captured by the subfigures.

a

# of Nodes	# of Edges	<K>	<L>	As_out	As_in
358	6824	587.288	1.562	0.135	-0.087

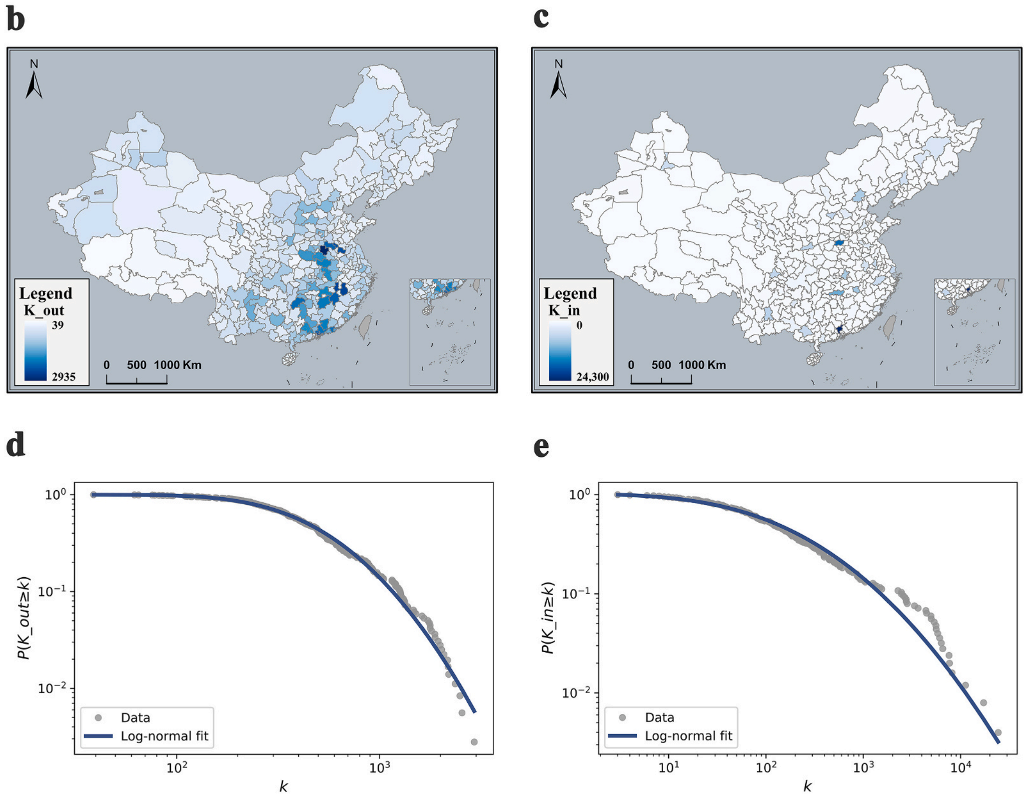

Fig. 4. Basic topological properties of the IPMN. a. Basic metrics of the IPMN, including the number of nodes and edges, the average of degrees and shortest path lengths, the assortativity coefficient of out-degrees and in-degrees. b–c. Cities' out-degrees and in-degrees, with city boundaries indicated in gray. d–e. Complementary cumulative distributions of out-degrees and in-degrees.

The patient outflow reflects the situation on the demand side (Fig. 7a and b); we first focus on the situation of cross-module outflows because they represent the behavior of long-distance healthcare seeking. Fig. 7a shows that the hubness of a city is facilitated by population and demand; that is, cities with large populations or low levels of medical care resources in this module tend to become demand hubs. The so-called four first-tier cities (Beijing, Shanghai, Guangzhou and Shenzhen) and Chongqing, which have the largest city-wide populations, were all identified as hubs. The diversity of cross-module healthcare-seeking choices is driven by cities' economic conditions. As shown in the figures, cities with more developed economies tend to have more choices and greater participation indices.

In comparison, we pay attention to the parameter space for cross-module inflows. Fig. 7c shows that cities identified as hubs have undertaken many incoming flows that do not originate from this module. Among them, exclusive hubs and inclusive hubs are regional supply centers that accept patients mainly from neighboring provinces. Extensive hubs accept more diverse inflows and thus can be regarded as national centers. Beijing, Shanghai and Guangzhou have the highest hubness index, showing their status as nationwide high-quality medical centers. As a rapidly developing city, Shenzhen's lack of medical resources is also reflected in its roles, and it was not identified as a hub from either the inside-module or outside-module supply perspectives.

This outshining of provincial capitals is also reflected in the role of cities from the perspective of inside-module outgoing and incoming flows, as shown in Fig. 7d. The two figures show a symmetrical trend, meaning that patients move inside the module with obvious preferences, so supply centers receive a wide range of inflows from cities with fewer choices for nonlocal healthcare. Most hubs identified in Fig. 7b are relatively underdeveloped cities, while in Fig. 7d, a considerable number of provincial capital cities are identified as supply centers, reflecting their primacy advantages.

4.6. Mobility models and influential factors

The results of model validation are shown in Fig. 8. We constructed a traditional gravity model and three machine learning models, namely, SVM, RF, and XGBoost, with a multisource city feature dataset to predict the number of patients traveling between cities. The XGBoost model achieved the best performance (R^2=0.77, RMSE = 59.0). To understand the impact of individual features on the model predictions, we adopted the SHAP framework to explain the XGBoost model.

We calculated each feature's contribution to the difference between the predicted value and the baseline value for each sample, i.e., the SHAP value. The fifteen variables with the highest mean absolute SHAP values are shown in Fig. 9a. The three variables with the largest impact on model prediction are distance (*Distance*) between the two cities, whether the two cities are in the same HSM (*Region_flag*) and the number of influential kidney doctors per capita at the destination (*Doc_renal_per*, data obtained from online healthcare platforms, see 2.2 for details).

