Approximating actual flows in physical infrastructure networks: the case of the Yangtze River Delta high-speed railway network

Weiyang Zhang[1, CDFMR], **Ben Derudder**[2, CMR], **Jianghao Wang**[3, D], **Frank Witlox**[4, C]

[1,2,4]*Ghent University,* Department of Geography, Krijgslaan 281-S8, 9000 Ghent, Belgium; [1]phone: +32 486 537 776, e-mail: weiyang. zhang@ugent.be; [2]phone: + 32 92 644 556, e-mail: ben.derudder@ugent.be *(corresponding author)*; [4]phone: + 32 92 644 553, e-mail: frank.witlox@ugent.be; [3]*Chinese Academy of Sciences,* State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Datun road 11 A, Chaoyang district, 100101 Beijing, China; phone: + 8 601 064 888 018, e-mail: wangjh@lreis.ac.cn

Abstract. Previous empirical research on urban networks has used data on infrastructure networks to guesstimate actual inter-city flows. However, with the exception of recent research on airline networks in the context of the world city literature, relatively limited attention has been paid to the degree to which the outline of these infrastructure networks reflects the actual flows they undergird. This study presents a method to improve our estimation of urban interaction in and through infrastructure networks by focusing on the example of passenger railways, which is arguably a key potential data source in research on urban networks in metropolitan regions. We first review common biases when using infrastructure networks to approximate actual inter-city flows, after which we present an alternative approach that draws on research on operational train scheduling. This research has shown that 'dwell time' at train stations reflects the length of the alighting and boarding process, and we use this insight to estimate actual interaction through the application of a bimodal network projection function. We apply our method to the high-speed railway (HSR) network within the Yangtze River Delta (YRD) region, discuss the difference between our modelled network and the original network, and evaluate its validity through a systemic comparison with a benchmark dataset of actual passenger flows.

Contents:

1. Introduction

In his groundbreaking book on 'the rise of the network society', Castells (1996: 377) examines the new spatial logic emerging from the 'complexity of the interaction between technology, society, and space'. This new spatial logic, which Castells famously terms 'the space of flows', has three layers: the electronic impulses in networks, the places which constitute the nodes and hubs of the different networks, and the spatial organization of people in terms of their work, play, and movement. The first layer provides the material support for the network society, i.e. it is the 'technological infrastructure of information systems, telecommunications and transportation lines' (1999: 295) that reflects, determines, supports, and/or enables the network society. Although Castells' book focuses on the information age and the electronic time-sharing practices through space this has brought about, his research can be envisaged as part of a wider metageographical shift emphasizing the importance of 'networks' in the organization of space. For instance, in urban geography we have seen a shift towards 'urban networks' as a major analytical lens which can understand 'urban systems' (e.g. Meijers, 2007; Camagni, Salone, 1993; Zhao et al., 2014).

Based on this general premise, we have seen the emergence of a rich empirical literature on the position of cities in networks at different scales, ranging from 'world city networks' at the global scale (e.g. Taylor, Derudder, 2015) to urban networks that con-

stitute polycentric metropolitan regions (e.g. Burger et al., 2014). The former literature highlights—in spite of its rich diversity—the role of major cities at the crossroads of multiple networks. For instance, when cast in terms of Castells' three-layered structure, the airline networks studied by Smith and Timberlake (2001) and Zook and Brunn (2006) can be understood as analyses of a key infrastructural network (the first layer) centered on world cities (the second layer) in order to facilitate the movement of capital, people, and information (the third layer). Similar observations can be made with respect to analyses of Internet backbone networks (Rutherford et al., 2004; Tranos, 2011), logistics networks (Ducruet, Notteboom, 2012; O'Connor, 2010), office networks of firms (Rozenblat, 2010; Derudder et al., 2013), or a combination of infrastructures (Devriendt et al., 2010; Ducruet et al., 2011).

In strict terms, infrastructure can be thought of as the basic physical and organizational structures and facilities (e.g. ports, buildings, roads, power supplies) needed for the operation of individual organizations and enterprises and/or society and the economy at large. However, infrastructure networks merely provide a material *basis* for tangible flows; they do not cover these tangible flows as such. A good example would be the analysis of urban networks through the lens of air traffic networks (Derudder, Witlox, 2005; Neal, 2014): although data on air traffic networks are widely used in urban network research (Matsumoto, 2004; Smith, Timberlake, 2001), in most cases data

tend to cover the *supply* of route structures between airports (e.g. the data from International Civil Aviation Organisation (ICAO) and International Air Transport Association (IATA)). This focus on the supply side of infrastructure networks does to some degree reflect demand for connectivity between city-pairs, especially in an increasingly deregulated air travel market, but there are of course major intervening effects. The most important one relates to the hub-and-spoke organization of global airline networks, where many passengers are routed via major airports to their destination. This overvaluing of 'major hubs' reveals that an analysis of supply of infrastructure provision does not directly match the actual demand or use. Neal (2014) has recently demonstrated the effect of this for urban network analysis, and prompted him to reveal the structural uniqueness of the networks of supply and demand.
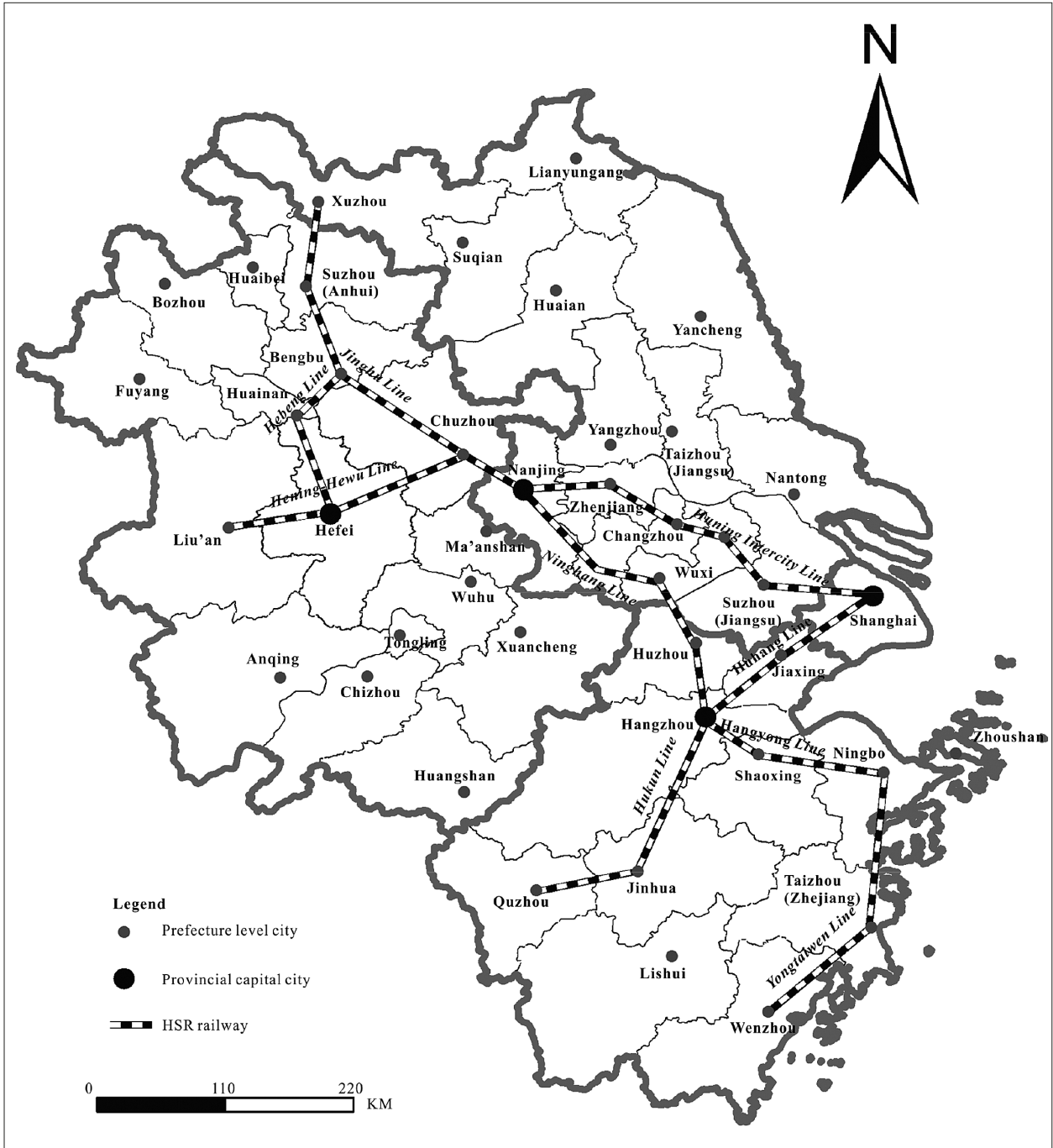

