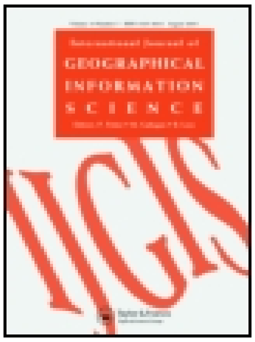

Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Detecting clusters over intercity transportation networks using K-shortest paths and hierarchical clustering: a case study of mainland China

Hanqiu Yue, Qingfeng Guan, Yongting Pan, Lirong Chen, Jianjun Lv & Yao Yao

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2019.1566551

RESEARCH ARTICLE

Detecting clusters over intercity transportation networks using K-shortest paths and hierarchical clustering: a case study of mainland China

Hanqiu Yue [a,b], Qingfeng Guan [a,b], Yongting Pan [a,b], Lirong Chen [a,b], Jianjun Lv [a,b] and Yao Yao [a,b]

[a]School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, China; [b]National Engineering Research Center of GIS, China University of Geosciences, Wuhan, Hubei, China

ABSTRACT

Intercity transportation infrastructures and services determine the depth and breadth of the spatial interactions among cities within an urban agglomeration, and have profound impacts on the spatial structure of the urban agglomeration. To evaluate whether the public intercity ground transportation infrastructures and services (i.e. passenger trains and long-distance buses) can support the integration and development of urban agglomerations, we propose a method for 'transportation cluster' detection (TCD), which has three unique features: (1) the K-shortest paths are used to quantify the proximity between cities, which is more in line with people's travel behaviors; (2) a dendrogram is obtained through hierarchical clustering to reveal the structural hierarchies of transportation clusters; and (3) the integration of geo-modularity and hierarchical clustering assures high strength of division of transportation networks. The proposed TCD method was applied to the network of passenger trains, the network of long-distance buses, and the combined network of both in mainland China, respectively. By comparing the resultant transportation clusters with the urban agglomerations delineated by the Chinese government, cities that have weak transportation connections with other cities within an urban agglomeration were identified, and such findings could help devise transportation planning to better support the integrated development of urban agglomerations.

ARTICLE HISTORY
Received 3 September 2018
Accepted 2 January 2019

KEYWORDS
Transportation networks; urban agglomeration; hierarchical clustering; K-shortest paths; geo-modularity

1. Introduction

With the development and improvement of transportation infrastructures, the interconnections and interactions between cities are growing stronger. The space and time constraints, which separate geographical locations from each other, are gradually ceasing. There is no doubt that our world has been 'shrinking' (Dudás 2013). The term megalopolitan cluster area refers to a region comprising a considerable number of cities clustered around the regional economic core of one or more super-large cities (Lang and Knox 2009, He *et al*. 2013). Other terms such as megalopolis, megacity, conurbation, metropolitan interlocking region, urban cluster, and urban agglomeration have also

CONTACT Qingfeng Guan guanqf@cug.edu.cn

been used in literature to denote this large urban landscape phenomenon (Gottmann 1957, Freeman and Dickinson 1967, Yu *et al*. 2014, Fang and Yu 2017). Despite the inconsistency in terminology, this type of extensive, multi-centered, and multi-city urban landscape has been well recognized (Yu *et al*. 2014). In this paper, we use the term urban agglomeration in order to emphasize the geographic agglomeration and linkages between individual cities.

The agglomeration of cities is the result of economic agglomeration within a region. Previous studies have suggested the organizational structures of urban agglomerations largely depend on the hierarchical transportation and socioeconomic networks, which enable and strengthen the coordinated developments of population, resources, environments, societies and economies of individual cities within a region (Fang and Yu 2017). Transportation plays a crucial role in shaping a region's spatial structure and affecting its accessibility (Downs and Horner 2012). As the skeleton of regional structure, the flows of people, goods, finance, and information are the fundamental connections between cities, and demand for necessary infrastructures and services of passenger and cargo transportations (Démurger 2001). Therefore, as one of the key enabling and driving factors for the development of urban agglomerations, transportation can be seen as an important indicator to evaluate the development conditions for a region (Li *et al*. 2016). Many studies on transportation systems focused on accessibility (Straatemeier 2008, Wilbanks 2010, Zhang *et al*. 2018), improving land-use values (Hurst 1970, Wagner 2010), promoting economic developments (Vickerman *et al*. 1999, Kopits and Cropper 2005, Yamaguchi 2007), and optimizing spatial structures (Montis *et al*. 2007, Downs and Horner 2012, Garcia-López 2012). However, one issue remains less attended: whether or not the existing transportation infrastructures (e.g. railways and highways) and services (e.g. passenger trains and long-distance buses) can support the development of regional integration. Specifically, this paper focuses on the strengths of transportation connections among the cities within urban agglomerations.

Many transportation networks (e.g. road networks, airline networks, and railway networks) are proved to have the properties of complex networks, such as small-world and scale-free (Latora and Marchiori 2001, Guimera *et al*. 2005, Li and Cai 2007, Kotegawa *et al*. 2014). The advancement of complex network theory has generated an increasing body of literature on applications in transportation systems (Barrat *et al*. 2004, Montis *et al*. 2007, Bagler 2008). One of the key characteristics of complex networks that many studies have been focusing on is the community structure (also termed cluster structure), which is the division of network nodes into groups such that the connections within groups are denser than the connections between groups (Girvan and Newman 2002, Newman and Girvan 2004, Newman 2006a). Detecting the clusters of cities over transportation networks (referred to as **transportation clusters** in the rest of the text) can help us discover the groups of cities that are tightly connected through transportation, which lay the foundation of the formation and development of urban agglomerations. On the other hand, the boundary of an urban agglomeration can be determined through varied methods. Taking China as an example, the government determines the boundaries of urban agglomerations based on a set of city attributes such as population, GDP, and administrative boundaries (Yao *et al*. 2016). Although most cities in an urban agglomeration are strongly connected, weak intercity transportation connections still exist, which impede the coordinated development of the region and should be strengthened. Therefore, the comparison between transportation clusters and urban agglomerations determined by the government can be used to evaluate whether the transportation networks are capable of supporting the integration of cities within urban agglomerations. Particularly, when a transportation cluster is consistent with an urban agglomeration, the transportation connections among the cities within the urban agglomeration are strong enough to support the integrated development. On the contrary, the inconsistent areas can be used to identify the weak connections where transportation infrastructures and/or services must be strengthened.

Many approaches have been developed for finding communities over networks, and most of them fall into two general categories based on the vertex connectivity. The first category is based on graph partitioning. A typical example is the Girvan-Newman (GN) algorithm, which repeatedly removes the edges of the maximum betweenness centrality in the network (Girvan and Newman 2002). Another example is the Kernighan-Lin algorithm which based on the minimum cut or normalized cut (Kernighan and Lin 1970). The other category is based on hierarchical clustering. Examples include the fast modularity maximization algorithm that uses the modularity maximization as an optimization condition (Clauset *et al*. 2004, Newman 2006b), and the structural clustering algorithm for networks (Xu *et al*. 2007). In addition, random walks in physics have also been introduced in community detection. For example, the Markov cluster algorithm (MCL) identifies network communities by adjusting the Markov chain and inflation (van Dongen 2000). The Infomap algorithm uses random walks and information coding to find communities in a network (Rosvall and Bergstrom 2007, 2008).

When dealing with geographic networks (e.g. transportation networks), not only the topological structures, but also the geospatial characteristics of networks must be taken into account. Considering the effect of geographical distance on connection strength, Chen *et al*. (2015) introduced geographical distance into the calculation of modularity (termed *geo-modularity*), and detected the community structure of the airline network in China using the fast modularity maximization algorithm. Based on the Infomap algorithm, Xu *et al*. (2017) identified 19 communities using the traffic travel data around the Chinese New Year of 2017, and analyzed the population flows among communities. Despite the good community division results, both of the above approaches cannot represent the hierarchical structures of communities, which are critical for revealing the pattern of connection strengths among cities. In a transportation network, the most strongly connected cities form the cores of clusters, while those relatively less connected cities may be either grouped into clusters as the fringe nodes or isolated from any group. Therefore, a hierarchical structure can be found in a cluster, as the cities are grouped into the cluster at various levels of connection strength. Such a hierarchy provides information about how cities are incrementally integrated, and helps identify the weakly connected cities within a cluster.

To delineate the hierarchical structures of clusters, hierarchical clustering methods have been introduced in detecting clusters in a network. Zhou and Lipowsky (2004) proposed the Netwalk algorithm, which uses random walking of particles over the network to calculate the average distance between nodes, i.e. the average number of steps for a particle to reach the target node for the first time from the source node. The proximity index is defined based on the average distance between nodes, and a hierarchical clustering algorithm is used to iteratively merge nodes into clusters based on the proximity index, starting from the most strongly connected nodes, followed by less connected nodes. The iterative clustering stops when the modularity stops increasing, and the final communities are obtained. Through such a clustering process, a bottom-up tree is constructed to represent the hierarchical structures of clusters based on the levels of connection strength. However, the Netwalk algorithm has a major disadvantage when dealing with transportation networks. It assumes that the walk of a particle is completely random, which is not how people and cargo travel through a transportation network. In a large-scale transportation network, travel behaviors are far from completely random as most people prefer to choose the optimal route, and a small proportion may choose other less optimal routes. In other words, the proximity index between nodes in the Netwalk algorithm can hardly reflect the travel behaviors in real-world transportation networks.