As seen in Fig. 9b, the effect of distance is nonlinear, with close distances adding the predicted values up to >600 but long distances reducing the predicted values by no >200. We use *Region_flag* to encode whether the two cities are in the same HSM, with 1 representing that

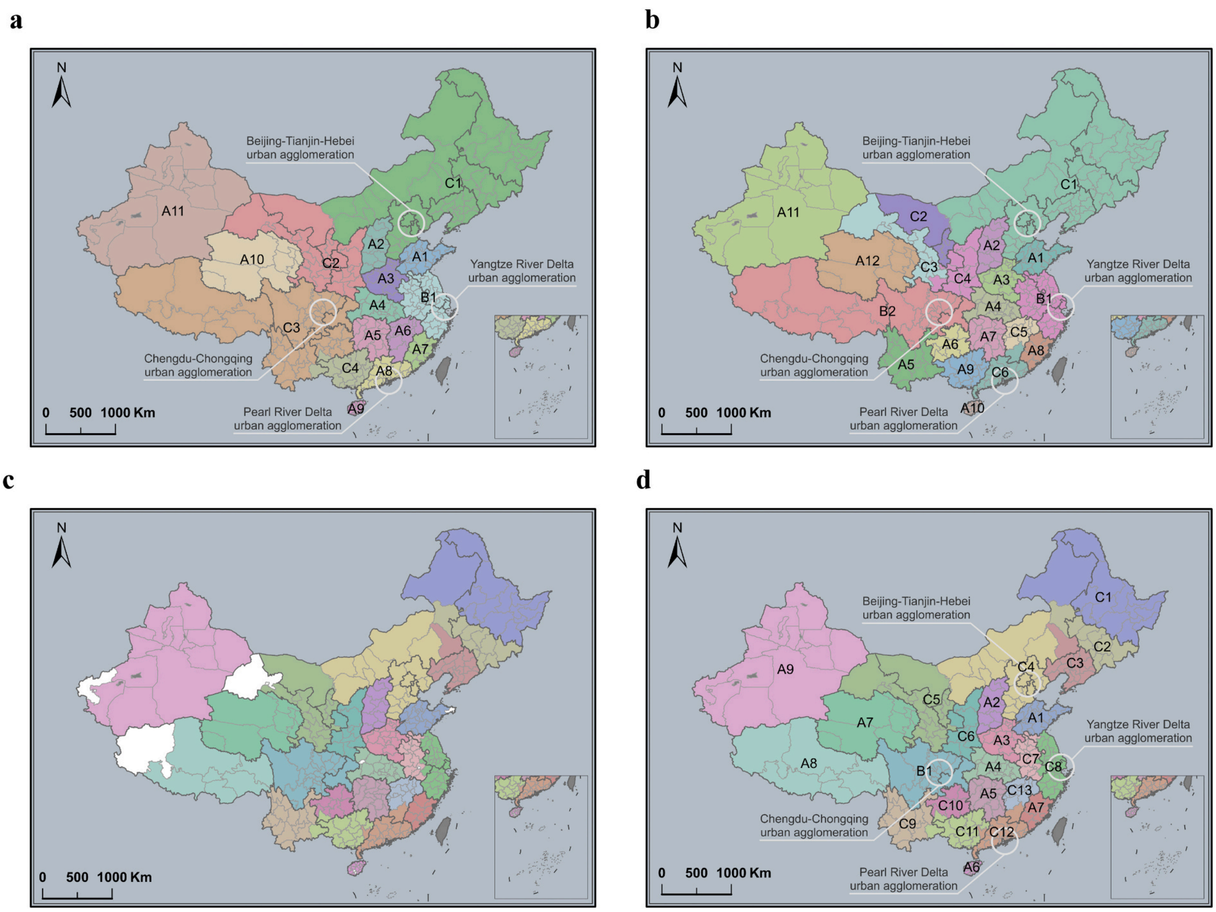

Fig. 5. Multiscale community structure of the IPMN a. The modules detected by modularity optimization (modularity $Q = 0.753$). b. The modules detected by Infomap ($Q = 0.748$). c. The original output of CPM optimization; cities in white were originally identified as individual clusters. d. The submodules given by modified c ($Q = 0.736$, resolution parameter $\gamma = 6.626$). Dark gray lines represent the boundaries of provincial-level administrative regions, and light gray lines represent the boundaries of cities. Cities with no mobility data are indicated in dark gray.

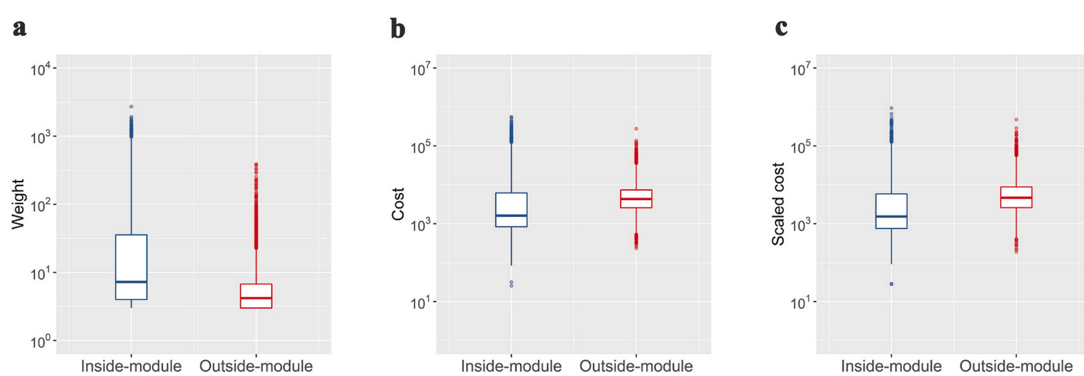

Fig. 6. Distribution of weight, cost and scaled cost of mobility flows inside and outside the HSMs. a. Weight b. Cost c. Scaled cost.