Fig. 1. The high-speed railway network within the Yangtze River Delta

Source: Own studies

To date, however, few studies of urban networks have analyzed the parallels and differences between physical infrastructure networks and the actual flows they enable. In most cases, the former is used as a proxy for the conceptually more meaningful latter. This implies that, in spite of a plethora of papers analysing urban networks through the lens of infrastructure networks, there remains scope for analytical improvement. There are some comparative studies on different layers of urban networks that may inform our understanding of their spatial outline (e.g. Taylor et al., 2007; Liu et al., 2012; Choi et al., 2006), but in this paper we focus more specifically on how data on infrastructure provision can be adapted so that it better reflects actual inter-city flows. To this end, we focus on the example of rail networks reflecting urban network-formation at the level of 'megaregions' (cf. Harrison, Hoyler, 2015). Previous research on this topic serves to clarify our research question. In a recent analysis of infrastructure networks in South Asia, Derudder et al. (2014) find that cities along transport corridors, often defined by road and train networks, are well connected. However, this may be an artifact of the network lay-out rather than 'real connectivity': the connectivity of cities located on a connection between two major interacting nodes may be vastly over-estimated. In the case of the Yangtze River Delta, which will be the empirical focus of this paper, this would result in overestimating the connectivity of Wuxi as it is on the Nanjing-Shanghai HSR line (which is officially called Huning Inter-city Line), granting the Wuxi-Nanjing and Wuxi-Shanghai links de facto equal status to the Shanghai-Nanjing connection (fig. 1). The purpose of this paper is to elaborate a method that would allow for an improved guesstimate of inter-city flows based on infrastructures. The paper focuses on urban networks at lower scales such those in mega-city-regions and countries, where road and rail networks are the key facilitators of inter-city flows.

The remainder of this paper is organized as follows. First, we give a brief overview of the methods for measuring inter-city interactions in railway networks in previous research, and survey the deficiencies encountered by the proxies of infrastructure networks for actual inter-city interactions in more detail. Following this discussion, we focus on setting out an alternative approach to approximating passenger flows in railway networks. This is followed by an empirical test of this approach by applying it to the HSR network within the Yangtze River Delta (YRD) and examining the difference between our transformed network and the original network. We then evaluate the validity of our method through a comparison with a benchmark dataset of actual flows of people, after which the paper is concluded with an overview of our main findings and a discussion of possible avenues for further research.

2. Methods for measuring inter-city interactions through railway networks

Railways constitute one of the main means for transporting people between cities, and thus play a major role in the structuring of inter-city interactions, especially at the regional and national level. Within the burgeoning literature on inter-city networks and spatial interactions, many researchers have thus tried to measure inter-city linkages through the lens of railway networks (e.g. Luo et al., 2011; Hall et al., 2006). However, few papers have mapped inter-city interactions using a *direct* measure of the volumes of inter-city passenger flows. This can be attributed to the lack of data on actual traffic volumes between train stations. As a consequence, a number of researchers have resorted to proxy strategies for measuring inter-city linkages. Two main solutions have been devised in the context of railway networks: (1) measuring the potential for interactions by train, and (2) measuring the volume of trains making inter-city connections.

2.1. Interaction potential

Interaction potential can be defined as the convenience and opportunity of inter-city travel through rail transport. The most commonly used indicator in this respect is travel time, which is often seen as an 'unproductive' cost (time) (Lyons et al., 2007) in a journey influencing potential inter-city interaction (see for example, Kramar, Kadi, 2013; Bruinsma, Rietveld, 1993; Murayama, 1994). Similarly, travel distance or the generalized cost of transport (distance

and time) can also be used as an indicator of measuring the possibility of inter-city journeys (see for example, Wang et al., 2009; Spence, Linneker, 1994). A major obstacle to using this proxy of interaction potential is that infrastructures merely enable the 'potential' of inter-city interactions; actual passenger volumes are co-determined by the 'demand' for inter-city interactions and this 'supply' of transport infrastructures. The 'demand' for inter-city travel can be attributed to the socio-economic attributes of cities and the distance between cities (Davies, 1979; Krings et al., 2009). Even having convenient and efficient transport infrastructures linking to each other does not guarantee that two (social or economic) proximate cities will also exchange a lot of passengers.

A related approach for assessing the potential is using a range of combined measures that not only reflect the quality of infrastructure networks, but also the demand for inter-city linkages. For instance, the indicator of weighted travel time suggested by Gutiérrez (1996; 2001) consists of travel times and urban mass which refers to, for example, gross domestic product or population. However, the 'demand' for inter-city linkages is using simulation approaches rather than more direct measures. Taken together, these indices expressing the potential of inter-city interaction by train mirror the quality or efficiency of train transport infrastructures itself.

2.2. A proxy based on infrastructure volumes

The number of daily or weekly trains has been used as a proxy (Derudder et al., 2014; Hall et al., 2006). Using this proxy instead of the measurements outlined in the previous section has two advantages. As the volume of carriages contains more *direct* information of inter-city flows, it seems to be a more suitable measure of passenger flows. In addition, the information on train numbers can be collected via open information platforms of transport companies much easier than through other ways such as surveys. This proxy also can be viewed as the assessment of transport infrastructures per se, which indicates the traffic supply of infrastructure networks at the level of carriages.

Using the volume of carriages assumes that every train holds similar passenger volumes, which

is of course is problematic. More importantly, this proxy also assumes that the number of trains is proportional to the volume of inter-city passengers between any pair of cities. This is problematic assumption because operationally, train networks are organized by chain structures, unlike air travel or bus trips where direct non-stop services are main organizational forms. A link from an origin to a destination produces $n(n-1)/2n(n-1)/2$ links between any pair of stations if there are n stations en route. In this case, the most important cities hold similar positions with smaller cities that can be found on same railway line, although this obviously does not conform to the actual distribution of inter-city flows of passengers. As a corollary, the volumes of passengers of 'major cities' tend to be underestimated, while the roles of 'small cities' located on major traffic arteries tend to be overstated. Consequently, this proxy structurally predetermines a flatter structure in the urban hierarchy than warranted.

3. An alternative approach to approximating passenger flows in railway networks

3.1. Dwell time

Dwell time, the time a train remains in a given station, is primarily determined by the number of boarding and alighting passengers, as well as some extra factors such as passenger behavior, platform and vehicle characteristics, and dispatching rules (Lin, Wilson, 1992; Wiggenraad, 2001; Jong, Chang, 2011). It is a key parameter of the capacity and performance of operation of trains as insufficient dwell time would lead to train delays, while excessive dwell time would result in inefficient operations (Jong, Chang, 2011). Dwell time, therefore, is set by scientific and efficient principles, which mainly follow the experience of the length of boarding and alighting processes from the past. A normal dwell time lasts between 2 and 5 minutes, with a dwell time of over 5 minutes often implying extraordinary dispatching such as coupling, decoupling, and meeting occurs in that station.