To solve the aforementioned problems, this paper proposes a transportation cluster detection (TCD) approach combining K-shortest paths, hierarchical clustering and geo-modularity, which aims to meet the following criteria: (1) the proximity of transportation connection between cities must reflect people's travel behaviors; (2) the hierarchy of transportation clusters should be explicitly described to determine the densely connected cities and weakly connected cities; and (3) the final result should achieve high strength of division of a transportation network. The proposed TCD approach was applied to the network of passenger trains, the network of long-distance buses, and the combined network of both in mainland China, respectively. The resultant transportation clusters were analyzed to evaluate whether these ground transportation infrastructures and services can support the integrated development of the urban agglomerations in mainland China.

The remainder of this paper is organized as follows. The next section describes the study area and dataset. The third section presents the proposed methodological framework of TCD, followed by a report and discussion of the analytical results. Conclusions are given in the last section.

2. Study area and datasets

The study area includes all prefecture-level cities in mainland China (Hong Kong, Macao and Taiwan are excluded in this study). With continuous construction and development, the transportation networks in mainland China have reached a considerable scale. By the end of 2017, the total mileage of China's ground transportation networks had reached 4.90×10^6 km, including 1.27×10^5 km of railways and 4.77×10^6 km of highways (National Statistics Bureau of China 2017). The rapid development of transportation plays a guiding role in the spatial organization of urban agglomerations. By 5 February 2018, the State Council of China had approved eight national urban agglomerations: Triangle of Central China (TCC), Harbin-Changchun (HCC), Chengdu-Chongqing (CC), Yangtze River Delta (YRD), Central Plains (CP), Beibu Gulf (BBG), Guanzhong Plain (GZP) and Central Inner Mongolia (CIM). In addition, there are five national urban agglomerations to be approved, namely, Pearl River Delta (PRD), Beijing-Tianjin-Hebei

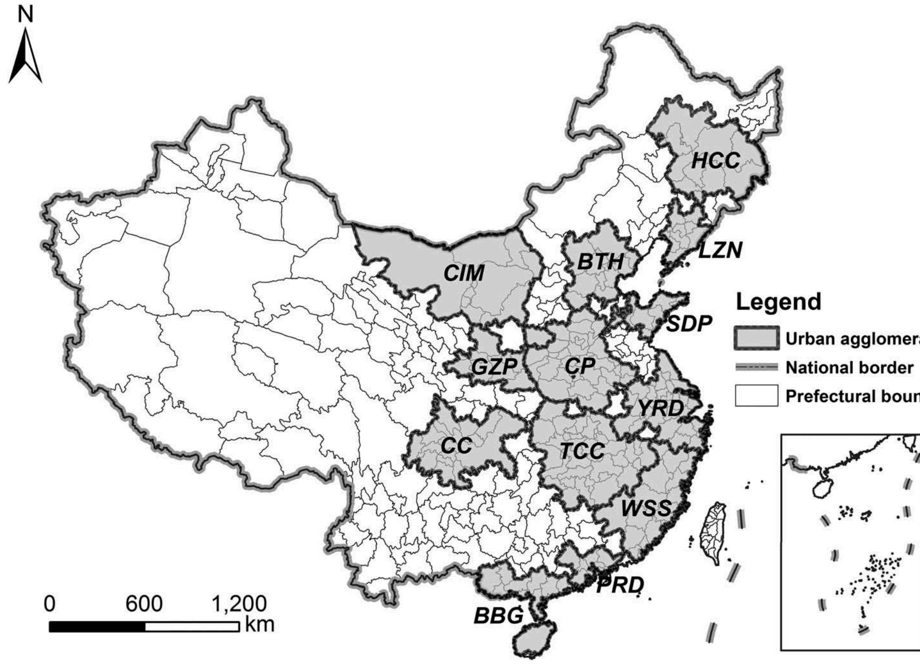

Figure 1. Thirteen urban agglomerations in mainland China designated by the Chinese government.

(BTH), Liaozhongnan (LZN), Shandong Peninsula (SDP), and West side of the Straits (WSS). The basic geographic data, such as prefecture boundaries and centers, were collected from the National Geomatics Center of China (http://ngcc.sbsm.gov.cn). All urban agglomerations are shown in Figure 1.

This study focuses on the ground transportation networks, including passenger trains and long-distance buses. All transportation data were collected during June 1 to 30, 2017. Specifically, the dataset of passenger trains in mainland China was collected from China's official train booking website (http://www.12306.cn), including 6,764 train lines (3,280 high-speed train lines and 3,484 low-speed train lines) and 2,919 railway stations. The dataset of long-distance passenger buses was collected from Ctrip (http://www.ctrip.com), one of the biggest travel booking websites in China, including 67,508 bus lines and 1,007 bus stations.

3. Methodology

A transportation cluster detection (TCD) approach combining K-shortest paths, hierarchical clustering, and geo-modularity is proposed in this study (Figure 2). The proximity index is calculated based on the K-shortest paths, which is used to quantify the strength of the connection between a pair of nodes. A hierarchical clustering algorithm is used to iteratively merge nodes into clusters according to the proximity index. In the process of clustering, the maximum of geo-modularity is used as an optimization condition to determine the final transportation clusters. The rest of this section describes the details of the process.

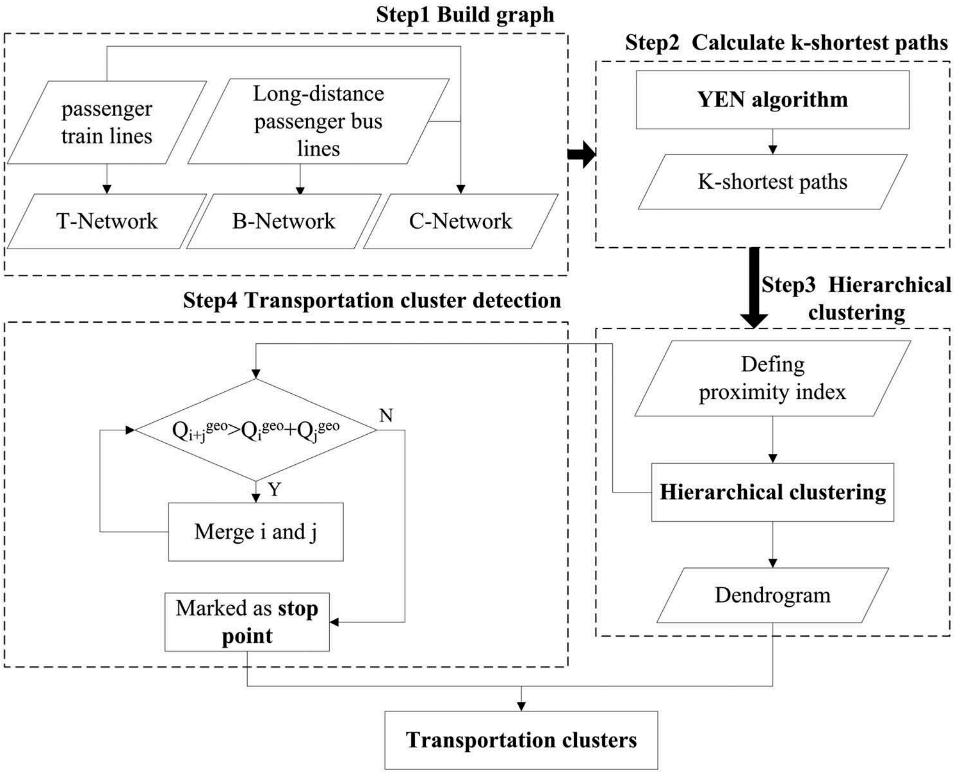

Figure 2. Flowchart of transportation cluster detection.

3.1. *Build graph*

From an original transportation dataset, a network must be constructed, consisting of nodes representing prefecture-level cities and weighted edges representing direct transportation connections between cities. For cities that have more than one station, the stations were merged as one node. Note that an edge between two nodes (cities) exists only if there is at least one transportation line directly connecting these two cities without any stop in between. In other words, two cities can be linked only if they are consecutive stops of one or more train/bus lines.

The passenger train network (denoted as *T-Network*) was constructed from the original dataset that includes both high-speed trains (denoted as *HS*) and low-speed trains (denoted as *LS*). For each type, the weight (w) of the edge between two cities is proportional to the number of train lines (m) connecting these two cities, and inversely proportional to the n-th power of the average traveling time of those trains (t). That is, the more commutes and the shorter the traveling time, the stronger is the connection between two cities. The total weight (W_T) between two cities is the sum of the weight of high-speed trains (W_{HS}) and the weight of low-speed trains (W_{LS}).

$$W_{HS} = m_{HS}/t_{HS}^{n} \tag{1}$$

$$W_{LS} = m_{LS}/t_{LS}^{n} \tag{2}$$

$$W_T = W_{HS} + W_{LS} \tag{3}$$

The traveling time (*t*) is used as the 'distance' between two cities, and the power (*n*) serves as the distance friction coefficient as in the gravity model. A large value of *n* leads to a strong distance decay effect (Liu *et al*. 2014). In this study, referring to Chen et al.'s study (2015), *n* in Equations (1) and (2) was set to be 2.

As the result, the *T-Network* includes 323 nodes and 999 weighted edges. Through the similar process, the bus network (*B-Network*) was constructed, including 356 nodes and 8,777 weighted edges.