Table 1
Proportion of weight, cost and scaled cost of mobility flows inside and outside the HSMs.

Mobility flow type	Weight %	Cost %	Scaled cost %
Inside-module	80.0	47.5	45.2
Outside-module	20.0	52.5	54.8

both are in the same HSM and 0 indicating the opposite. As a distance-related variable, it remains the second most important feature, indicating that the model captured other information implied by it, such as the fact that the same module tends to have similar cultural habits or be in the same administrative unit. We obtained the number of renal doctors in each city from popular online paid healthcare platforms in China and considered them to be high-level and well-recognized. The results suggest that this kind of online data source can reflect the high-quality medical service capabilities of a city. Our results show that cities with more influential doctors per capita (*Doc_renal_per*) tend to become a popular destination for patients, while lower values result in slightly

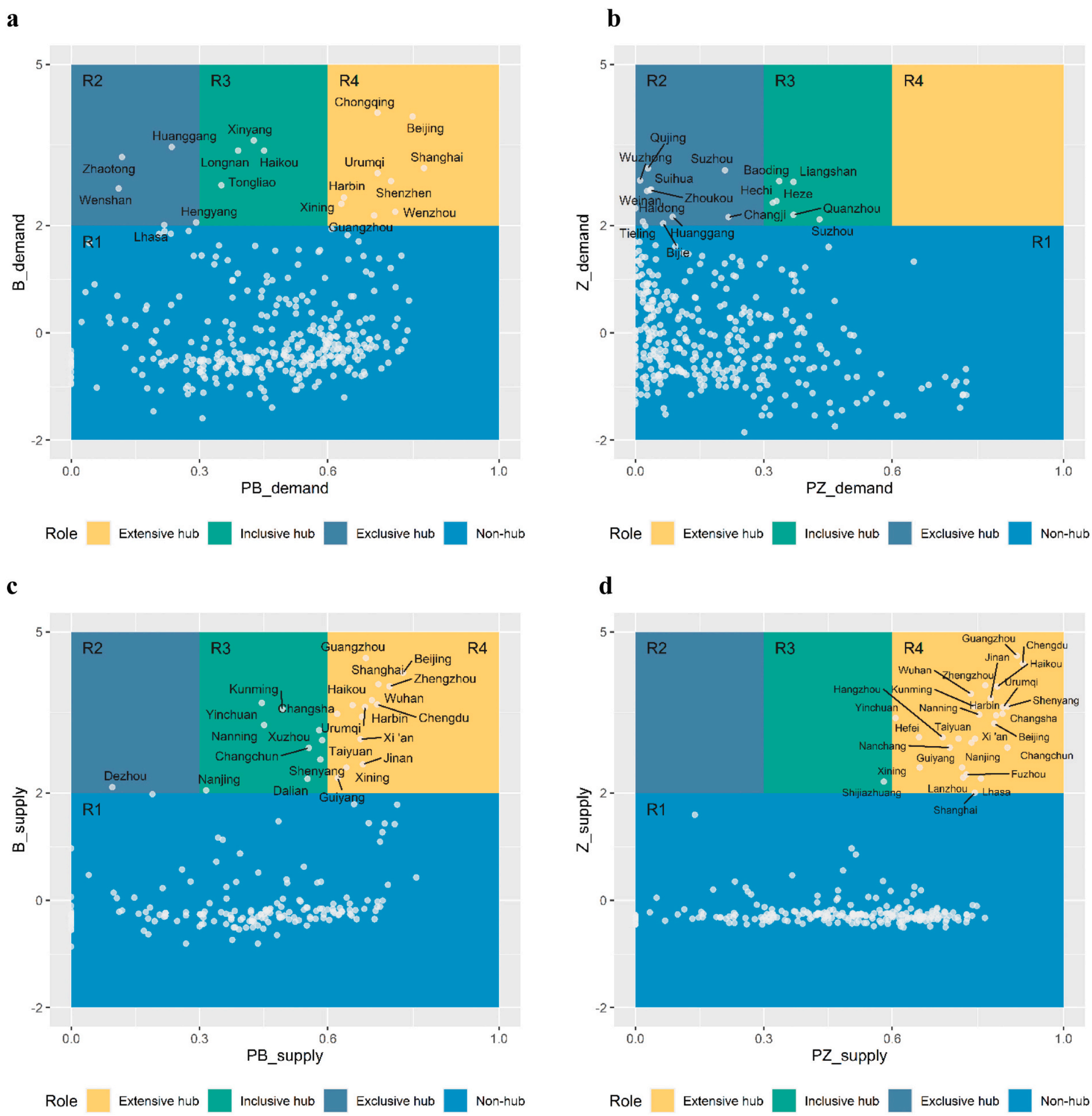

Fig. 7. Roles of cities in each parameter space. R1, R2, R3, and R4 represent nonhubs, exclusive hubs, inclusive hubs, and extensive hubs, respectively. a. Outside-module demand perspective b. Inside-module demand perspective c. Outside-module supply perspective d. Inside-module supply perspective.

reduced traffic between two cities.