These underlying principles suggest that there is a potential for modelling passenger flows based on the corresponding dwell time in a certain station.

However, eliminating the influence of extraordinary dispatching rules on dwell time is needed: special dispatching (e.g. overtaking, meeting, insufficient headway) clearly biases the interpretation of dwell times, and thus represent outliers. In our research, we will adopt the strategy of replacing outliers with mean values. This is mainly based upon two considerations: (i) simply deleting outliers would be equal to suggesting that trains did not stop in these stations, which is obviously unreasonable; and (ii) as the cause of producing outliers is known in our case, it is possible to replace these outliers using reasonable values to eliminate the effect of abnormal dispatching.

After dealing with outliers, the adjusted dwell times thus correspond with the time of boarding and alighting. According to Jong and Chang's research (2011), the linear relation between the time of passenger flows and the volume of passenger flows is statistically significant. We thus introduce a dummy parameter 'r', which refers to the correlation coefficient between passenger volumes and the boarding and alighting time, to simulate the volume of passenger flows. That is, the volume of passenger flows 'v' is dependent on its adjusted dwell time 't', so that:

$$v = t \times r \qquad (1)$$

The stations of origin and destination do not have dwell times, albeit that they are often the main sources of passengers. To this end, we impose an as-signed value by setting a relatively reliable boarding and alighting time in starting and terminal stations for empirical regions. In our case, the HSR network within the Yangtze River Delta region, most maximal dwell times (after replacing outliers) are around 5 minutes . We posit that the passenger volume from original or to terminal station resemble (or slight exceed) the passenger volume in the largest transit station as a general rule. Thus, we assign the dummy dwell time as 3 minutes.

3.2. Approximating passenger flows between city-pairs

As our object of research is cities rather than train stations, we combine multiple stations into one city through summing adjusted dwell times in the case of there being multiple stations in a single city. From an operational perspective, the distribution of passenger flows for a given train that passes 'n' cities can be summarized by means of an upper triangular matrix as shown in table 1, where 'v_{ij}' is the number of passengers boarding in city 'i' and alighting in city 'j'. In table 1, each row indicates the distribution of alighting for passengers boarding in city 'i'; each column indicates the distribution of boarding for passengers alighting in city 'j'. As a consequence, the sum of each row (V_{ix}) is the number of boarding passengers in city 'i', and the sum of each column (V_{xj}) is the number of alighting passengers in city 'j'.

Table 1. The distribution of passenger flows for a certain train

Alighting city Boarding city	City $_1$	City $_2$	…	City $_j$	…	City $_{(n-1)}$	City $_n$
City $_1$	0	$v_{1,2}$	…	$v_{1,j}$	…	$v_{1,n-1}$	$v_{1,n}$
City $_2$	0	0	…	$v_{2,j}$	…	$v_{2,n-1}$	$v_{2,n}$
…	…	…	…	…	…	…	…
City $_i$	0	0	…	$v_{i,j}$	…	$v_{i,n-1}$	$v_{i,n}$
…	…	…	…	…	…	…	…
City $_{(n-1)}$	0	0	…	0	…	0	$v_{n-1,n}$
City $_n$	0	0	…	0	…	0	0

Source: Own studies

Following equation (1), passenger volumes in city 'i' and 'j' can be obtained by:

$$v_{ix} + v_{xi} = v_i = t_i \times r \qquad (2)$$

$$v_{jx} + v_{xj} = v_j = t_j \times r \qquad (3)$$

If we hypothesize that in the course of a day the boarding and alighting passengers are equivalent in any of the transit cities, then can be formulated as[1]:

$$v_{ix} = v_{xi} = v_i/2 \qquad (4)$$

$$v_{jx} = v_{xj} = v_j/2 \qquad (5)$$

To evaluate the number of passengers boarding at city 'i' and alighting at city 'j' (i.e. v_{ij} in table 1), we first survey the probability of boarding in city 'i' and alighting in city 'j' for all passengers. For any passenger in the passenger distribution 'V', we believe that boarding in city 'i' and alighting in city 'j' are two mutual independent events. According to the rule of the probability of two mutual independent events happening together, the probability of boarding in city 'i' and alighting in city 'j' can be obtained by multiplying the probability of boarding in city 'i' by the probability of alighting in city 'j'. As a corollary, we can approximate the number of passengers boarding at city 'i' and alighting at city 'j' ($v_{i,j}$) from multiplying the number of boarding passengers at city 'i' (v_{ix}) by the number of alighting passengers at city 'j' (v_{xj}):

$$v_{ij} = \alpha \times v_{ix} \times v_{xj} \qquad (6)$$

where α is a dummy parameter describing the relation between v_{ij} and the product of v_{ix} and v_{xj}.

Combining the different equations, v_{ij} can be expressed as the function of dwell times:

$$v_{ij} = \alpha \times r^2 \times (t_i \times t_j) / 4 \qquad (7)$$

However, for the cities of origin or destination, the number of boarding or alighting passengers (v_{ox} or v_{xd}) is equal to the total passenger volumes (v). In these cases, the number of passengers of boarding in origin city 'o' and alighting in transit city 'i' (or boarding in transit city 'i' and alighting in destination city 'd') is given by:

$$v_{oi} = \alpha \times r^2 \times (t_o \times t_i) / 2 \qquad (8)$$

$$v_{id} = \alpha \times r^2 \times (t_i \times t_d) / 2 \qquad (9)$$

Similarly, the number of passengers of boarding in origin city 'o' and alighting in destination city 'd' is given by:

$$v_{od} = \alpha \times r^2 \times (t_o \times t_d) \qquad (10)$$

Based on equations (7-10), the volume of inter-city passengers can be obtained in the form of multiples of '$\alpha \times r^2$'. In the next section, we operationalize this approach by means of a case study.

4. Approximating the flows of high-speed railway (HSR) passengers within the Yangtze River Delta

4.1. Case region, data and transformed network

High-speed railway (HSR) travel plays an important role in facilitating individual movements, thus enabling the formation of larger labor markets in regions and fostering wholesale regional integration (Zheng, Kahn, 2013; Blum et al., 1997; Chen, 2012). Since the first HSR[2] in China became operational in 2007, China's HSR network has been growing rapidly. By the end of 2013, its length reached 10463 km, constituting the longest HSR network in the world. The Yangtze River Delta region is one of the main mega-city regions with intensive HSR networks, where 22 major cities—54% of the entire 41 cities within the YRD[3] (i.e., Shanghai, Nanjing, Hangzhou, Suzhou (Jiangsu), Hefei, Changzhou, Wuxi, Zhenjiang, Bengbu, Chuzhou, Huainan, Lu'an, Quzhou, Suzhou (Anhui), Xuzhou, Jinhua, Ningbo, Huzhou, Shaoxing, Taizhou (Zhejiang), Wenzhou and Jiaxing)—are interconnected through HSR (fig. 1). Our empirical analysis focuses on the passenger flows of HSR among these 22 cities.

