

PAPER • OPEN ACCESS

Betweenness centrality in urban networks: revealing the transportation backbone of the country from the demographic data

To cite this article: M V Goremyko *et al* 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **177** 012017

View the [article online](#) for updates and enhancements.

Related content

- [Notes on the Algorithm for Calculating Betweenness](#)
Zhou Tao, Liu Jian-Guo and Wang Bing-Hong
- [Betweenness centrality of an edge in tree-like components with finite size](#)
Dongchao Guo, Mangui Liang and Li Wang
- [The effect of attack cost on network robustness](#)
Chen Hong, Xian-Bin Cao, Wen-Bo Du et al.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Betweenness centrality in urban networks: revealing the transportation backbone of the country from the demographic data

M V Goremyko¹, V V Makarov¹, A E Hramov¹, D V Kirsanov¹,
V A Maksimenko¹, A V Ivanov¹, I A Yashkov¹ and S Boccaletti²

¹ REC 'Artificial Intelligence Systems and Neurotechnology', Yuri Gagarin State Technical University of Saratov, 77 Polytechnicheskaja str., 410054, Russia

² CNR-Institute of Complex Systems, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Florence, Italy

E-mail: hramovae@gmail.com

Abstract. The theory of complex networks plays an important role in the modelling and analysis of processes in urban systems, for example, in studies of urban transport networks, the evolution of street networks, passenger flows in the city, etc. In this study we analyze a set of demographic data taken from official population census of Republic of Kazakhstan considering the complex network approach based on spatio-populational principles, namely, the estimation of the distance between cities and taking into account their populations. To determine the geopolitical importance of particular city we use very common network characteristics such as the degree of the node and the betweenness centrality. We show how the values of betweenness can reveal the main transport routes in the country and evaluate the wealth of transportation network.

1. Introduction

In recent years, the interest of researchers dealing with the problems of modeling various complex socio-economic systems is attracted by the network theory as an adequate mathematical tool capable of describing the structure and dynamics of complex processes in many real systems [1]. The formation of social, demographic, economic and other links inevitably leads to the formation of specific structures characterized by the presence of separate elements (or, as they are called in the theory of networks, nodes), between which various connections are established, which means the formation and development of various networks structures [2]. It should be noted that when building network models, communication between nodes can be very diverse and depend on the modeled object [3]. For example, links between network elements can be specified by existing material objects, for example, a network of railway or motorway connections, where links are determined by the presence of a corresponding path, power grids, etc [4]. However, links between network nodes can be defined in abstract space, such as an airline networks, when there are no physical connection between the nodes of the network (in this case between airports), and links are determined by the presence and intensity of air communication between two network vertices [5].

The theory of complex networks plays an increasingly important role in the modeling and analysis of processes in urban systems, for example, in studies of urban transport networks, the evolution of



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

street networks, passenger flows in the city, and so on [6]. In fact, the city is a network of networks where different structures are "nested" in each other and closely intertwined, dynamically changing in the process of urban development [7]. At the same time, the majority of urban studies, which use the complex networks theory, are aimed at the analysis of particular cities or urban area, its economic and transport system. The use of network theory is also of great interest in the study of socio-economic indicators of cities, where a number of interesting results have also been achieved [8]. In particular, various approaches have been proposed to identify the links between cities that are of a socio-economic nature [9]. At the same time, cities are also connected with modern communications and telecommunications, it has been shown that different types of networks are formed between cities, both hierarchical in nature and distributed, and interaction between nodes can be both cooperative and competitive.

Analysis of the available literature allows us to conclude that the network models are promising for understanding the evolution of the urban structure of a particular territory (country) and to identify the features of urban organization, in particular, to analyze the growth or degradation processes (due to the extinction of settlements) of the city network, the network characteristics during the administrative reorganizing processes (for example, transfer of the capital or changing the administrative division).

2. Urban network construction

2.1. Overview of demographic data

Studies of real systems are often complicated by a limited set of statistical data available. This, in turn, encourages us to search for new approaches to extrapolate and to transform the existing data in more vivid representation. In this study we analyze a set of demographic data taken from official population census of Republic of Kazakhstan for three time points, 1989, 1999 and 2009. For each year, the data includes the population of each city, its name, geographical coordinates, and administrative affiliation.