Each subplot in Fig. 10 provides details of the impact of each feature on model prediction. We pay particular attention to the cutoff where the SHAP value is 0, as this point determines whether the feature's impact on the prediction is positive (predicted value should be higher than the baseline) or negative. Here, we provide the results for the three most influential features, which represent the impact of geographic distance (*Distance, Region_flag*) and medical resources (*target.Doc_renal_per*). We found that the effect of distance on model prediction shows an interesting power-law-like trend. The negative effect on the model stabilizes with increasing distance, and 200 km is a cutoff that determines the direction of the effect. Fig. 10c illustrates the attractiveness of *Doc_renal_per*. We observed that when a city has more than approximately 20 high-level renal doctors per million people, it will become an attractive destination for nonlocal patients.

5. Discussion and conclusion

To the best of our knowledge, our study is one of the pioneering works that employed a real-world dataset to investigate the intercity mobility behavior of patients nationwide. This paper provides a methodological reference for patient mobility studies using interdisciplinary approaches. The proposed framework can also be extended to other countries and regions to promote the equalization of healthcare resources. The most important contributions of this study are the new understanding of the behavioral patterns of patients as a specific population and their implications for public health policies.

In terms of the travel distance of patients, we found that the distribution differs from previous studies on human mobility (Brockmann et al., 2006; Gonzalez et al., 2008; Han et al., 2011; Noulas et al., 2012). The truncated power law fits better than other functions, but the tail is steeper, which indicates that healthcare-seeking behavior is constrained by financial cost. From the perspective of complex networks, we find

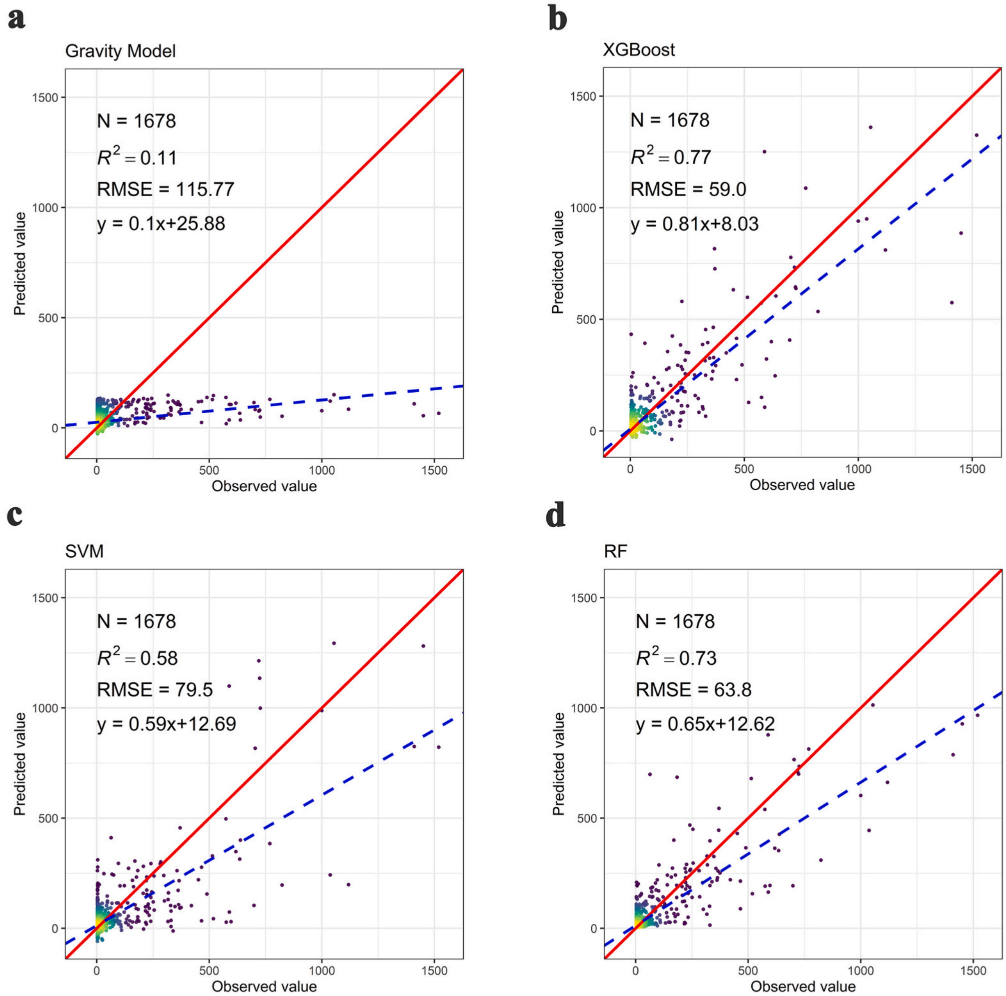

Fig. 8. Performance of patient mobility models. a. Gravity model b. XGBoost c. SVM d. RF.