Data were gathered from the official website of the customer service center of China's railway (www.12306.cn). This website offers precise information on train operations, which includes prices, transit stations, and dwell times. To iron out the possible effects of operational fluctuations, we mined the information of all HSR trains transiting any city of the YRD region on a fixed day (February 24th, 2014). For every train, we recorded cities of origin and destination, transit cities and their dwell times. The end product that

details the situation of transits (dwell times) is a city-train matrix of 657 trains across the 22 cities. Applying our method, the transformed inter-city network is shown in fig. 2, in which edge thickness reflects the flow strength of city-pairs and node size reflects cities' volumes of passenger flows.
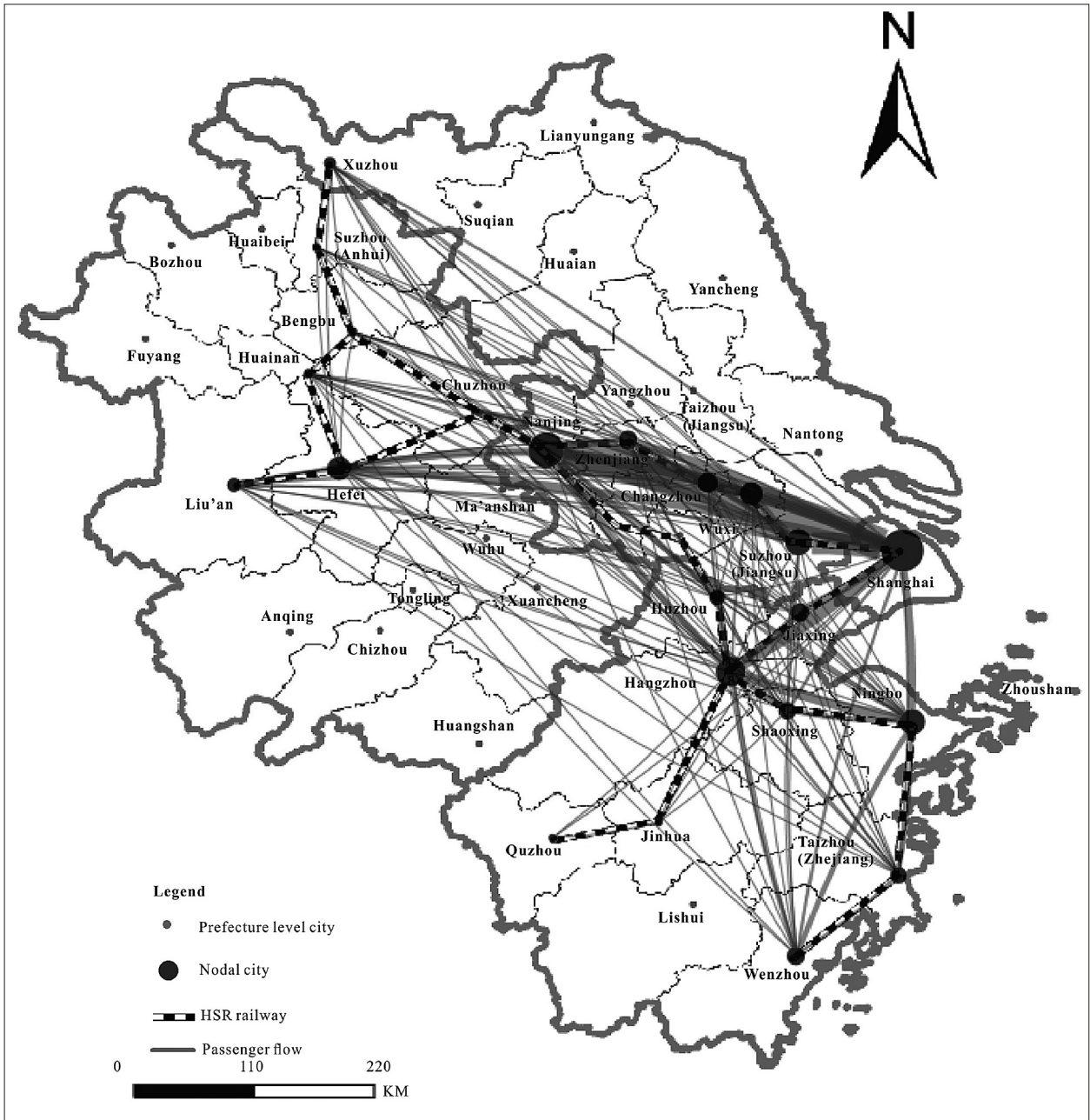

Fig. 2. The transformed network of passenger flows within the Yangtze River Delta

Source: Own studies

The transformed network only connects cities along the HSR network; therefore, only 207 valid (nonzero) inter-city connections in terms of HSR passenger flows are presented in this network. The largest flow is between Shanghai and Nanjing, with 344 HSR trains operating between them daily; the smallest flow is between Changzhou and Quzhou, where only two HSR trains operate on a daily basis. Parallelling the central corridor of the YRD urban agglomerations (Gu et al., 2007), we can observe the

geographic concentration of passenger flows along the Nanjing-Shanghai-Hangzhou-Ningbo belt, where the main HSR lines lie, i.e. Shanghai-Nanjing HSR line, Shanghai-Hangzhou HSR line and Hangzhou-Ningbo HSR line. In addition, Shanghai, Nanjing and Hangzhou emerge as the most connected cities in the network of passenger flows; Suzhou (one of the most dynamic cities that attract foreign direct investment in YRD (Zhao, Zhang, 2007)), Ningbo (the main gateway city in the southern part of the YRD) and Hefei (the administrative and economic centre of Anhui province that has been looking to join the YRD regional collective) are three sub-centre nodes of the network of passenger flows.

4.2. Comparison between the original network generated by the proxy of the number of daily trains and the transformed network

Our alternative approach is devised to address the obstacle of overly flat structures produced by train schedule-based methods for assessing urban networks. Here, we examine the changes put forward by applying the transformation set out in section 3 by comparing original and transformed networks at the level of nodes, linkages, and network structures.

We first offer a direct comparison of cities' degree centralities in both networks (fig. 3). Degree centrality is a measure of nodes' position, which represents the (valued) number of passenger flows of cities. The first obvious change to note is that the degree centralities of a range of cities, which can be separated into two categories, seem lower in the transformed network. The first category is Nanjing, the sub-center city within the YRD. There are 444 HSR trains operating across Nanjing on a daily basis, which is almost the same as the number of HSR trains operating across Shanghai (490 per day). However, part of these trains only transit across Nanjing, while most of them depart from or have their final stop at Shanghai. That means Shanghai contributes most of the passengers, whereas Nanjing only contributes part of the passengers. In this case, Nanjing's position in the original network is obviously overestimated. The other category includes Suzhou (Jiangsu), Wuxi,

Changzhou, Zhenjiang, Shaoxing, and Xuzhou, which are transit cities in main corridors: Suzhou (Jiangsu), Wuxi, Changzhou, and Zhenjiang are on the Shanghai-Nanjing HSR railway line, Shaoxing is on the Hangzhou-Ningbo HSR railway line, and Xuzhou is on the Beijing-Shanghai HSR railway line (fig. 1). This is consistent with the theoretical illustration of over-estimations of the position of transit cities in section 2.2. On the other hand there are also nodes becoming relatively more important in the transformed network. The most dramatic change is the higher rankings of Hefei, Ningbo, Hangzhou and Wenzhou.