We should note some basic features of the data under study. Firstly, the number of cities decreases during the considered 20 years from 292 to 139 in 2009. Secondly, this process is accompanied by the sharp growth of the average population of the city from 31,769 to 64,096. Such dynamics implies that the cease of the number of cities involved mostly the small towns together with overall outflow of people to regional centers. This also correlates with the relatively stable value of summary urban population of the country: from 9276510 in 1989 it decreased only to 8909333 in 2009.

2.2. The mathematical model

We consider approach to construct a mathematical model of the urban network, which is based on spatio-populational principles, namely, the estimation of the distance between cities and taking into account their populations. We are inspired by early works which attempted to estimate the interaction between cities in the gravity framework [10] together with the principles of spatial growth [11].

In this approach, in the first stage, for all pairs of cities i and j we calculate the weight of connection, w_{ij} , as

$$w_{ij} = \frac{\lg P_i \lg P_j}{R_{ij}}, \text{ for all } i \neq j, \quad (1)$$

where P_i is the population of the i -th city, R_{ij} is the Euclidian distance between the i -th and j -th cities, $i, j = 1..N$, N – the total number of cities in the considered year. We also normalize the set of w_{ij} on the maximal value, so $\max_{i \in 1..N, j \in 1..N} w_{ij} = 1$ and all other values of w_{ij} are in range (0:1). Before proceeding to the construction and analysis of the characteristics of the network, let's consider the question of a specific choice of relation (1) for calculating the weights of links between nodes. The weight of the connection between the two nodes (cities) of the network is proportional to the logarithms of the populations of the corresponding cities and inversely proportional to the distance between the cities. At first glance, it would be more natural to evaluate the weight of connection as $w_{ij} = P_i P_j / R_{ij}$, i.e. in proportion to the product of the population of cities. Why there is a need for complication of the form

of the function? To answer this question, we will consider two fragments of the network in different scale, both including a node corresponding to the city of Almaty (Fig. 1 *a, b*). The Fig. 1 *a* shows a fragment of the nearest cities of Almaty. To the left is a fragment of a map with a marked characteristic scale of 20 km and four cities, three of which are rather small ones having the population of less than 40 thousand people. The diameter of blue circles is proportional to the population, P_i , of corresponding city (the number is also indicated on the map under the city name). The Fig. 1 *b* shows another situation – four large cities, located at large distances from each other (the characteristic scale on the map is 200 km).

On the middle and right panels on Fig. 1 we show the corresponding parts of the network with the values of the weights (marked by the thickness of the lines, and also by the numbers next to the corresponding link) for the two equations for w_{ij} . It is clearly seen that in the first case (the middle row), we have a very large spread of weight values, although it can be assumed that in such a compact urban agglomeration the strength of communication between cities should not be very different. Taking into account that we want to isolate and carry out the study of the formation of the backbone of strongly connected clusters of cities, we introduce nonlinear functions (by adding logarithm in the equation) that grow more slowly than the linear function, as the influence of the population of cities effects on the weight of links, as can be seen from the right figures.