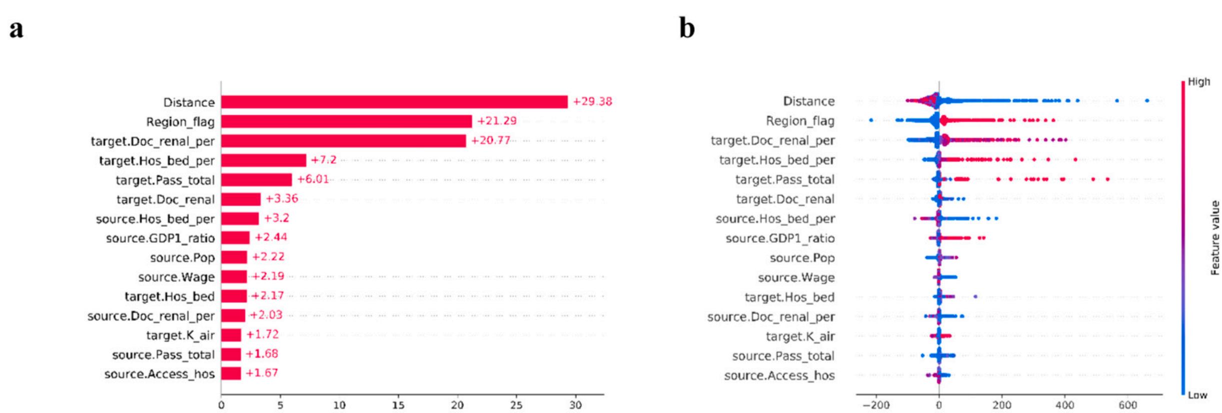

Fig. 9. Top 15 variables with the greatest impact on model prediction. a. Summary plot of the importance of the features ranked by the mean absolute SHAP value b. The impact of the features, where each point represents a sample of data in the training set, colored according to its value.

that the in- and out-degrees of the IPMN follow a log-normal distribution, which is consistent with many real-world networks (Broido & Clauset, 2019).

We examined the self-organized community structures of the IPMN using community detection methods. We summarize the results of community detection as "universal administrative constraints and a few boundary breaches". It is influenced by a mixed effect of administrative regions, urban agglomerations, geographic locations and healthcare resources. The mismatch between the behavioral community and administrative boundary is also found in the division of HSAs in the United States (The Center for the Evaluative Clinical Sciences, Dartmouth Medical School, 1996). However, unlike HSAs and HRRs in the United States, few studies have focused on healthcare policy zoning units in China. When applying healthcare policies, zoning units play an

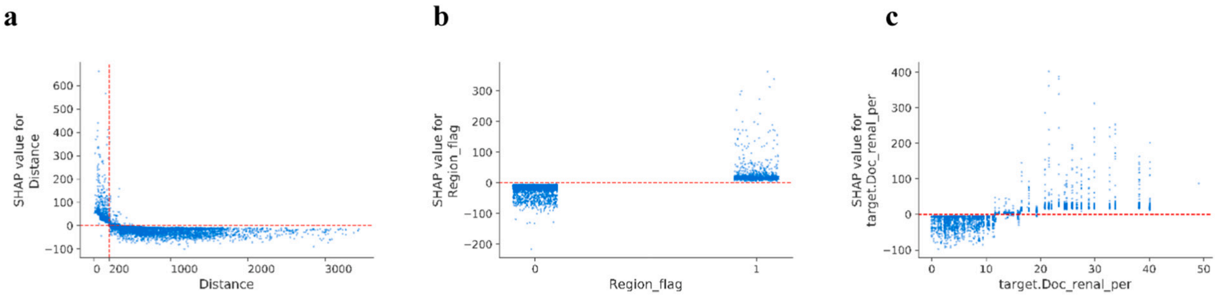

Fig. 10. Details of SHAP values for selected features a. Distance b. Whether the two cities are in the same module c. Number of influential renal doctors per capita at destination.

important role in decision-making, such as the construction of regional medical centers or the designation of counterpart support hospitals, both of which require delineating areas of responsibility. Our results show that HSMs can better describe the clustering pattern of patient behavior than provincial boundaries.

After constructing HSMs, we focused on what role each city plays. We identified cities' roles according to their hubness and participation indices inside and outside HSMs. Considering the imbalanced development of China's cities, it is necessary to consider spatial heterogeneity when applying policies based on the different characteristics of each city. For instance, counterpart support can be established between supply and demand hub cities in the same HSM. Regional medical centers can be prioritized to be built in supply hub cities to reduce their pressure or to cover multiple demand hub cities to avoid unnecessary long-distance mobility. Healthcare resources in demand hub cities can be strengthened to reduce patients' burden of long-distance travel, especially for outside-module demand hubs. For cities with low participation indices, more attention should be given to improving the settlement mechanism of medical treatment in popular destinations.

The modeling results indicate that distance and healthcare resources are the most important factors driving patient flows. *Distance* has a power-law-like effect on patient travel behavior, and as an indicator correlated with *Distance*, *Region_flag* serves as the second most important factor, which validates the rationality of the data-driven HSMs as a basic zoning unit. We used a novel data source to measure each city's capability of high-quality healthcare services. Additionally, the performance of the mobility model validates that such data indeed have good explanatory power for real-world patient behaviors. As revealed by the results, high-quality healthcare resource development plays a key role in solving the issue of intercity patient mobility.

Due to limited data availability, only data from patients with CKD were used. However, according to our calculations, the rate of nonlocal visits for CKD patients (22 %) is much higher than the average (8 %). Moreover, the allocation of medical resources needs to be tailored to different kinds of diseases, and our framework can be fully applicable to other diseases as well. In the future, improvements can be made on the present work, starting with integrating larger-scale datasets for diverse or universal patient mobility patterns of multiple diseases. Additionally, this study was based on cities as nodes of the complex network, leading to the loss of some location information in the aggregation process. Analysis at a finer-resolution spatial unit similar to the HRRs in the US could be conducted to promote cooperation among cities (Jia et al., 2020b).