Second, fig. 4, in which edge thickness reflects the flow strength of city-pairs, maps the 15 most connected city-dyads in the original network as well as the transformed network. City-dyads along the Nanjing-Shanghai HSR line are the most connected city-dyad—with the exception of Shanghai-Hangzhou—in the original network (fig. 4a). This reflects the fact that any pair of cities along the Shanghai-Nanjing HSR line will have similar number of inter-city trains. Compared with the pattern of concentrating on the Shanghai-Nanjing corridor in the original network, the backbone of the transformed network (fig. 4b) consists of the key cities along the Nanjing-Shanghai-Hangzhou-Ningbo belt, which is more consistent with the central corridor of YRD urban agglomerations (Gu et al., 2007). More specifically, the original network tends to overvalue inter-city connections, such as Nanjing-Wuxi and Shanghai-Zhenjiang, along the Nanjing-Shanghai HSR line, but on the other hand there are also inter-city connections that are being undervalued. These connections can be divided into two simple categories: the connections between Shanghai and sub-centers that are not on the Shanghai-Nanjing corridor (i.e. Hangzhou, Ningbo and Hefei), and the connections between pairs of proximate sub-centers (i.e. Nanjing-Hefei, Hangzhou-Ningbo). In the latter cases, the dense flows of people between Nanjing and Hefei—the closest pair of provincial capitals in China—are apparent, especially in the context of the regional integration of Yangtze Economic Zone. The Hangzhou-Ningbo corridor, along which long-running and dynamic peri-urbanization process has occured (Webster, Muller, 2002), typifies the cooperative pattern of core city (Hangzhou) and sub-centre & port city (Ningbo): Ningbo—Hang-
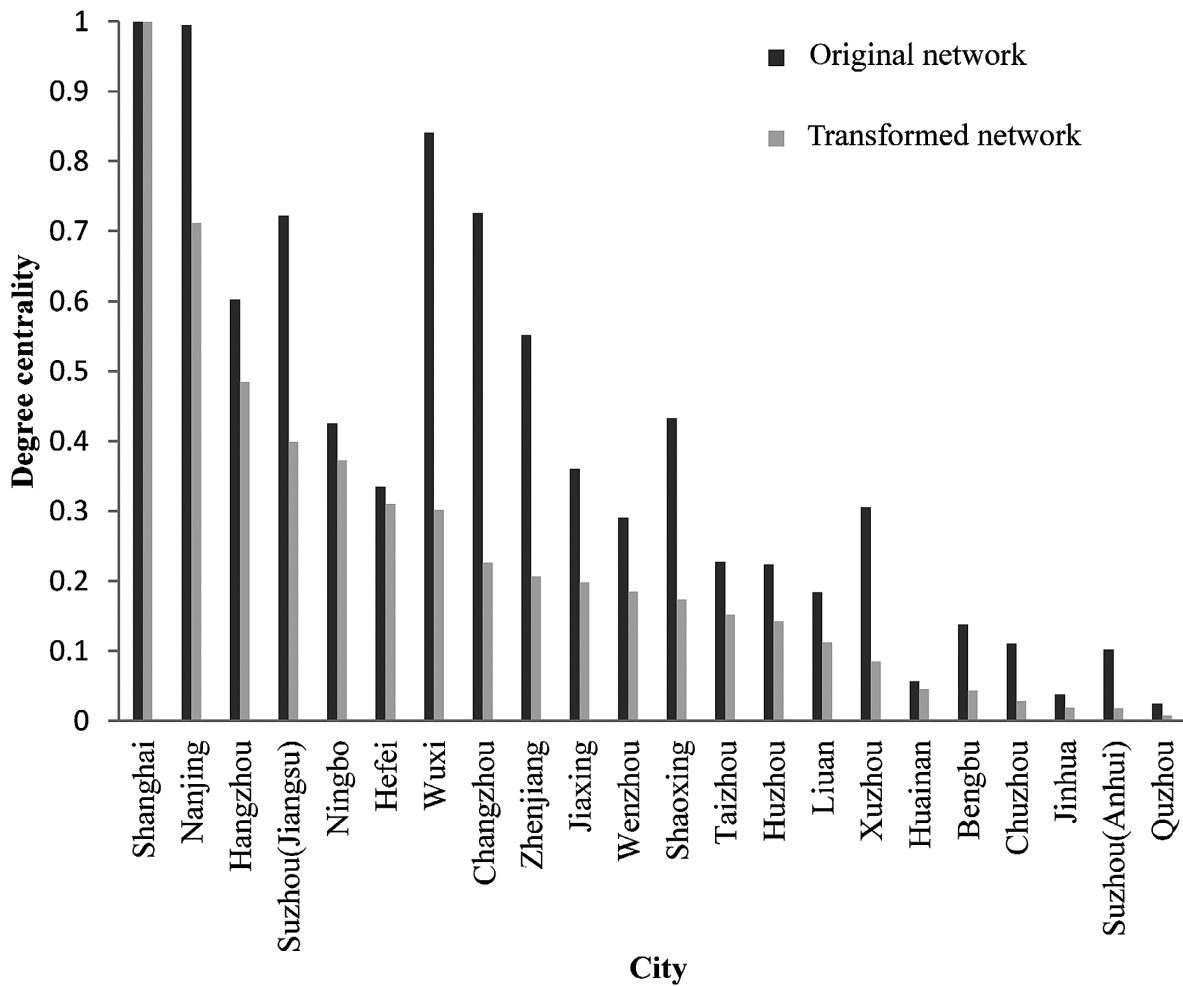
Fig. 3. Cities' degree centralities in the original network and the transformed network

Source: Own studies

zhou's vicinity having more attractive labor, land and tax costs—attracts many manufacturing functions to moving from Hangzhou with keeping R&D and sales functions in Hangzhou (Webster et al., 2003); on the other hand, Ningbo's deep-sea container port provides Hangzhou with more wide international market and hinterland. This provides fundamental bases for the dense inter-city flows between Hangzhou and Ningbo.

And third and finally, to explore the structural difference between both networks, we compare the rank-size distributions of cities' degree centralities. The posited flatter structure of the original network is indeed shown by the much steeper drop-off in the cities' degree centralities in the transformed network, shown in fig. 5. We calculate the integration of rank-size curve of cities' degree centralities to measure

the flat degree of both networks. After normalizing cities' ranks into the interval [0,1], the flattening ratio (F) of networks can be calculated as:

$$F = \int_0^1 L(X)dX \quad F=$$ (11)