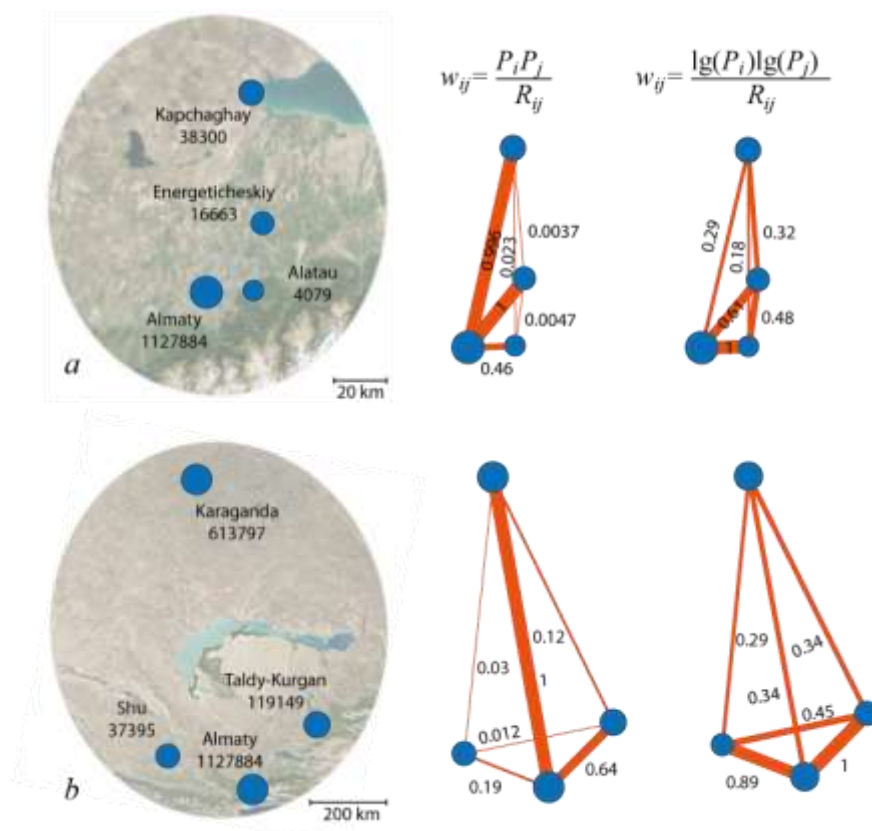


Figure 1. Fragments of a network of urban settlements in the vicinity of Almaty in 1989 explaining the choice of relation (1) for the weights between the nodes

A similar situation arises when we consider large cities located at a great distance from each other, as shown in Fig. 1 *b*. Similarly, when considering the linear connection (middle figure), we get a significant spread in the link weight between cities located at approximately the same distance with each other, while closely located cities are characterized by significantly less ties. The right illustration shows the same network fragment constructed using relation (1), where the links are more

homogeneous, which seems to be more correct. Based on these estimates, we have designed the equation (1) to calculate the weights of the urban network.

However, to obtain more simple model, which can be easily studied using the complex network approach, we need to binarize the links between the cities. A large number of weights, w_{ij} , is characterized by negligibly small values, for example, for pairs of cities with small populations, located at a large distance, R_{ij} , from each other. This is illustrated in Fig. 2, which shows the weight distribution of the 1989, when the number of settlements is the largest ($N = 292$), and hence the number of links. It should be noted that the distributions obtained for other years are qualitatively similar to the distribution presented in Fig. 2. On this basis, it is convenient to enter some threshold value Q , which defines the threshold weight of the connection, which is necessary to connect the two nodes (cities) with each other. The choice of threshold value can be determined from the weight distribution. One can see, that the most of it fits the power law (solid line) nicely. This marks the fact, that the obtained network demonstrates the scale-free property [12], which is essential for the majority of real networks [13]. However, the first point, corresponding to the negligibly small weights does not fits the power law, hence this weights can be removed from the system. More specifically, we choose the value of Q at which one of the networks nodes becomes disconnected, and restore the removed link. This value of Q accurately vanishes the last bin of distribution as seen from Fig. 2.

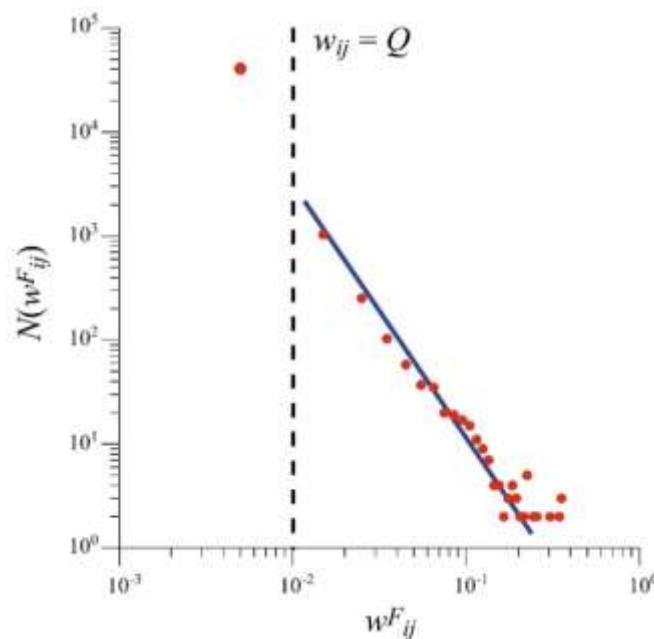


Figure 2. Distribution of weights (red dots) shown in double logarithmic scale. The distribution is based on the data of 1989 year. The vertical dashed line corresponds to a threshold value of Q , by which the weak links were cut off. A solid line corresponds to a power law approximation