Urbanization has facilitated the agglomeration of resources, including high-quality healthcare service resources. This inequality has led to the cross-city mobility of patients. With the continuing development of urbanization, the flows of population, resources and information between cities will be further accelerated (Hu et al., 2020; Pan & Lai, 2019; Ye & Liu, 2019; Zhang et al., 2020a; Zhao et al., 2014). As we can observe from the structure of the patient mobility network, it has different characteristics from other kinds of networks between cities. For example, in the IPMN, the edges with the highest weight are mostly confined to the provincial administrative region, while for the national population flow network, the cross-provincial high-volume population flows are much more common (Pan & Lai, 2019; Zhang et al., 2020a). The differences between the IPMN and the population flow network are also shown in the distribution of in-degrees. The cities with the highest in-degrees in the IPMN absorb far more patient inflows than other cities, reflecting a more severe polarization of healthcare resources. Han (Han et al., 2011) proved that the mechanism of scaling law in human mobility is related to the hierarchical nature of traffic systems. However, as we found in this paper, the displacement distribution of patients does not follow this law, suggesting a different structure of the IPMN. Patient mobility reflects the uneven development of different resources amidst China's urbanization. As shown in our study, patients are now able to seek healthcare in more distant cities thanks to the convenience of intercity traffic systems. However, cross-module visits which accounted for only 20 % of total visits, accounted for >50 % of the total travel distance, which also demonstrates the potential of relevant policies such as regional medical centers to avoid unnecessary long-distance travel. Both the government and academics should strive to further understand and optimize the patient mobility network. Efforts should be made to reduce nonlocal visits and unnecessary long-distance travel for patients to promote the equality and accessibility of public health resources.

In conclusion, the wide use of information technology has made it possible to study patient mobility at a large spatial scale. As a subset of human mobility, patient mobility has strong practical relevance. In particular, for developing countries with relatively poor equity in healthcare resources, patient mobility reflects the mismatch between resources and population, and optimizing patient mobility should be a goal for policymakers. Several emerging data sources, such as electronic hospital records, mobile phone and taxi data, provide us with the unprecedent opportunity to perform a detailed analysis of patient mobility at multiple spatial scales, ultimately serving to build a more people-centered and cost-effective healthcare system.

CRediT authorship contribution statement

Jiaqi Ding: Conceptualization, Methodology, Software, Writing – original draft. **Chao Yang:** Methodology, Software, Writing – original draft. **Yueyao Wang:** Software, Visualization. **Pengfei Li:** Methodology, Data curation. **Fulin Wang:** Data curation. **Yuhao Kang:** Writing – review & editing. **Haoyang Wang:** Methodology. **Ze Liang:** Conceptualization. **Jiawei Zhang:** Writing – review & editing. **Peien Han:** Writing – review & editing. **Zheng Wang:** Visualization. **Erxuan Chu:** Writing – review & editing. **Shuangcheng Li:** Supervision, Project administration. **Luxia Zhang:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

None.

Data availability

The authors do not have permission to share data.

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (72125009, 41590843, 91846101, 82003529, 81771938, 82090021), National Key Research and Development Program of the Ministry of Science and Technology of China (2019YFC2005000), Chinese Scientific and Technical Innovation Project 2030 (2018AAA0102100), the University of Michigan Health System-Peking University Health Science Center Joint Institute for Translational and Clinical Research (BMU2018JI012, BMU2019JI005, 71017Y2027), CAMS Innovation Fund for Medical Sciences (2019-I2M-5-046), and PKU-Baidu Fund (2019BD017, 2020BD004, 2020BD005, 2020BD032). The authors thank the Bureau of Medical Administration and Medical Service Supervision, National Health Commission of China for the support to this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cities.2022.103975.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Aggarwal, A., et al. (2018). Determinants of patient mobility for prostate cancer surgery: A population-based study of choice and competition. *European Urology, 73*, 822–825.

Alstott, J., Bullmore, E., & Plenz, D. (2014). Powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE, 9*, Article e85777.

Andritsos, D. A., & Tang, C. S. (2014). Introducing competition in healthcare services: The role of private care and increased patient mobility. *European Journal of Operational Research, 234*, 898–909.

Azzopardi-Muscat, N., et al. (2018). The role of the 2011 patients' rights in cross-border health care directive in shaping seven national health systems: Looking beyond patient mobility. *Health Policy, 122*, 279–283.

Baeten, R. (2014). Cross-border patient mobility in the European Union: In search of benefits from the new legal framework. *Journal of Health Services Research & Policy, 19*, 195–197.

Barber, R. M., et al. (2017). Healthcare access and quality index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990–2015: A novel analysis from the global burden of disease study 2015. *The Lancet, 390*, 231–266.

Bazzani, A., Giorgini, B., Rambaldi, S., Gallotti, R., & Giovannini, L. (2010). Statistical laws in urban mobility from microscopic GPS data in the area of Florence. *Journal of Statistical Mechanics: Theory and Experiment, 2010*, Article P05001.

Beal, E. W., et al. (2019). Association between travel distance, hospital volume, and outcomes following resection of cholangiocarcinoma. *Journal of Gastrointestinal Surgery, 23*, 944–952.

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*, Article P10008.

Borno, H. T., Zhang, L., Siegel, A., Chang, E., & Ryan, C. J. (2018). At what cost to clinical trial enrollment? A retrospective study of patient travel burden in cancer clinical trials. *The Oncologist, 23*, 1242–1249.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Bright, E. A., Rose, A. N., & Urban, M. L. (2016). *LandScan 2015* (2015 ed.). Oak Ridge National Laboratory.

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature, 439*, 462–465.

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications, 10*, Article 1017.

Chen, T., & Guestrin, C. (2016). XGBoost. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM.

Chikanda, A., & Crush, J. (2019). south-south cross-border patient travel to South Africa. *Global Public Health, 14*, 326–339.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51*, 661–703.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

Diaz, A., Burns, S., Paredes, A. Z., & Pawlik, T. M. (2019). Accessing surgical care for pancreaticoduodenectomy: Patient variation in travel distance and choice to bypass hospitals to reach higher volume centers. *Journal of Surgical Oncology, 120*, 1318–1326.