where the function $Y = L(X)$ represents the rank-size curve. The flattening ratio varies between 0 for completely even and 1 for completely uneven networks. In our measures, the flattening ratio of the original network (0.39) is much higher than the flattening ratio of the transformed network (0.23): more precisely, the flattening ratio of the transformed network has decreased to 60% of the original flattening ratio in the case of the HSR network within the YRD.
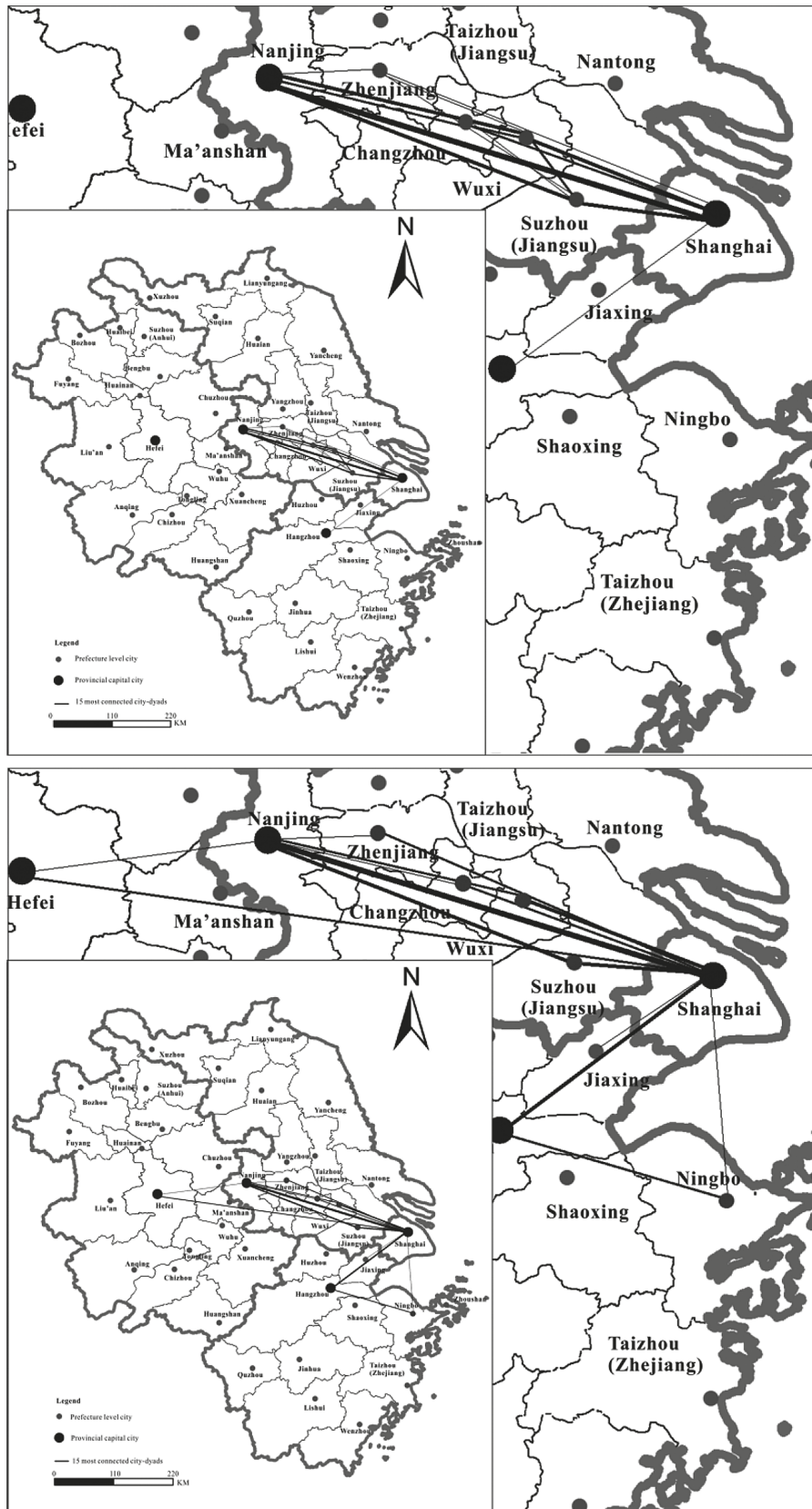
Fig. 4. The 15 largest inter-city links in the original network and the transformed network (up figure: The original network; down figure: The transformed network)
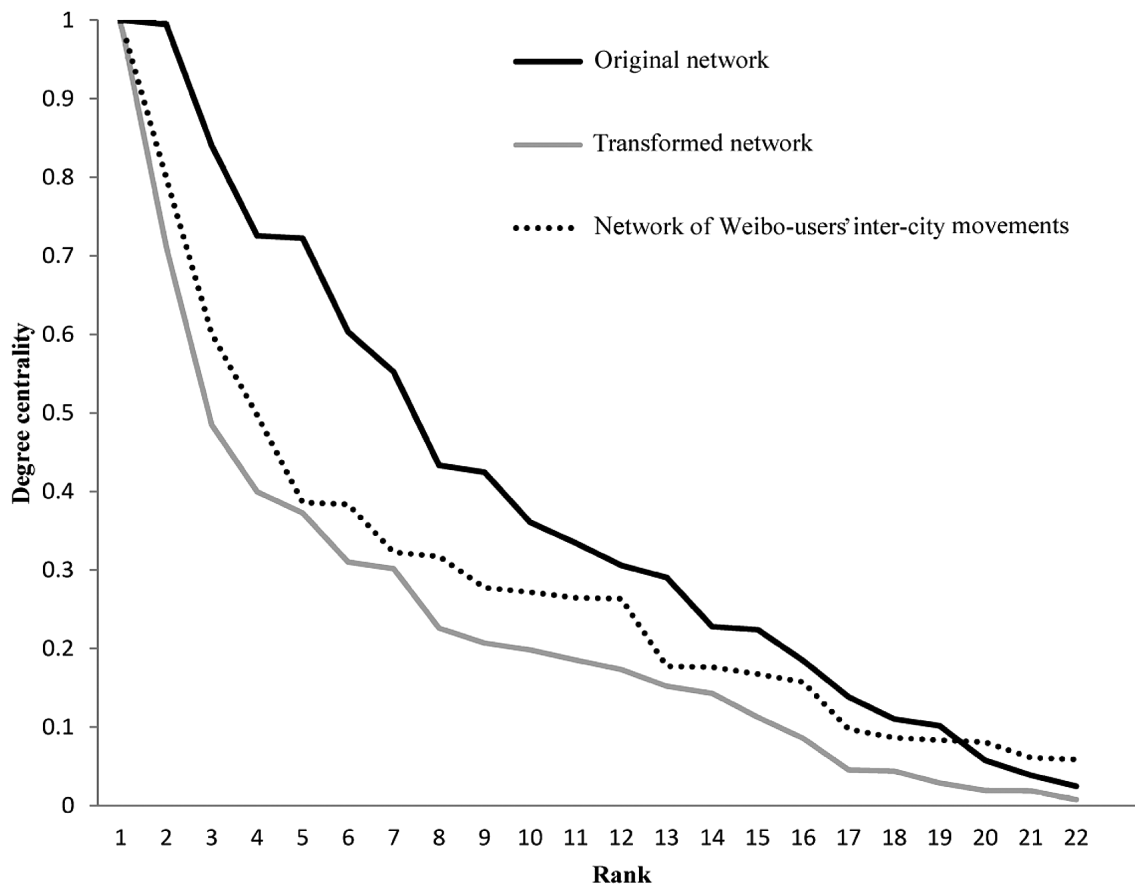
Source: Own studies

Fig. 5. The rank-size distributions of cities' degree centralities in the original network and the transformed network

Source: Own studies

5.   A benchmark test using the data on Weibo users' inter-city movements

In order to evaluate the validity of our method, we need compare the transformed network to a measure of actual flows of people. Due to the difficulty of finding a corresponding database of flows of HSR passengers, here we utilize a database of Weibo[4]-users' inter-city movements, which represents a specific part of tangible flows of people. It can be argued that the flows of HSR passengers and Weibo-users' inter-city movements have similar characteristics. The reason is that they serve relatively similar user markets: the market of HSR is mainly oriented to business travel and leisure tourism of citizens with certain economic means (Wu et al., 2013; also see

Zheng and Kahn, 2013: 'poor rural migrants would not choose bullet trains'); and most of social media users are young adults who have certain economic capacities that include the use of smartphones. Both the collection and subsequent transformation of Weibo data follow the methodology developed in Zhang et al. (2015); here we summarize the main tenets.

Similar to other social media services (such as Facebook and Gowalla), Weibo users are allowed to share their location through a mobile application that is commonly known as a geo-tagged server, thus generating massive location records contributed by millions users. We transform the geo-tagged information into individual travelling trajectories by connecting users' registered place and their visited cities. In practice, we employ an Application

Programming Interface (API) provided by Weibo to crawl all Weibo-users' travel records submitted within the YRD region from March to August 2014. This dataset contains 3 million inter-city footprints; each record represents a directional inter-city flow of a person. Finally, the directional network of Weibo users' flows was converted to an undirected one by combining opposite directional links.

By means of a Pearson correlation measure, we first compare the similarity between both networks (the transformed network and the original network) and the benchmark network of Weibo-users' inter-city movements in terms of cities' connectivities. The correlation coefficients show that, in general, there is a more similar relationship between the transformed network and the benchmark network (r = 0.87 at the 0.001 significance level) which exceeds the coefficient for the original network (r = 0.76 at the 0.001 significance level). We also plot the rank-size distribution of cities' degree centralities in the Weibo-users' movements network (fig. 5) to compare networks' structural similarity. An intuitive sense is that the curve of the Weibo-users' network is closer with the curve of the transformed network. We mathematically compute the flattening ratio of the Weibo-users' network (0.29), which is indeed closer with the flattening ratio (0.23) of the transformed network.