Also, a combined network (*C-Network*) of *B-Network* and *T-Network* was constructed. The weight W_C of an edge is calculated as:

$$W_C = \begin{cases} W_T + \mu W_B & Both\ train\ lines\ and\ long-distance\ bus\ lines\ between\ two\ cities \\ W_T & Only\ train\ lines\ between\ two\ cities \\ \mu W_B & Only\ long-distance\ buslines\ between\ two\ cities \\ 0 & No\ lines\ between\ two\ cities \end{cases} \tag{4}$$

where W_T and W_B are the weights of edges in *T-Network* and *B-Network*, respectively. A ratio (μ) is be used to represent the relative importance between train lines and bus lines when combining their weights. Such a ratio may vary across city-pairs, as the relative importance between trains and buses may vary for different pairs of cities. In this study, due to the limitation of data, it is arduous to obtain the ratios for all city-pairs in mainland China. Therefore, the edge weights of the combined network were computed using a universal ratio. According to the reports from the National Bureau of Statistics of China in 2015, 2016, and 2017, the ratio between the number of passengers over long-distance buses and that over trains is about 5.12, which was used to combine the weights of *B-Network* and *T-Network* (i.e. $\mu = 5.12$ for the whole study area). As the result, the *C-Network* includes 356 nodes and 8,913 edges (Figure 3).

3.2. *Calculate K-shortest paths*

Before detecting the clusters over a transportation network, the proximity index must be derived to represent the strength of transportation connection between any pair of nodes (i.e. cities) in the network (possibly connected by chains of edges). In order to derive the proximity index that reflects people's travel behaviors, TCD uses the K-shortest paths (KSP). The purpose of the K-shortest paths is to find multiple alternative paths with various costs (ordered ascendingly) between the source and the destination in a network to satisfy the user's selection of different paths to the greatest extent (Eppstein 1994, Aljazzar and Leue 2011).

As the example shown in Figure 4, there are three routes from A to B (i.e. A-B, A-C-B, and A-D-E-B), and the traveling cost (time, distance, or payment) is 4, 5, and 6, respectively. Most people would choose route AB to save cost, but there are also a small number of people who would take a detour and choose route ACB or ADEB for special reasons. That is to say, for the choice of traveling route between any pair of cities, the optimal route usually has the greatest probability while the sub-optimal routes have smaller probabilities. Therefore, when quantifying the proximity of two cities, we must

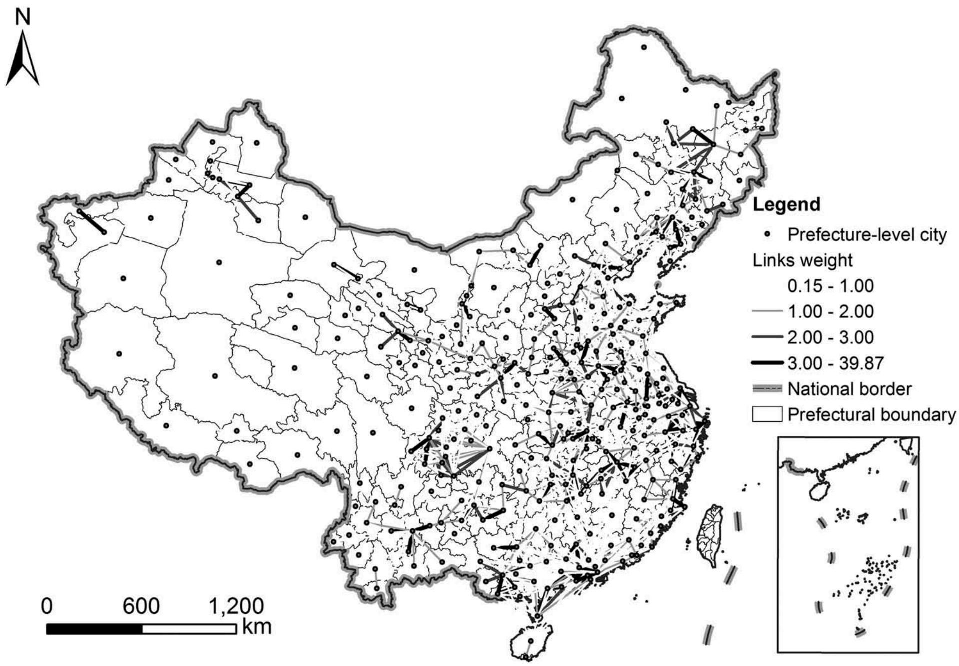

Figure 3. The combined ground transportation network (C-network) (For display clarity, the edges with a weight less than 0.15 are not shown).

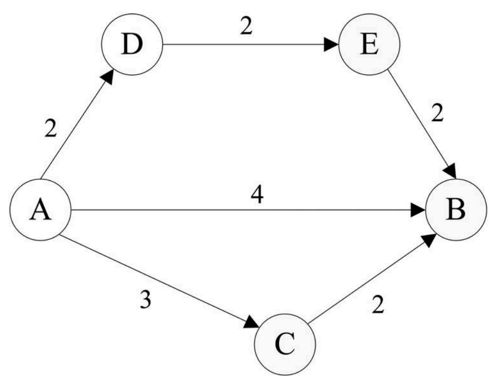

Figure 4. K-shortest paths from A to B (K = 3). The value on an edge indicates the cost between the two linked nodes.

consider not only the optimal route, but also other routes. KSP can provide such a collection of alternative routes with various connection strengths. Many algorithms have been developed to solve the KSP problem, and more detailed information can be found in the review paper by Brander and Sinclair (1996).

TCD uses the YEN algorithm (Yen 1971) to generate the KSP for all pairs of nodes in a transportation network. Supposing there are N nodes and M edges in the network, the computational complexity of calculating K shortest paths between a pair of nodes is $O(KN(M + NlogN))$(Eppstein 1994). The YEN algorithm assumes that there are

K paths with different lengths between a pair of nodes, sorted by length from small to large. The k-th shortest path is d_k, which contains a series of edges, and the total length is l_k. The core idea of the YEN algorithm is to use the shortest paths of d_1, d_2, …, d_{k-1} $(l_1 \leq l_2 \leq \ldots \leq l_{k-1})$ that have been obtained to generate d_k. The YEN algorithm can be broken down into two parts: determining the shortest path d_1, and then determining all other $k-1$ shortest paths. The Dijkstra's algorithm is used to determine d_1 in this study (Dijkstra 1959). To find d_k, the algorithm assumes that d_1 to d_{k-1} have already been found. Taking find d_2 as an example, suppose d_1 is composed of nodes (n_1, n_2, … n_i…n_j). Next, the edge between n_i and n_{i+1} ($1 \leq i<j$) are sequentially removed, and the deviation path d_2^i between n_1 and n_j is calculated again by Dijkstra's algorithm. Then a deviation path sets $\boldsymbol{S}$ is formed, which contains $j-2$ deviation paths. Finally, the path with the smallest length in $\boldsymbol{S}$ is selected as d_2. By repeating the above steps, other paths ($k > 2$) can also be determined.

It should be noted that in TCD, the 'length' of an edge (l) is the reciprocal of the weight W (i.e. $l = \frac{1}{W}$). That is to say, the larger is the weight of an edge (i.e. the stronger is the connection between the two linked nodes), the shorter is the 'length' of the edge.

3.3. *Hierarchical clustering*

Once the KSP between all pairs of nodes in the transportation network are determined, the proximity indices between these pairs are calculated and a hierarchical clustering process is applied to incrementally group nodes into clusters.

As mentioned above, when quantifying the proximity between a pair of nodes (cities), it is necessary to take people's travel behaviors into account. Therefore, the proximity index $\boldsymbol{\Phi}$ between node i and node j is defined as a weighted combination of K shortest paths(l_k), and the weight of the k-th shortest path $\boldsymbol{\omega_k}$ decreases as k increases. The lower is the proximity index, the stronger is the transportation connection between a pair of nodes.

$$\boldsymbol{\Phi}(i,j) = \sum_{k=1}^{K} \boldsymbol{\omega}_k l_k \tag{5}$$

where K is the number of paths generated by KSP. The value of K and the values of $\boldsymbol{\omega_k}$ must be set carefully, such that the following hierarchical clustering can generate a good division of the nodes in a network. Please see section 4.1 for an example.

For a network with N nodes, the nodes are initially divided into N clusters, each containing a single node, and the proximity index between all pairs of clusters are calculated. In TCD, a hierarchical clustering process is used to iteratively merge clusters of nodes, and the proximity index between cluster α and β is updated as follows:

$$\boldsymbol{\Phi}(\alpha,\beta) = \frac{1}{n_{\alpha,\beta}} \sum\nolimits_{(i,j):i\in\alpha,j\in\beta} \boldsymbol{\Phi}(i,j) \tag{6}$$

where $n_{\alpha,\beta}$ is the number of node pairs between cluster α and β, $\boldsymbol{\Phi}(i,j)$ is given by Equation (5). At each iteration, the hierarchical clustering merges the pair of clusters with the minimum proximity index into a single cluster, and then updates the proximity indices between this new cluster and all remaining clusters. Finally, a dendrogram (i.e.

a cluster tree) is formed to represent the bottom-up hierarchy of clusters of nodes. The hierarchical clustering approach used in TCD is the same as the full-order constrained average linkage clustering (Full-Order-ALK) method proposed by Guo (2008), and the computational complexity is $O(N^2 logN)$ with N nodes in the network. As the example shown in Figure 5, cluster α contains node 1 and 2, cluster β contains node 3 and 4, and cluster γ contains node 5. When cluster α and β are merged into $\alpha + \beta$, the proximity index between cluster $\alpha + \beta$ and γ is calculated as follows:

$$\boldsymbol{\Phi}_{(\alpha+\beta,\gamma)} = \frac{\boldsymbol{\Phi}_{(1,5)}+\boldsymbol{\Phi}_{(2,5)}+\boldsymbol{\Phi}_{(3,5)}+\boldsymbol{\Phi}_{(4,5)}}{4} \tag{7}$$

3.4. *Transportation cluster detection*

The modularity has been used in many studies to evaluate the division result of communities (clusters). As a quantitative index, the modularity computes the difference between the number of edges within communities and the expected number. A good partition of a network should result in a significantly greater number of edges within communities than expected. The mathematical expression of modularity is as follows (Newman and Girvan 2004):

$$Q = \sum_i \left[e_{ii} - a_i^2\right] \tag{8}$$

where e_{ii} is the fraction of edges in the network that connect nodes in the same community i, and a_i is the fraction of edges that are attached to the nodes in community i.