Diaz, A., et al. (2020). Accessing surgical care for esophageal cancer: Patient travel patterns to reach higher volume center. *Diseases of the Esophagus, 33*, 1–10.

Dugué, N., Labatut, V., & Perez, A. (2015). A community role approach to assess social capitalists visibility in the twitter network. *Social Network Analysis and Mining, 5*, Article 26.

Fang, C. (2015). Important progress and future direction of studies on China`s urban agglomerations. *Journal of Geographical Sciences, 25*, 1003–1024.

Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences, 104*, 36–41.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks, 1*, 215–239.

Fu, L., Xu, K., Liu, F., Liang, L., & Wang, Z. (2021). Regional disparity and patients mobility: Benefits and spillover effects of the spatial network structure of the health Services in China. *International Journal of Environmental Research and Public Health, 18*(3), Article 1096.

Glinos, I. A., Baeten, R., Helble, M., & Maarse, H. (2010). A typology of cross-border patient mobility. *Health & Place, 16*, 1145–1155.

Glinos, I. A., Baeten, R., & Maarse, H. (2010). Purchasing health services abroad: Practices of cross-border contracting and patient mobility in six european countries. *Health Policy, 95*, 103–112.

Gong, S., Gao, Y., Zhang, F., Mu, L., Kang, C., & Liu, Y. (2021). Evaluating healthcare resource inequality in Beijing, China based on an improved spatial accessibility measurement. *Transactions in GIS, 25*, 1504–1521.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*, 779–782.

Good, B. H., De Montjoye, Y.-A., & Clauset, A. (2010). Performance of modularity maximization in practical contexts. *Physical Review E, 81*(4), Article 046106.

Guimerà, R., & Nunes Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature, 433*, 895–900.

Han, X.-P., Hao, Q., Wang, B.-H., & Zhou, T. (2011). Origin of the scaling law in human mobility: Hierarchy of traffic systems. *Physical Review E, 83*(3), Article 036117.

Hanefeld, J., Lunt, N., Smith, R., & Horsfall, D. (2015). Why do medical tourists travel to where they do? The role of networks in determining medical travel. *Social Science & Medicine, 124*, 356–363.

Hillier, B., Turner, A., Yang, T., & Park, H.-T. (2010). Metric and topo-geometric properties of urban street networks: Some convergences, divergences and new results. *Journal of Space Syntax, 1*(2), 258–279.

Hu, X. Q., Wang, C., Wu, J. J., & Stanley, H. E. (2020). Understanding interurban networks from a multiplexity perspective. *Cities, 99*, Article 102625.

Huang, Y. M., et al. (2019). Spectrum of chronic kidney disease in China: A national study based on hospitalized patients from 2010 to 2015. *Nephrology, 24*, 725–736.

Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature, 582*, 389–394.

Jia, P., Wang, F., & Xierali, I. M. (2020). Evaluating the effectiveness of the hospital referral region (HRR) boundaries: A pilot study in Florida. *Annals of GIS, 26*, 251–260.

Jia, P., Xierali, I. M., & Fahui, W. (2015). Evaluating and re-demarcating the hospital service areas in Florida. *Applied Geography, 60*, 248–253.

Jiang, B., Yin, J., & Zhao, S. (2009). Characterizing the human mobility pattern in a large street network. *Physical Review E, 80*(2), Article 021136.

Kawamoto, T., & Rosvall, M. (2015). Estimating the resolution limit of the map equation in community detection. *Physical Review E, 91*(1), Article 012809.

Kitamura, R., Chen, C., Pendyala, R. M., & Narayanan, R. (2000). Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation, 27*, 25–51.

Klimm, F., Borge-Holthoefer, J., Wessel, N., Kurths, J., & Gorka, Z.-L. (2014). Individual node's contribution to the mesoscale of complex networks. *New Journal of Physics, 16*(12), Article 125006.

Kornelsen, J., et al. (2021). The rural tax: Comprehensive out-of-pocket costs associated with patient travel in British Columbia. *BMC Health Services Research, 21*, 854.

Laugesen, M. J., & Vargas-Bustamante, A. (2010). A patient mobility framework that travels: European and United States-mexican comparisons. *Health Policy, 97*, 225–231.

Legido-Quigley, H., Glinos, I., Baeten, R., & McKee, M. (2007). Patient mobility in the European Union. *BMJ, 334*, 188–190.

Leicht, E. A., & Newman, M. E. J. (2008). Community structure in directed networks. *Physical Review Letters, 100*(11), Article 118703.

Li, Y., Yan, X., & Song, X. (2019). Provision of paid web-based medical consultation in China: Cross-sectional analysis of data from a medical consultation website. *Journal of Medical Internet Research, 21*, Article e12126.

Liang, X., Zheng, X., Lv, W., Zhu, T., & Xu, K. (2012). The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics and its Applications, 391*, 2135–2144.

Liu, Y., Gong, L., & Tong, Q. (2014). Quantifying the distance effect in spatial interactions. *Acta Scientiarum Naturalium Universitatis Pekinensis, 50*, 526–534.

Liu, Z. (2021). E-healthcare sees sound development in country. *China Daily*. http://global.chinadaily.com.cn/a/202107/05/WS60e24aeaa310efa1bd65f9f5.html, 2021. (Accessed 7 September 2022).

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*. Curran Associates Inc.

Lundberg, S. M., et al. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering, 2*, 749–760.

Lundberg, S. M., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*, 56–67.

Lunt, N., & Mannion, R. (2014). Patient mobility in the global marketplace: A multidisciplinary perspective. *International Journal of Health Policy and Management, 2*, 155–157.

National Bureau of Statistics of China. (2015). *China statistical yearbook (2015)*. China Statistics Press.