6. Conclusion

The purpose of this paper has above all been methodological: we propose to rethink some of the discrepancies between physical infrastructure networks and actual flows occurring in these networks, focusing on the lens of the railway system. We did so by (1) assessing some limitations in commonly used measures of inter-city rail connections and (2) setting out an alternative approach to approximating passenger flows in railway networks.

Previous empirical research on measuring inter-city linkages through the lens of railway infrastructures has tended to use proxy strategies, where (1) measuring the potential for interactions by train and (2) measuring the volume of trains making inter-city connections stand out as the two main strands due to the lack of data on actual

passenger flows. However, the method of measuring the potential for interactions only mirrors the quality or efficiency of train transport infrastructures itself rather than considering the 'direct demand' for inter-city linkages. And, the proxy of using the volume of trains structurally predetermines a flatter structure in the urban hierarchy than warranted.

This research has shown that 'dwell time' at train stations reflects the length of the alighting and boarding process, and we use this insight to estimate actual interaction through the application of a bimodal network projection function. The empirical application to the high-speed railway (HSR) network within the Yangtze River Delta (YRD) region revealed that the transformed network varies from the original network to a large extent: (i) the positions of transit cities along main transport corridors in the YRD urban system are drove down while some arguable sub-central cities stand out; (ii) inter-city connectivities tend to be more hierarchical; and (iii) the flattening ratio has decreased to 60% of the original flattening ratio. Moreover, the validity of our method has been evaluated through a comparative analysis with Weibo-users' inter-city movements, verifying that the transformed network more parallels tangible flows of people.

We believe our paper makes two contributions to the literature. The first is to remind researchers to re-examine the validity of proxy strategies when measuring inter-city transport flows. With the exception of recent research on airline networks in the context of the world city literature, relatively limited attention has been paid to the degree to which these infrastructure networks reflect the actual flows they undergird. In this regard, this article offers empirical evidence for the structural determinism of using train networks per se, as these tend to result in flatter networks. Second, the central contribution of this paper has been to set out an alternative method of approximating actual flows in railway networks, which permits practical applications in simulating flows of railway passengers in other cases.

Apart from empirical applications in other cases, further research issues also include: discussing other modes of constructing equations, discovering alternative perspectives to approximating actual flows in railway networks, investigating the biases between the infrastructure provision and corporeal flows in

other networks such as Internet backbone and bus networks, and studying how data on these infrastructure operations can be adapted to better reflect actual inter-city interactions.

Notes

(1) This hypothesis is, of course, implausible in any of the transit cities. However, the operational logic of trains is vested in there being round-trip. In this case, the average volumes of boarding and alighting in a daily basis will be roughly equal. For easy operationalization, we adopt an equal weight for boarding passengers and alighting passengers in every transit city for a train.

(2) In the Chinese context, HSR refers to train services with an average speed of 200 km/h or higher, which include D category trains (high-speed trains in conventional railways), G Category trains (high-speed trains in high-speed railways), and C Category trains (short inter-city express trains).

(3) The YRD has various boundaries according to different definitions and research purposes. Throughout this paper, we adopt the largest scope including Shanghai Municipality, Jiangsu Province, Zhejiang Province and Anhui Province, which is also in conformity with the administrative boundary of the Shanghai Railway Bureau.

(4) With more than 212 million monthly active users and 93 million daily active users (see http://goo.gl/ovGvYO), Weibo is the most mainstream social media in China.

References

Blum, U., Haynes, K.E. and Karlsson, C., 1997: Introduction to the special issue The regional and urban effects of high-speed trains. In: *The Annals of Regional Science*, Vol. 31(1), pp. 1-20. DOI: 10.1007/s001680050036

Bruinsma, F. and Rietveld, P., 1993: Urban agglomerations in European infrastructure networks. In: *Urban Studies*, Vol. 30(6), pp. 919-934. DOI: 10.1080/00420989320080861

Burger, M.J., Meijers, E.J. and Oort, F.G., 2014: Multiple perspectives on functional coherence: Heterogeneity and multiplexity in the Randstad. In: *Tijdschrift Voor Economische En Sociale Geografie*, Vol. 105(4), pp. 444-464. DOI: 10.1111/tesg.12061

Camagni, R.P. and Salone, C., 1993: Network urban structures in northern Italy: elements for a theoretical framework. In: *Urban Studies*, Vol. 30(6), pp. 1053-1064. DOI: 10.1080/00420989320080941

Castells, M., 1996: *The Rise of the Network Society,* Oxford: Blackwell.

Castells, M., 1999: Grassrooting the space of flows. In: *Urban Geography*, Vol. 20(4), pp. 294-302. DOI: 10.2747/0272-3638.20.4.294

Chen, C-L., 2012: Reshaping Chinese space-economy through high-speed trains: opportunities and challenges. In: *Journal of Transport Geography*, Vol. 22, pp. 312-316. DOI: 10.1016/j.jtrangeo.2012.01.028

Choi, J.H., Barnett, G.A. and Chon, B.S., 2006: Comparing world city networks: a network analysis of Internet backbone and air transport intercity linkages. In: *Global Networks*, Vol. 6(1), pp. 81-99. DOI: 10.1111/j.1471-0374.2006.00134.x

Davies, W.K., 1979: Urban connectivity in Montana. In: *The Annals of Regional Science*, Vol. 13(2), pp. 29-46. DOI: 10.1007/BF01284778

Derudder, B., Liu, X., Kunaka, C. and Roberts, M., 2014: The connectivity of South Asian cities in infrastructure networks. In: *Journal of Maps*, Vol. 10(1), pp. 47-52. DOI: 10.1080/17445647.2013.858084

Derudder, B., Taylor, P.J., Hoyler, M., Ni, P., Liu, X., Zhao, M., Shen, W. and Witlox, F., 2013: Measurement and interpretation of connectivity of Chinese cities in world city network, 2010. In: *Chinese Geographical Science*, Vol. 23(3), pp. 261-273. DOI: 10.1007/s11769-013-0604-y

Derudder, B. and Witlox, F., 2005: An appraisal of the use of airline data in assessing the world city network: a research note on data. In: *Urban Studies*, Vol. 42(13), pp. 2371-2388. DOI: 10.1080/00420980500379503

Devriendt, L., Derudder, B. and Witlox, F., 2010: Conceptualizing digital and physical connectivity: The position of European cities in Internet backbone and air traffic flows. In: *Telecommunications Policy*, Vol. 34(8), pp. 417-429. DOI: 10.1016/j.telpol.2010.05.009

Ducruet, C., Ietri, D. and Rozenblat, C., 2011: Cities in worldwide air and sea flows: A multiple networks analysis. In: *Cybergeo: European Journal of Geography* 528: https://cybergeo.revues.org/23603. DOI: 10.4000/cybergeo.23603, DoA: 01.12.2015

Ducruet, C. and Notteboom, T., 2012: The worldwide maritime network of container shipping: spatial structure and regional dynamics. In: *Global Networks*, Vol. 12(3), pp. 395-423. DOI: 10.1111/j.1471-0374.2011.00355.x

Gu, C., Zhang, M., Zhang, C., Zhang, X., Wang, C. and Chen, L., 2007: Prospects of urban agglomeration in the Changjiang (Yangtze) River Delta. In: *Scientia Geographica Sinica*, Vol. 27(1), pp. 1. (in Chinese)

Gutiérrez, J., 2001: Location, economic potential and daily accessibility: An analysis of the accessibility impact of the high-speed line Madrid–Barcelona–French border. In: *Journal of Transport Geography*, Vol. 9(4), pp. 229-242. DOI: 10.1016/S0966-6923(01)00017-5

Gutiérrez, J., Gonzalez, R. and Gomez, G., 1996: The European high-speed train network: predicted effects on accessibility patterns. In: *Journal of Transport Geography*, Vol. 4(4), pp. 227-238. DOI: 10.1016/S0966-6923(96)00033-6

Hall, P.G., Pain, K. and Kapoen, L., 2006: The connectivity of the European heartland. In: Hall PG and Pain K, editors, *The Polycentric Metropolis: Learning from Mega-city Regions in Europe,* London: Earthscan, pp. 65-70.