A large body of literature has been generated on the evaluation of community detection using modularity (Gustafsson *et al*. 2006, Newman 2006a, Schwarz *et al*. 2008, Dinh and Thai 2013). Many geographic networks are spatially constrained, in which the nodes and edges are embedded in space (Barthélemy 2011). To quantify the effect of distance on the strength of the connection between nodes, Chen *et al*. (2015) proposed the geo-modularity (Q^{geo}), which adds a weight of distance and connection frequency to edges when calculating modularity.

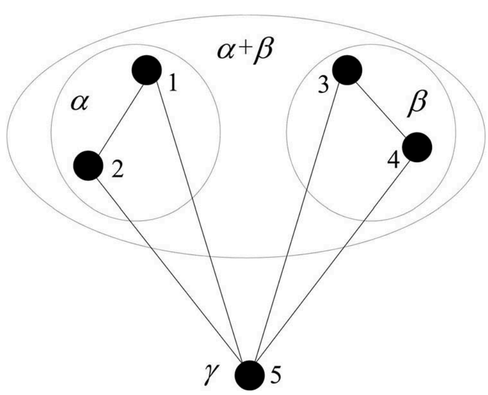

Figure 5. An example of cluster merging and proximity index updating.

$$Q^{geo} = \sum_i \left[e_{ii}^{geo} - \left(a_i^{geo}\right)^2 \right] = \sum_i \left[\frac{\sum_{u,v \in i} w_{uv}}{2W_{sum}} - \left(\frac{\sum_{u \in i} w_{uv}}{2W_{sum}} \right)^2 \right] \tag{9}$$

where w_{uv} is the weight of the edge between nodes u and v, and is given by Equation (1), W_{sum} is the sum of edge weights in the network.

In TCD, geo-modularity is used to evaluate the quality of transportation cluster detection, and as the optimization condition of hierarchical clustering. In the hierarchical clustering process (step 3, Section 3.3), the Q^{geo} of each cluster is calculated at each iteration. Taking cluster C_1 and cluster C_2 as an example, if $Q^{geo}_{merge} > Q^{geo}_{c_1} + Q^{geo}_{c_2}$, then C_1 and C_2 are merged. Otherwise, the cluster C_1 and cluster C_2 are marked as 'stop points'. After the cluster dendrogram has been constructed, each branch of this dendrogram is backtracked from the top level, until a cluster with a 'stop-point' mark is encountered. This cluster is then considered to represent an elementary module of the network. As a result of this backtrack process, the network is partitioned into a set of mutually exclusive elementary modules $a_1, a_2 \ldots a_n$, such that the total geo-modularity $Q^{geo} = \sum_{a_i} Q^{geo}_{a_i}$ achieves its global maximum value. The integration of Q^{geo} in the hierarchical clustering helps generate clusters with high strength of division.

Taking a simple transportation network ***G*** as an example (Figure 6), which contains 6 nodes and 7 weighted edges.

The transportation cluster detection process is as follows:

(1) The 6 nodes are initialized into 6 clusters, then the proximity index between each pair of clusters ($K = 2$, $\omega_1 = 0.95$, and $\omega_2 = 0.05$), and the geo-modularity of each cluster is calculated.
(2) The two clusters with the smallest proximity index are merged into a new cluster. In the example, cluster 1 and 2 are merged into a new cluster, cluster 7.
(3) The proximity indices between cluster *7* and all other clusters are updated.
(4) Steps (2) and (3) are repeated until all nodes are merged. The order in which nodes are merged is (1, 2), (3, 7), (4, 5), (6, 9), (8, 10). Finally, a dendrogram is formed, as shown in Figure 7.
(5) In step (2), only merging clusters 8 and 10 will result in the value of the merged geo-modularity being less than the value before the merge. That is, $Q^{geo}_{merge} < Q^{geo}_8 + Q^{geo}_{10}$ (0.0 < 0.209 +0.209). Therefore, clusters 8 and 10 are marked as 'stop points'. Finally, two clusters are generated. One of them contains nodes 1, 2, and 3, and the other contains nodes 4, 5, and 6 (Figure 8).

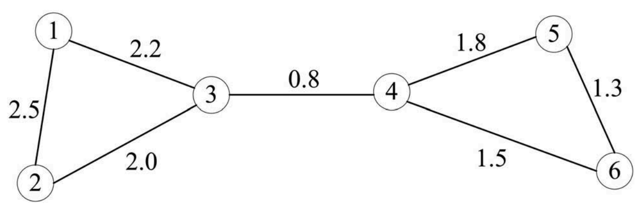

Figure 6. A simple schematic diagram of network G (The value on an edge represents the weight of the edge, and the greater the weight, the stronger is the connection between the nodes).

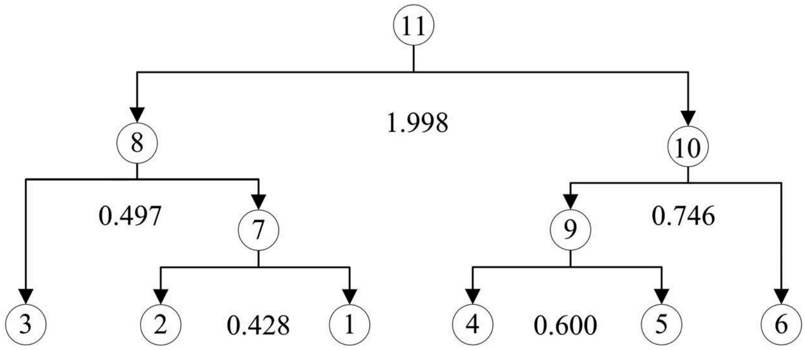

Figure 7. A dendrogram showing the cluster structure of G (The value between two branches represents the proximity index of the two clusters).

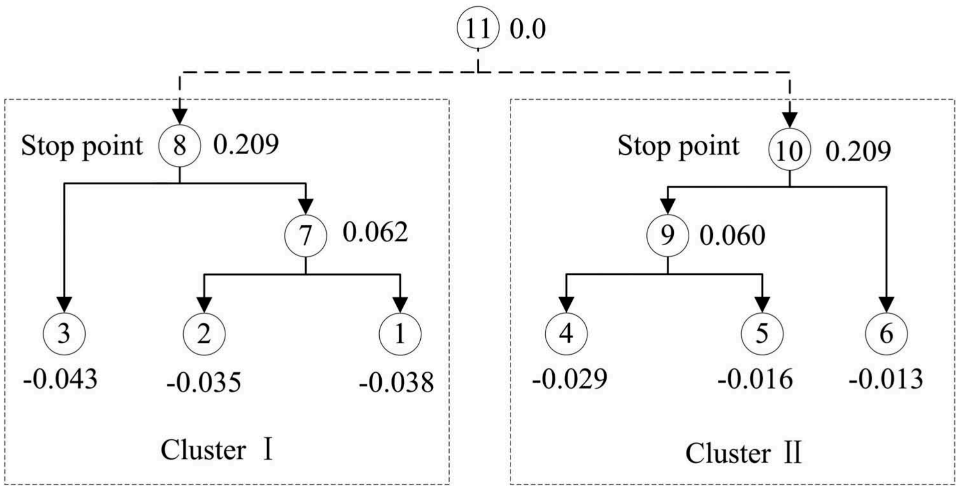

Figure 8. The TCD results of G (The value next to the node represents the value of the geo-modularity of each cluster).

4. Results

The algorithms in TCD, including YEN algorithm, hierarchical clustering, and the calculations of proximity index and geo-modularity, were implemented using Python. The TCD was then applied on the *T-Network, B-Network* and *C-Network*, respectively, and the resultant transportation clusters were analyzed. Particularly, the transportation clusters of *C-Network* were compared with the urban agglomerations delineated by the Chinese government to evaluate whether the ground transportation infrastructures and services can support the integrated developments of these urban agglomerations.

4.1. *Hierarchical structures of clusters*

As mentioned before, for the K-shortest paths and proximity index between a pair of nodes, the value of *K* (i.e. the number of alternative paths) and the values of $\boldsymbol{\omega_k}$ (i.e. the weights for combining the 'lengths' of paths) must be set carefully to achieve a good division of the nodes. In our experiments, for each transportation network, a series of trials were conducted with various values of K (from 1 to 3, step = 1) and $\boldsymbol{\omega_k}$ (from 0.00 to 1.00, step = 0.01), and the setting that generated the maximum geo-modularity was chosen.

For the *T-Network*, when $K = 1$, and $\boldsymbol{\omega}_1 = 1.0$, Q^{geo} reached a high value. Based on the weighted combination of KSP, the proximity index Φ between any two cities was calculated. Through the hierarchical clustering, a dendrogram was generated which showed the hierarchical cluster structure of the *T-Network* (Figure 9). At each proximity level, cities were grouped into multiple clusters. Small clusters were incrementally integrated into larger clusters as the inter-cluster proximity index increased.

In order to observe the trend of the merger of cities with the increase of proximity index, the number of clusters and the corresponding proximity index were fitted into a curve, which was found to conform to the power law distribution (Figure 10). When the inter-cluster proximity index is less than 10, the curve drops quickly, and most of the cities are grouped into the railway transportation clusters within this range. However, a small number of cities do not enter any cluster until the inter-cluster proximity index is greater than 30, such as Aksu, Bazhong, Yining, Holingola, Alxa League, Daxinganling, Hotan, Shigatse, and Qitaihe. It indicates that these cities have relatively weak railway transportation connections with other cities.