National Health Commission of the People's Republic of China. (2019). *Implementation plan for the establishment of national medical centers and national regional medical centers*.

Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 74*, Article 036104.

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E, 67*(2), Article 026126.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE, 7*, Article e37027.

Pan, J. H., & Lai, J. B. (2019). Spatial pattern of population mobility among cities in China: Case study of the National day plus mid-autumn festival based on tencent migration data. *Cities, 94*, 55–69.

Pedersen, M., Omidvarnia, A., Shine, J. M., Jackson, G. D., & Zalesky, A. (2020). Reducing the influence of intramodular connectivity in participation coefficient. *Network Neuroscience, 4*, 416–431.

Pekala, K. R., et al. (2021). The centralization of bladder cancer care and its implications for patient travel distance. *Urologic Oncology: Seminars and Original Investigations, 39*(12), 834.e9–834.e20.

Riccardo, G., Armando, B., & Sandro, R. (2012). Towards a statistical physics of human mobility. *International Journal of Modern Physics C, 23*, 1250061.

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2010). The map equation. *The European Physical Journal Special Topics, 178*, 13–23.

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences, 105*, 1118–1123.

Sharma, K., Vangaveti, V., & Larkins, S. (2016). Geographical access to radiation therapy in North Queensland: A retrospective analysis of patient travel to radiation therapy before and after the opening of an additional radiotherapy facility. *Rural and Remote Health, 16*(1), Article 3640.

Song, C. M., Qu, Z. H., Blumm, N., & Barabasi, A. L. (2010). Limits of Predictability in Human Mobility. *Science, 327*(5968), 1018–1021.

Tao, W., et al. (2020). Towards universal health coverage: Lessons from 10 years of healthcare reform in China. *BMJ Global Health, 5*, Article e002086.

The Center for the Evaluative Clinical Sciences, Dartmouth Medical School. (1996). *The Dartmouth atlas of health care*. Amer Hospital Pub.

Tizzoni, M., et al. (2014). On the use of human mobility proxies for modeling epidemics. *PLoS Computational Biology, 10*, Article e1003716.

Traag, V. A., Van Dooren, P., & Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 84*, Article 016114.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports, 9*, Article 5233.

Wang, C., Wang, F., & Onega, T. (2021). Network optimization approach to delineating health care service areas: Spatially constrained Louvain and Leiden algorithms. *Transactions in GIS, 25*, 1065–1081.

Wang, C., Wang, F., & Onega, T. (2021). Spatial behavior of cancer care utilization in distance decay in the northeast region of the U.S. *Travel Behaviour and Society, 24*, 291–302.

Wang, F., & Wang, C. (2021). *GIS automated delineation of hospital service areas*. CRC Press.

Wang, F., Wang, C., Hu, Y., Weiss, J., Alford-Teaster, J., & Onega, T. (2020). Automated delineation of cancer service areas in northeast region of the United States: A network optimization approach. *Spat Spatiotemporal Epidemiol, 33*, Article 100338.

Wang, J. E., Du, F. Y., Huang, J., & Liu, Y. (2020). Access to hospitals: Potential vs. observed. *Cities, 100*, Article 102671.

Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., & González, M. C. (2012). Understanding road usage patterns in urban areas. *Scientific Reports, 2*, Article 1001.

Weiss, D. J., et al. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature, 553*, 333–336.

Xing, J. D., & Ng, S. T. (2022). Analyzing spatiotemporal accessibility patterns to tertiary healthcare services by integrating total travel cost into an improved E3SFCA method in Changsha, China. *Cities, 122*, Article 103541.

Xiong, X., & Luo, L. (2020). Inpatient flow distribution patterns at Shanghai hospitals. *International Journal of Environmental Research and Public Health, 17*(7), Article 2183.

Xu, M., Pan, Q., Muscoloni, A., Xia, H., & Cannistraci, C. V. (2020). Modular gateway-ness connectivity and structural core organization in maritime network science. *Nature Communication, 11*, 2849.

Xu, Z., et al. (2020). Centralizing rectal cancer surgery: What is the impact of travel on patients? *Diseases of the Colon and Rectum, 63*, 319–325.

Yan, X., Shan, L., He, S., & Zhang, J. (2022). Cross-city patient mobility and healthcare equity and efficiency: Evidence from Hefei, China. *Travel Behaviour and Society, 28*, 1–12.

Yang, C., et al. (2020). CKD in China: Evolving spectrum and public health implications. *American Journal of Kidney Diseases, 76*, 258–264.

Yang, C., et al. (2022). Healthcare resource utilisation for chronic kidney disease and other major non-communicable chronic diseases in China: A cross-sectional study. *BMJ Open, 12*, Article e051888.

Yang, Y., Zhang, X., & Lee, P. K. C. (2019). Improving the effectiveness of online healthcare platforms: An empirical study with multi-period patient-doctor consultation data. *International Journal of Production Economics, 207*, 70–80.

Ye, X., & Liu, X. (2019). *Cities as spatial and social networks*. Springer.

Zhang, L., et al. (2020). China kidney disease network (CK-NET) 2016 annual data report. *Kidney International. Supplement, 2011*(10), e97–e185.

Zhang, W. L., Chong, Z. H., Li, X. J., & Nie, G. B. (2020). Spatial patterns and determinant factors of population flow networks in China: Analysis on tencent location big data. *Cities, 99*, Article 102640.

Zhao, M., Liu, X., Derudder, B., Zhong, Y., & Shen, W. (2014). Mapping producer services networks in mainland Chinese cities. *Urban Studies, 52*, 3018–3034.

Zipf, G. K. (1946). The P1P2/D hypothesis: On the intercity movement of persons. *American Sociological Review, 11*, 677.