Harrison, J. and Hoyler, M., 2015: *Megaregions: Globalization's New Urban Form?,* Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

Jong, J-C. and Chang, E-F., 2011: Investigation and estimation of train dwell time for timetable planning. In: *Proceedings of 9th World Congress on Railway Research,* Lille, http://www.railway-research.org/IMG/pdf/f1_jong_jyh-cherng.pdf, DoA: 01.12.2015

Kramar, H. and Kadi, J., 2013: Polycentric city networks in Central-Eastern Europe: Existing concepts and empirical findings. In: *Geographia Polonica*, Vol. 86(3), pp. 183-198.

Krings, G., Calabrese, F., Ratti, C. and Blondel, V.D., 2009: Urban gravity: a model for inter-city telecommunication flows. In: *Journal of Statistical Mechanics: Theory and Experiment* 2009(07), pp. L07003. DOI: 10.1088/1742-5468/2009/07/L07003

Lin, T-M. and Wilson, N.H., 1992: Dwell time relationships for light rail systems. In: *The sixth national conference on Light rail transit.* Calgary, Alberta: National Academy Press.

Liu, X., Neal, Z. and Derudder, B., 2012: City networks in the United States: a comparison of four models. In: *Environment and Planning A*, Vol. 44(2), pp. 255-256. DOI: 10.1068/a44496

Luo, Z., He, H. and Geng, L., 2011: Analysis of the polycentric structure of Yangtze River Delta based on passenger traffic flow. In: *Urban Planning Forum*, Vol. (2), pp. 16-23. (in Chinese) DOI: 1000-3363(2011)02-0016-08

Lyons, G., Jain, J. and Holley, D., 2007: The use of travel time by rail passengers in Great Britain. In: *Transportation Research Part A: Policy and Practice*, Vol. 41(1), pp. 107-120. DOI: 10.1016/j.tra.2006.05.012

Matsumoto, H., 2004: International urban systems and air passenger and cargo flows: some calculations. In: *Journal of Air Transport Management*, Vol. 10(4), pp. 241-249. DOI: 10.1016/j.jairtraman.2004.02.003

Meijers, E., 2007: From central place to network model: theory and evidence of a paradigm change. In: *Tijdschrift Voor Economische En Sociale Geografie*, Vol. 98(2), pp. 245-259. DOI: 10.1111/j.1467-9663.2007.00394.x

Murayama, Y., 1994: The impact of railways on accessibility in the Japanese urban system. In: *Journal of Transport Geography*, Vol. 2(2), pp. 87-100. DOI: 10.1016/0966-6923(94)90015-9

Neal, Z., 2014: The devil is in the details: Differences in air traffic networks by scale, species, and season. In: *Social networks* 38, pp. 63-73. DOI: 10.1016/j.socnet.2014.03.003

O'Connor, K., 2010: Global city regions and the location of logistics activity. In: *Journal of Transport Geography*, Vol. 18(3), pp. 354-362. DOI: 10.1016/j.jtrangeo.2009.06.015

Rozenblat C., 2010: Opening the black box of agglomeration economies for measuring cities' competitiveness through international firm networks. In: *Urban Studies*, Vol. 47(13), pp. 2841-2865. DOI: 10.1177/0042098010377369

Rutherford, J., Gillespie, A. and Richardson, R., 2004: The territoriality of pan-European telecommunications backbone networks. In: *Journal of Urban Technology*, Vol. 11(3), pp. 1-34. DOI: 10.1080/10630730500064166

Smith, D.A. and Timberlake, M.F., 2001: World city networks and hierarchies, 1977-1997: An empirical analysis of global air travel links. In: *American Behavioral Scientist*, Vol. 44(10), pp. 1656-1678. DOI: 10.1177/00027640121958104

Spence, N. and Linneker, B., 1994: Evolution of the motorway network and changing levels of accessibility in Great Britain. In: *Journal of Transport Geography*, Vol. 2(4), pp. 247-264. DOI: 10.1016/0966-6923(94)90049-3

Taylor, P.J. and Derudder, B., 2015: World City Network: A Global Urban Analysis (2nd. edition), London: Routledge.

Taylor, P.J., Derudder, B. and Witlox, F., 2007: Comparing airline passenger destinations with global service connectivities: A worldwide empirical study of 214 cities. In: *Urban Geography*, Vol. 28(3), pp. 232-248. DOI: 10.2747/0272-3638.28.3.232

Tranos, E., 2011: The topology and the emerging urban geographies of the Internet backbone and aviation networks in Europe: A comparative study. In: *Environment and Planning*, Vol. 43(2), pp. 378-392. DOI: 10.1068/a43288

Wang, J., Jin, F., Mo, H. and Wang, F., 2009: Spatiotemporal evolution of China's railway network in the 20th century: An accessibility approach. In: *Transportation Research Part A: Policy and Practice*, Vol. 43(8), pp. 765-778. DOI: 10.1016/j.tra.2009.07.003

Webster, D., Cai, J., Muller, L. and Luo, B., 2003: Emerging third stage peri-urbanization: Functional specialization in the Hangzhou peri-urban region. *Working paper. Stanford, CA: Asia-Pacific Research Center, Stanford University.*http://fsi.stanford.edu/sites/default/files/webster_hangzhou_2003.pdf, DoA: 01.12.2015

Webster, D. and Muller, L., 2002: Challenges of peri-urbanization in the Lower Yangtze Region: The case of the Hangzhou-Ningbo Corridor. Working paper. Stanford, CA: Asia-Pacific Research Center, Stanford University.http://aparc.fsi.stanford.edu/sites/default/files/Webster-Muller.pdf, DoA: 01.12.2015

Wiggenraad, P., 2001: Alighting and boarding times of passengers at Dutch railway stations. In: *Papers of the TRAIL workshop train delay at stations and network stability.* Delft: TRAIL Research School.

Wu, K., Fang, C. and Zhao, M., 2013: The intercity space of flow influenced by high-speed rail: A case study for the rail transit passenger behavior between Beijing and Tianjin. In: *Acta Geographica Sinica*, Vol. 68(2), pp. 159-174. (in Chinese)

Zhang, W., Derudder, B., Wang, J. and Witlox, F., 2015: Assessing spatial patterns using Chinese location-based social media: the case of Weibo-users' intercity movements in the Yangtze River Delta. *Working paper available from authors.*

Zhao, M., Liu, X., Derudder, B., Zhong, Y. and Shen, W., 2014: Mapping producer services networks in mainland Chinese cities. In: *Urban Studies*: 0042098014558541. DOI: 10.1177/0042098014558541

Zhao, S. and Zhang, L., 2007: Foreign direct investment and the formation of global city-regions in China. In: *Regional Studies*, Vol. 41(7), pp. 979-994. DOI: 10.1080/00343400701281634

Zheng, S. and Kahn, M.E., 2013: China's bullet trains facilitate market integration and mitigate the cost of megacity growth. In: *Proceedings of the National Academy of Sciences*, Vol. 110(14), pp. E1248-E1253. DOI: 10.1073/pnas.1209247110

Zook, M.A. and Brunn, S.D., 2006: From podes to antipodes: positionalities and global airline geographies. In: *Annals of the Association of American Geographers*, Vol. 96(3), pp. 471-490. DOI: 10.1111/j.1467-8306.2006.00701.x