To further explore the hierarchical structure of *T-Network* clusters, we selected some inflection points in the curve in Figure 10. The values of the inter-cluster proximity index at these points are 0.11, 0.60, 2.04, 10.80 and 30.77, corresponding to the number of clusters being 290, 200, 150, 80 and 40, respectively (Figure 10). Taking three typical urban agglomerations in mainland China as examples: Yangtze River Delta (YRD), Pearl River Delta (PRD), and Beijing-Tianjin-Hebei (BTH), the results are shown in Figure 11. In the YRD region, three small clusters formed at the beginning of clustering, each containing a small number of cities that are most tightly connected through passenger trains (e.g. Shanghai, Nanjing and

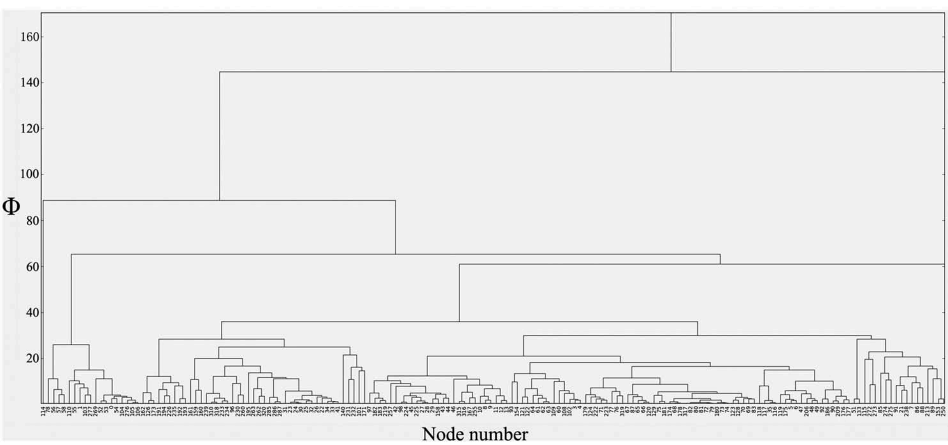

Figure 9. Transportation cluster structure of T-Network (only part of the dendrogram is showed for clarity).

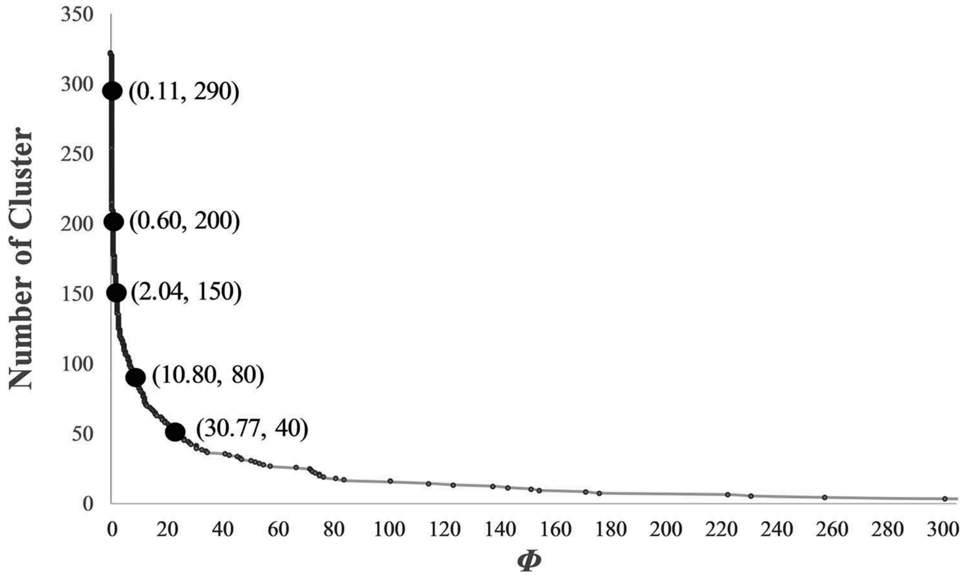

Figure 10. The curve between the minimum inter-cluster proximity index and the number of T-Network clusters.

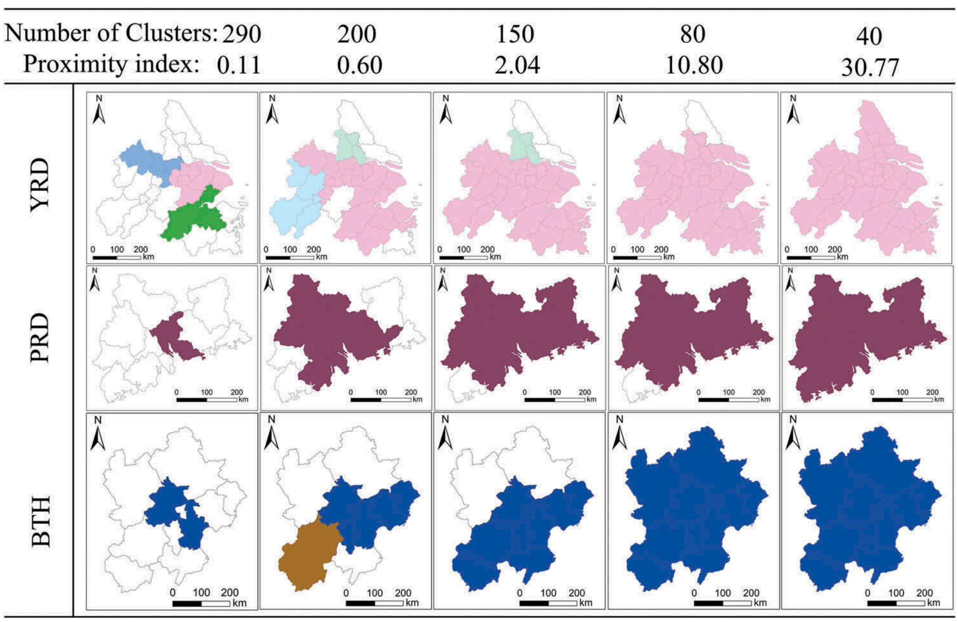

Figure 11. The clustering process of T-Network.

Hangzhou). These clusters can be seen as the railway cores in YRD. As the inter-cluster proximity index increased, these core clusters were merged together, and other cities (fringe cities) were also gradually merged into a big regional cluster. The southern cities entered the regional cluster before northern cities, indicating the railway connections between the southern cities and the cores are stronger than those of the northern cities. In PRD, Guangzhou and Shenzhen formed the core transportation cluster. With the increase of the inter-cluster proximity index, the northern cities were merged into the regional cluster, followed by the southwestern cities. In BTH, Beijing and Tianjin formed the core transportation cluster. With the increase of the inter-cluster proximity index, the southern cities were merged into the regional cluster, and the northern cities entered the regional cluster later on.

The same method was applied to the *B-Network*. When K = 2, $\boldsymbol{\omega}_1$ = 0.80, and $\boldsymbol{\omega}_2$ = 0.20, the Q^{geo} reached a high value. Using YRD, PRD, and BTH as examples, the hierarchical structures of transportation clusters over the *B-Network* are shown in Figure 12. In YRD, three core clusters formed in the beginning, and grew in parallel before they were merged together. Two cities in the south, Taizhou and Jinhua, formed their own cluster in the middle of clustering, and did not enter the big regional cluster until the inter-cluster proximity index increased to 0.54. In the PRD region, cities are strongly connected through long-distance buses, and formed a big regional cluster in the early stage of clustering. BTH has a single core cluster centered in Beijing, which gradually merged other cities throughout the clustering and grew into a big regional cluster. Compared with the *T-Network*, the cities within these three regions are more strongly connected over the *B-Network*, indicated by the fact that the regional clusters over the *B-Network* formed earlier than those over the *T-Network*.

For the *C-Network*, when K = 2 and $\boldsymbol{\omega}_1$ = 0.99, and $\boldsymbol{\omega}_2$ = 0.01, Q^{geo} reached a high value. The clustering process of the cities in YRD, PRD and BTH are shown in Figure 13. In YRD, two core clusters centered in Shanghai and Nanjing, and three small clusters formed in the early stage of clustering. They gradually merged together into one regional cluster. The northern and southern cities (e.g. Yancheng and Jinhua) entered the regional cluster in the late stage of clustering. In PRD, three core clusters formed in

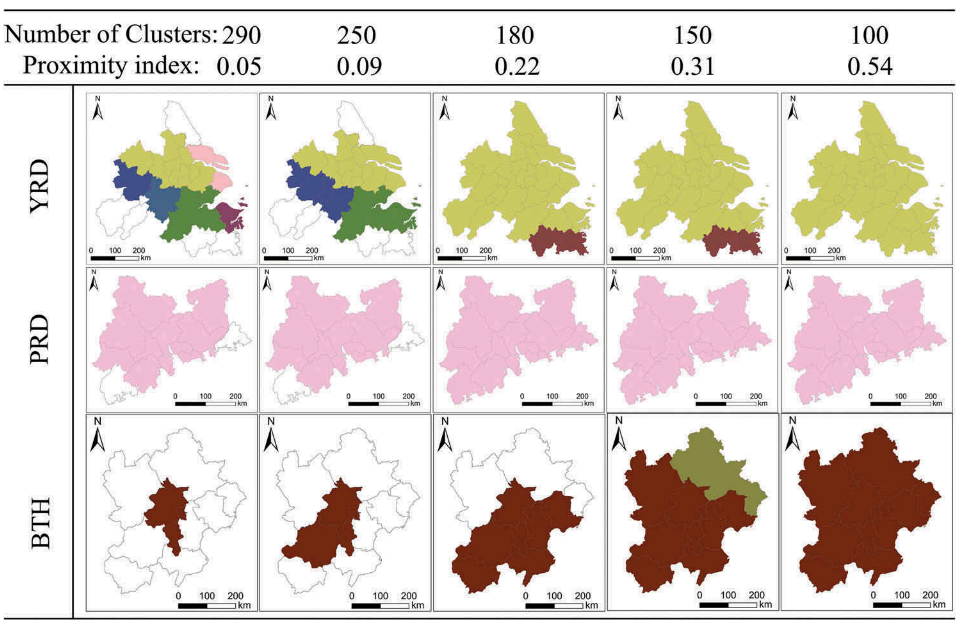

Figure 12. The clustering process of B-Network.

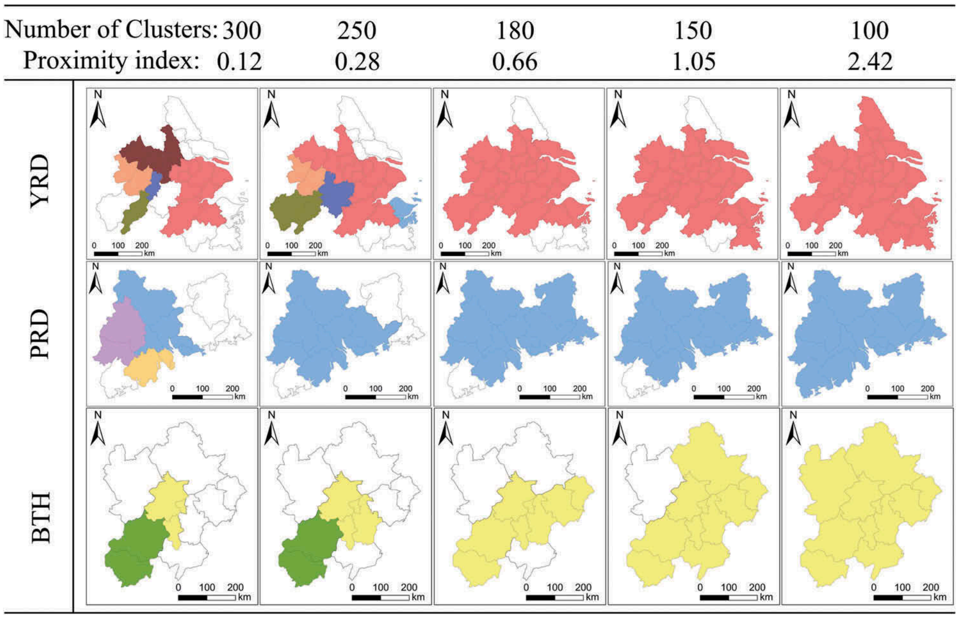

Figure 13. The clustering process of C-Network.

the beginning, centered in Guangzhou, Zhuhai, and Zhaoqing, and quickly merged into a single regional cluster. The southwestern part (i.e. Yangjiang) entered the regional cluster in the late stage of clustering. In BTH, two core clusters formed in the beginning, centered in Beijing and Shijiazhuang, and then merged together. Other cities entered the regional cluster, northeastern ones earlier than western ones.

4.2. *Division of transportation clusters*

According to step 4 of the method (Section 3.4), with the maximum value of Q^{geo} as the stopping condition for hierarchical clustering, the final transportation clusters over the transportation networks were determined. We compared our method with Netwalk (Zhou and Lipowsky 2004), Infomap (Rosvall and Bergstrom 2007, 2008) and G-N algorithm (Girvan and Newman 2002). The results showed that the proposed approach generated higher Q^{geo} than other algorithms for all three networks (Table 1), indicating TCD is capable of generating strong divisions of transportation networks.

This section focuses on the clusters over the *C-Network*, as the *C-Network* includes both passenger train lines and long-distance bus lines. As shown in Figure 14, a total of 31 clusters over the *C-Network* were detected in mainland China.

It can be observed that the average proximity indices of cities in the clusters on the east side of the Huhuanyong Line (also known as the 'Heihe-Tengchong Line') are all less than 3, and those on the west side are all greater than 3, indicating that the transportation connections in the western China are far weaker than those in the eastern China. There are two continuous regions containing several isolated cities that do not belong to any cluster in western and northern China, because of their extremely sparse transportation links with other cities. The Hu Line was first proposed by Huanyong Hu in 1935

Table 1. Q^{geo} by different approaches.

Q^{geo}

Algorithm	*T-Network*	*B-Network*	*C-Network*
Netwalk	0.7660	0.7884	0.7882
Infomap	0.7523	0.7300	0.7756
G-N algorithm	0.6131	0.1473	0.1592
TCD	**0.7805(K =1)**	**0.7917(K =2)**	**0.8025(K =2)**

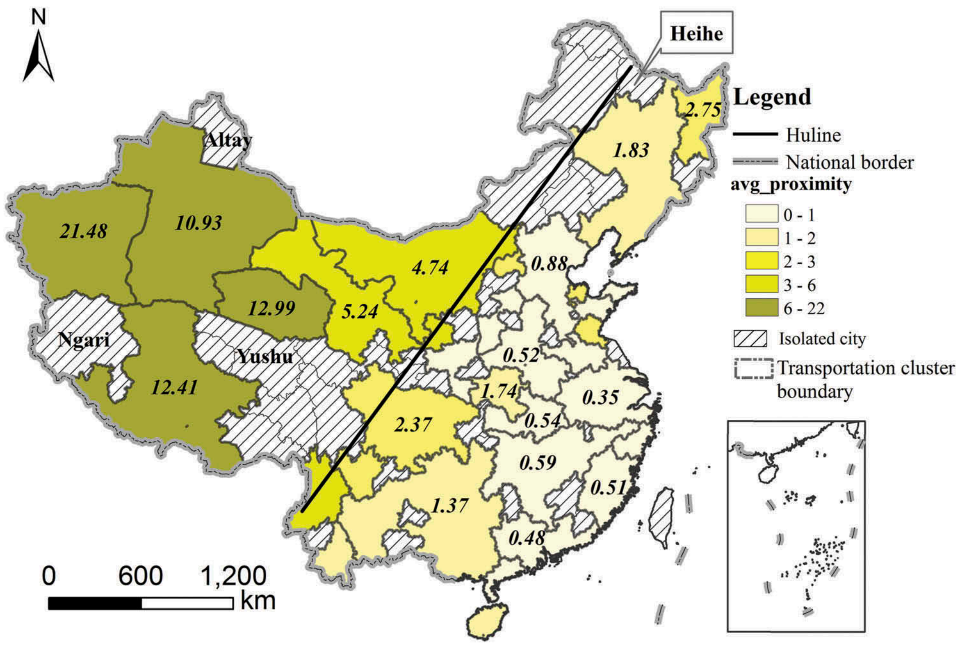

Figure 14. Transportation clusters over C-Network (the numbers in the clusters are the average proximity index of cities in the cluster, and small clusters were not labeled for clarity; the slashed area are isolated cities that do not belong to any cluster, only a few isolated cities were marked for clarity).

to describe the significant difference of population distribution in China (Hu 1935), and has also been widely used to describe the division of urbanization and economic development in China. Our analysis reveals that the ground transportation infrastructures and services in the western China are also greatly weaker than those in the east, which should be closely related with the less amount and density of population, slower urbanization and economic development, and more complex topographic conditions in the west.

To evaluate whether the transportation infrastructures and services can support the integrated developments of urban agglomerations, we overlaid the resultant transportation clusters with the urban agglomerations delineated by the Chinese government (Figure 15). The following phenomena can be observed:

(1) The transportation clusters in the PRD, BTH, CIM, CC and LZN urban agglomerations expand to the surrounding cities. In addition, the BTH and SDP urban agglomerations are merged into one large transportation cluster, same as LZN and most cities within HCC. This indicates that the cities within these regions are strongly connected through ground transportations. The transportation integrations have gone beyond the inter-city level and reached the inter-agglomeration level.
(2) There are some isolated cities that do not belong to any cluster in the HCC, GZP, CP, YRD, TCC and WSS urban agglomerations. As discussed above, this indicates that those cities have weaker transportation connections with other cities in their corresponding regions. The cities in the eastern, western and southern parts of BBG form three transportation clusters, which indicates that the transportation connections between the eastern and western parts, the southern and northern parts of BBG are still weak.
(3) In this study, the Wuhan metropolitan area (WMA), the Changsha-Zhuzhou-Xiangtan city group (CZT), and the Poyang Lake city group (PL) were considered as a single urban agglomeration. The Chinese government has merged these three urban agglomerations into a large national urban agglomeration, called the Triangle of Central China (TCC). As shown in Figure 16, three transportation clusters were detected within TCC. The northern cluster covers WMA, while the southern cluster covers CZT and PL. This indicates that the transportation

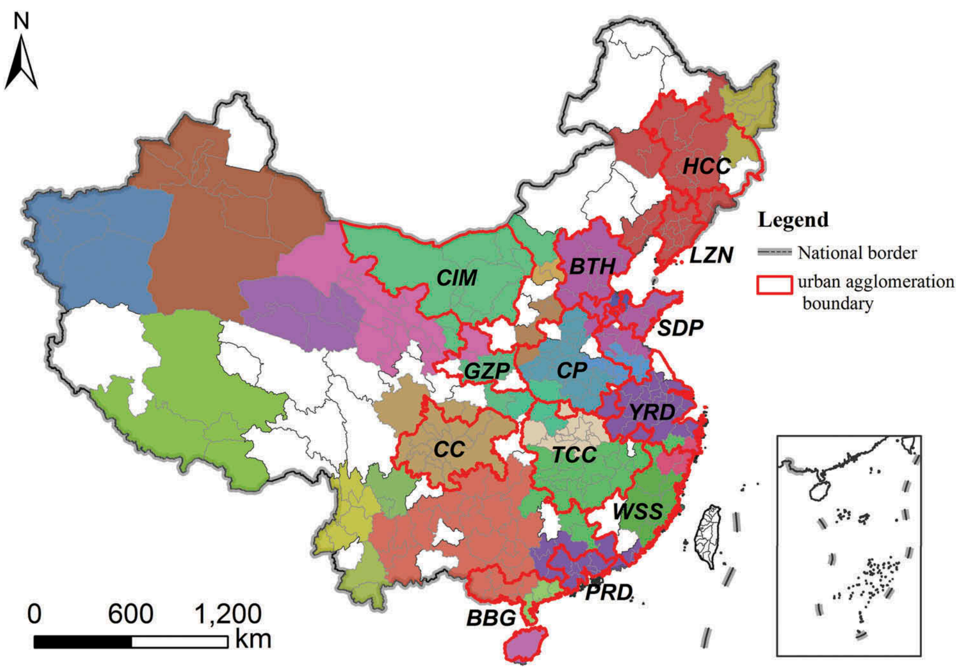

Figure 15. The overlay of the transportation clusters over C-Network and the urban agglomerations in mainland China.

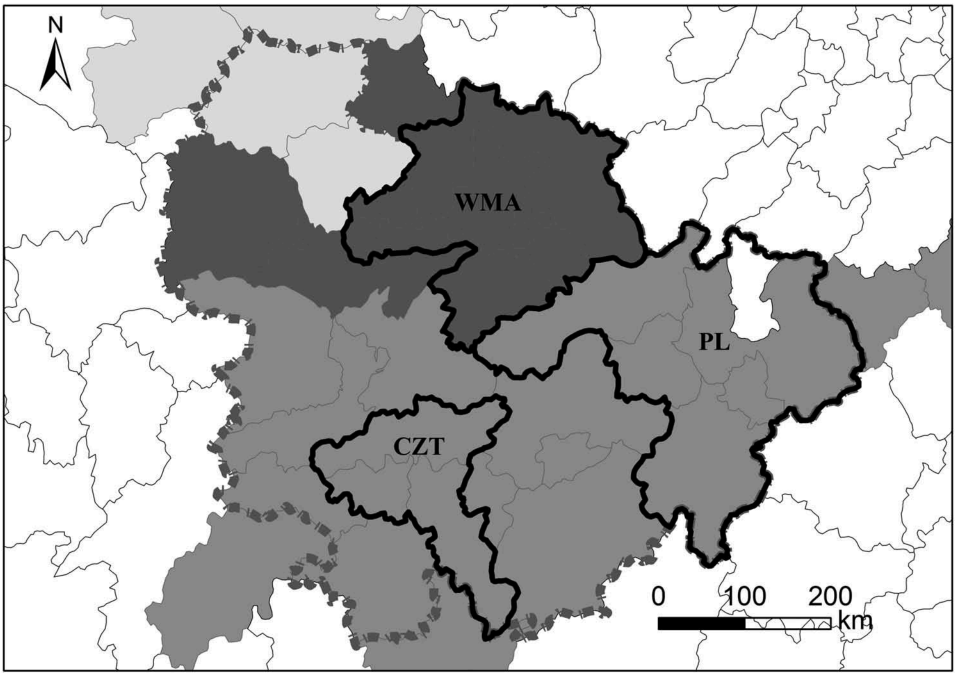

Figure 16. The three transportation clusters covering the TCC urban agglomeration(the gray dotted line represents the border of TCC; and the black solid lines represent the borders of WMA, CZT and PL).

connections between the southern and northern parts of TCC are still weak and cannot support the integrated development of the whole agglomeration.

In general, each of the 13 urban agglomerations designated by the Chinese government has at least one transportation cluster, and the core cities of each urban agglomeration are strongly connected as expected. However, weak connections were found within some urban agglomerations, especially indicated by the isolated cities. To support the integrated developments of those urban agglomerations, it is recommended that the transportation infrastructures (e.g. railways and highways) and services (e.g. passenger train lines and long-distance bus lines) between the isolated cities and other cities should be improved. Taking WSS as an example, in order to connect isolated cities and to integrate multiple dispersed small transportation clusters into one large transportation cluster, new transportation infrastructures and/or services along the northeast-southwest direction should be added.

In addition, the average proximity indices in HCC, LZN, CC, and CIM are much larger than those of other urban agglomerations, indicating relatively weaker transportation connections within these urban agglomerations. Therefore, transportation improvements are also recommended in these regions to enhance the convenience and efficiency of intercity flows, which is conducive to the integration and development of urban agglomerations.

5. Conclusion

The clusters of cities over transportation networks (i.e. transportation clusters) can be detected based on the strengths of transportation connections between cities, and can be used to evaluate whether the existing transportation infrastructures and services can support the integrated development of regions. This study proposes a novel approach for transportation cluster detection (TCD). The main contributions are as follows: (1) the K-shortest paths are used to quantify the proximity (i.e. connection strength) between cities, which is more in line with people's travel behaviors; (2) a dendrogram is obtained through the hierarchical clustering to reveal the structural hierarchies of transportation clusters; and (3) the integration of geo-modularity and hierarchical clustering assures high strength of division of a transportation network.

The proposed TCD approach was applied to the network of passenger trains (*T-Network*), the network of long-distance buses (*B-Network*), and the combined network of both (*C-Network*) in mainland China. The results showed that TCD outperformed some existing approaches for community detection over networks, by achieving the highest geo-modularity in all three transportation networks.

Taking the YRD, PRD, and BTH urban agglomerations as examples, we explored the bottom-up clustering process of cities and the hierarchical structures of these transportation clusters. The core cities and fringe cities of the clusters were explicitly identified according to the proximity levels, at which the cities were merged into the regional clusters.

The final resultant transportation clusters of the *C-Network* were further analyzed. A significant difference in the average proximity index was found between the clusters on the east side and the clusters on the west side of the Hu Line, indicating the imbalance of ground transportation distribution in mainland China. Several isolated cities that failed to be merged into any cluster were found because of their extremely sparse transportation connections with other cities. By overlaying the transportation clusters with the urban agglomerations delineated by the Chinese government, well connected and integrated agglomerations were identified, while some agglomerations were split into multiple transportation clusters and/or isolated cities. Those weakly connected regions require the strengthening of transportation infrastructures and services in order to improve the integration of the cities and support the collaborative developments.

The limitations of this study and our future work are as follows:

(1) This study evaluated the public ground transportation infrastructures and services for urban agglomerations, but did not consider the actual flows between cities. Therefore, our future work aims to integrate various flow data (e.g. human flow, cargo flow, financial flow, and information flow) into the cluster detection to analyze the socioeconomic connections between cities.

(2) Due to the limitation of data, a universal ratio (i.e. $\mu = 5.12$) was used in this study to combine the weights of *B-Network* and *T-Network* for the edge weights in *C-network*, which can hardly reflect the variance of the relative importance between railway and bus transportations. Data of passenger flows via trains/buses can provide the key basis for determining the specific ratio for each pair of cities.

(3) As discussed in Section 3, the computational complexity of the YEN algorithm to generate K short paths between a pair of nodes is $O(KN(M + NlogN))$, and there are $(N(N-1)/2)$ pairs of nodes in a network. The computational performance of TCD mainly depends on the computational complexity of the KSP algorithm in the network. A KSP algorithm with higher performance (either through the optimization of the algorithm itself or through the parallel implementation) can greatly improve TCD's capability of dealing with large networks.

Acknowledgments

The authors thank the editor and anonymous reviewers for their constructive comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Key R&D Program of China [2017YFB0503800]; National Natural Science Foundation of China [41671408, 41801306]; Natural Science Foundation of Hubei Province [2017CFA041].

Notes on contributors

Hanqiu Yue is a PhD Candidate in the School of Geography and Information Engineering at China University of Geosciences. His research focuses on graph learning, complex network analysis and urban computing.

Qingfeng Guan received his Ph.D. degree from the Department of Geography, University of California, Santa Barbara in 2008. He is currently a professor in the School of Geography and Information Engineering at China University of Geosciences. His research interests include spatio-temporal data mining and modeling, spatial computational intelligence and high-performance spatial computing.

Yongting Pan is a PhD Candidate in the School of Geography and Information Engineering at China University of Geosciences. Her research focuses on spatio-temporal data mining, and urban computing.

Lirong Chen is a PhD Candidate in the School of Geography and Information Engineering at China University of Geosciences. Her research focuses on spatio-temporal data mining and modeling.

Jianjun Lv is currently a professor in the School of Geography and Information Engineering at China University of Geosciences. His research focuses on smart city, geospatial calculation and analysis.

Yao Yao is currently an associate professor in the School of Geography and Information Engineering at China University of Geosciences. His research focuses on spatio-temporal data mining, urban computing and social sensing.

ORCID

Hanqiu Yue http://orcid.org/0000-0002-5088-2633
Qingfeng Guan http://orcid.org/0000-0002-7392-3709
Yongting Pan http://orcid.org/0000-0002-7726-9302
Lirong Chen http://orcid.org/0000-0002-7302-5458
Jianjun Lv http://orcid.org/0000-0002-8144-1929
Yao Yao http://orcid.org/0000-0002-2830-0377

References

Aljazzar, H. and Leue, S., 2011. K*: a heuristic search algorithm for finding the k shortest paths. *Artificial Intelligence*, 175, 2129–2154. doi:10.1016/j.artint.2011.07.003

Bagler, G., 2008. Analysis of the airport network of India as a complex weighted network. *Physica A: Statistical Mechanics and Its Applications*, 387 (12), 2972–2980. doi:10.1016/j.physa.2008.01.077

Barrat, A., *et al.*, 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101, 3747–3752. doi:10.1073/pnas.0400087101

Barthélemy, M., 2011. Spatial networks. *Physics Reports*, 499, 1–101. doi:10.1016/j.physrep.2010.11.002

Brander, A.W. and Sinclair, M.C., 1996. A comparative study of K-shortest path algorithms. *In*: M. Merabti, M. Carew, and F. Ball, eds. *Performance engineering of computer and telecommunications systems*. London: Springer London, 370–379.

Chen, Y., *et al.*, 2015. Finding community structure in spatially constrained complex networks. *International Journal of Geographical Information Science*, 29 (6), 889–911. doi:10.1080/13658816.2014.999244

Clauset, A., *et al.*, 2004. Finding community structure in very large networks. *Physical Review E*, 70 (6), 1–6. doi:10.1103/PhysRevE.70.066111

Démurger, S., 2001. Infrastructure development and economic growth: an explanation for regional disparities in China? *Journal of Comparative Economics*, 29 (1), 95–117. doi:10.1006/jcec.2000.1693

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1 (1), 269–271. doi:10.1007/BF01386390

Dinh, T.N. and Thai, M.T., 2013. Community detection in scale-free networks: approximation algorithms for maximizing modularity. *IEEE Journal on Selected Areas in Communications*, 31, 997–1006. doi:10.1109/jsac.2013.130602

Downs, J.A. and Horner, M.W., 2012. Enhancing habitat connectivity in fragmented landscapes: spatial modeling of wildlife crossing structures in transportation networks. *Annals of the Association of American Geographers*, 102 (1), 17–34. doi:10.1080/00045608.2011.600190

Dudás, G., 2013. *Spatial analysis of the world city network using air traffic data*. Thesis (PhD). University of Szeged, Hungary.

Eppstein, D., 1994. Finding the K shortest paths. *In*: *Proceedings 35th annual symposium on foundations of computer science presented at the 35th annual symposium on foundations of computer science*. Santa Fe, NM, USA: IEEE Comput. Soc. Press, 154–165. doi:10.3168/jds.S0022-0302(94)77044-2

Fang, C. and Yu, D., 2017. Urban agglomeration: an evolving concept of an emerging phenomenon. *Landscape and Urban Planning*, 162, 126–136. doi:10.1016/j.landurbplan.2017.02.014

Freeman, T.W. and Dickinson, R.E., 1967. The city region in Western Europe. *The Geographical Journal*, 133 (4), 521. doi:10.2307/1794491

Garcia-López, M.-À., 2012. Urban spatial structure, suburbanization and transportation in Barcelona. *Journal of Urban Economics*, 72 (2–3), 176–190. doi:10.1016/j.jue.2012.05.003

Girvan, M. and Newman, M.E.J., 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99 (12), 7821–7826. doi:10.1073/pnas.122653799

Gottmann, J., 1957. Megalopolis or the urbanization of the Northeastern Seaboard. *Economic Geography*, 33 (3), 189. doi:10.2307/142307

Guimera, R., *et al*., 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102 (22), 7794–7799. doi:10.1073/pnas.0407994102

Guo, D., 2008. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *International Journal of Geographical Information Science*, 22 (7), 801–823. doi:10.1080/13658810701674970

Gustafsson, M., *et al*., 2006. Comparison and validation of community structures in complex networks. *Physica A: Statistical Mechanics and Its Applications*, 367, 559–576. doi:10.1016/j.physa.2005.12.017

He, C., *et al*., 2013. Modeling the urban landscape dynamics in a megalopolitan cluster area by incorporating a gravitational field model with cellular automata. *Landscape and Urban Planning*, 113, 78–89. doi:10.1016/j.landurbplan.2013.01.004

Hu, H., 1935. Distribution of China's population - with statistical tables and density maps(in Chinese). *Acta Geographica Sinica*, 2 (2), 33–74.

Hurst, M.E.E., 1970. An approach to the study of nonresidential land use traffic generation. *Annals of the Association of American Geographers*, 60 (1), 153–173. doi:10.1111/j.1467-8306.1970.tb00709.x

Kernighan, B.W. and Lin, S., 1970. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 49 (2), 291–307. doi:10.1002/j.1538-7305.1970.tb01770.x

Kopits, E. and Cropper, M., 2005. Traffic fatalities and economic growth. *Accident Analysis & Prevention*, 37 (1), 169–178. doi:10.1016/j.aap.2004.04.006

Kotegawa, T., *et al*., 2014. Impact of service network topology on air transportation efficiency. *Transportation Research Part C: Emerging Technologies*, 40, 231–250. doi:10.1016/j.trc.2013.11.016

Lang, R. and Knox, P.K., 2009. The new metropolis: rethinking megalopolis. *Regional Studies*, 43 (6), 789–802. doi:10.1080/00343400701654251

Latora, V. and Marchiori, M., 2001. Efficient behavior of small-world networks. *Physical Review Letters*, 87 (19), 198701. doi:10.1103/PhysRevLett.87.198701

Li, W. and Cai, X., 2007. Empirical analysis of a scale-free railway network in China. *Physica A: Statistical Mechanics and Its Applications*, 382 (2), 693–703. doi:10.1016/j.physa.2007.04.031

Li, X., *et al*., 2016. Exploring the impact of high speed railways on the spatial redistribution of economic activities - Yangtze River Delta urban agglomeration as a case study. *Journal of Transport Geography*, 57, 194–206. doi:10.1016/j.jtrangeo.2016.10.011

Liu, Y., *et al*., 2014. Analyzing relatedness by toponym co-occurrences on web pages: analyzing relatedness by toponym co-occurrences on web pages. *Transactions in GIS*, 18 (1), 89–107. doi:10.1111/tgis.12023

Montis, A.D., *et al*., 2007. The structure of interurban traffic: a weighted network analysis. *Environment and Planning B: Planning and Design*, 34 (5), 905–924. doi:10.1068/b32128

National Statistics Bureau of China, 2017. *China city statistical yearbook*. Beijing, China: China Statistics Press.

Newman, M.E.J., 2006a. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74. doi:10.1103/physreve.74.036104

Newman, M.E.J., 2006b. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103 (23), 8577–8582. doi:10.1073/pnas.0601602103

Newman, M.E.J. and Girvan, M., 2004. Finding and evaluating community structure in networks. *Physical Review E*, 69, 2. doi:10.1103/PhysRevE.69.026113

Rosvall, M. and Bergstrom, C.T., 2007. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104 (18), 7327–7331. doi:10.1073/pnas.0611034104

Rosvall, M. and Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105 (4), 1118–1123. doi:10.1073/pnas.0706851105

Schwarz, A.J., *et al*., 2008. Community structure and modularity in networks of correlated brain activity. *Magnetic Resonance Imaging*, 26, 914–920. doi:10.1016/j.mri.2008.01.048

Straatemeier, T., 2008. How to plan for regional accessibility? *Transport Policy*, 15 (2), 127–137. doi:10.1016/j.tranpol.2007.10.002

van Dongen, M., 2000. *Graph clustering by flow simulation*. Thesis (PhD). University of Utrecht, Netherlands.

Vickerman, R., *et al*., 1999. Accessibility and economic development in Europe. *Regional Studies*, 33 (1), 1–15. doi:10.1080/00343409950118878

Wagner, T., 2010. Regional traffic impacts of logistics-related land use. *Transport Policy*, 17 (4), 224–229. doi:10.1016/j.tranpol.2010.01.012

Wilbanks, T.J., 2010. Accessibility and technological change in Nortean India. *Annals of the Association of American Geographers*, 62, 427–436. doi:10.1111/j.1467-8306.1972.tb00876.x

Xu, J., *et al*., 2017. Difference of urban development in China from the perspective of passenger transport around Spring festival. *Applied Geography*, 87, 85–96. doi:10.1016/j.apgeog.2017.07.014

Xu, X. *et al*., 2007. SCAN: a structural clustering algorithm for networks. *In*: *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining presented at the the 13th ACM SIGKDD international conference*. San Jose, California, USA: ACM Press, 824.

Yamaguchi, K., 2007. Inter-regional air transport accessibility and macro-economic performance in Japan. *Transportation Research Part E: Logistics and Transportation Review*, 43, 247–258. doi:10.1016/j.tre.2006.10.004

Yao, S., *et al*., 2016. *New perspectives on urban agglomerations in China (in Chinese)*. Beijing, China: Science Press.

Yen, J.Y., 1971. Finding the K shortest loopless paths in a network. *Management Science*, 17, 712–716. doi:10.1287/mnsc.17.11.712

Yu, B., *et al*., 2014. Object-based spatial cluster analysis of urban landscape pattern using nighttime light satellite images: a case study of China. *International Journal of Geographical Information Science*, 28 (11), 2328–2355. doi:10.1080/13658816.2014.922186

Zhang, T., *et al*., 2018. Quantifying multi-modal public transit accessibility for large metropolitan areas: a time-dependent reliability modeling approach. *International Journal of Geographical Information Science*, 1–28. doi:10.1080/13658816.2018.1459113

Zhou, H. and Lipowsky, R., 2004. Network brownian motion: a new method to measure vertex-vertex proximity and to identify communities and subcommunities. *In*: M. Bubak, *et al*., eds, *4th international conference on computational science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1062–1069.