The corresponding networks constructed on the basis of our model are shown in Fig. 3 a-c. One can clearly see from them that the model reflects the connections between nearby cities that make up urban agglomerations (clusters) in the north of Kazakhstan, in the south it is tied to the former capital of Almaty, to the west, in the vicinity of the city of Uralsk, etc. There is also the establishment of chains of links through nodes (cities) which are located between such strongly connected clusters, being in good agreement with the form of transport arteries (roads and railways) in Kazakhstan making the connection between large cities through small towns located along roads. In our opinion, this indicates the correctness of the constructed model, which correlates, among other things, with the network of real transport flows.

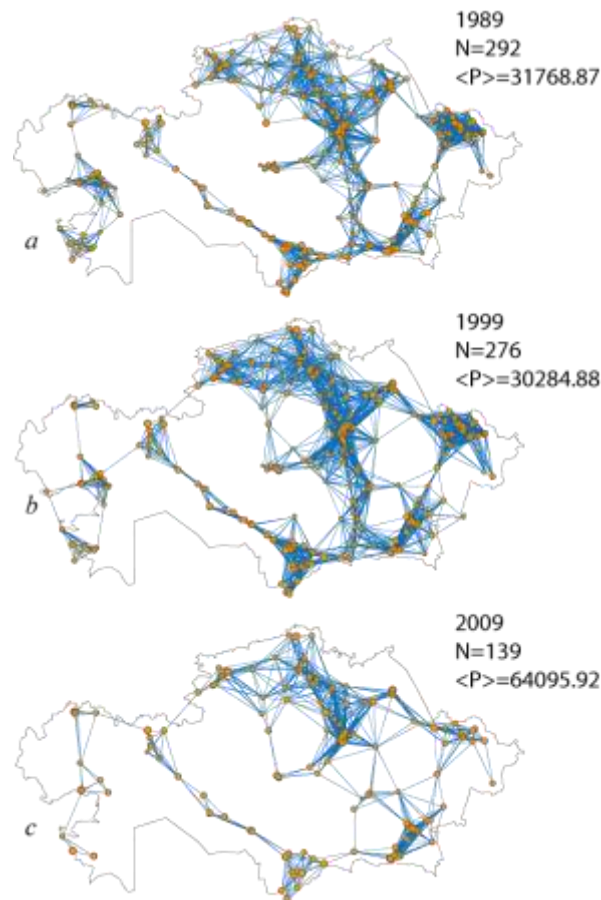


Figure 3. Spatial visualizations of the obtained networks

3. Betweenness centrality in the urban network

To determine the geopolitical importance of particular city we use very common network characteristics such as the degree of the node and the betweenness centrality. The power of node i is a simple sum of all its connections:

$$K_i = \sum_{j=1}^N w_{ij}, \quad (2)$$

which allows to determine whether the given node is a structural hub aggregating connections with other elements, or a peripheral node. Nevertheless, this characteristic is very simple and does not allow to evaluate the real value of that or another network node as a transport hub. To assess the significance of a node in terms of transport flows we use the betweenness centrality, which is the number of shortest paths, g_i , between any two other nodes that pass through the node i [14]:

$$g_i = \sum_{j \neq i} \sum_{l \neq i} \sigma_{jl}, \quad (3)$$

where

$$\sigma_{jl} = \begin{cases} 1, & \text{if the shortest path between nodes } j \text{ and } l \text{ passes through node } i, \\ 0, & \text{if the shortest path between nodes } j \text{ and } l \text{ does not pass through node } i. \end{cases}$$

This type of centrality has shown itself very useful in urban studies, especially while studying street networks [15] or transportation paths [16].

First, we investigate how the betweenness centrality of the node correlates with its degree, i.e. number of links, connected to the node. The corresponding correlation plots are shown in Fig. 3 for each network: (a) 1989, (b) 1999, (c) 2009. Easy to see, that quantities, which we investigate are completely uncorrelated between each other. This occurs as very surprising result, because, intuitively, the nodes with high number of links should aggregate the shortest paths, which are going through them. This also is an essential property of scale-free networks, where hubs accumulate connections with other hubs [17]. On the contrary, in our urban network we observe the existence of small-degree nodes, which demonstrate very high value of betweenness. At the same time, we can also see nodes with high degree, which have almost zero shortest paths going through them.

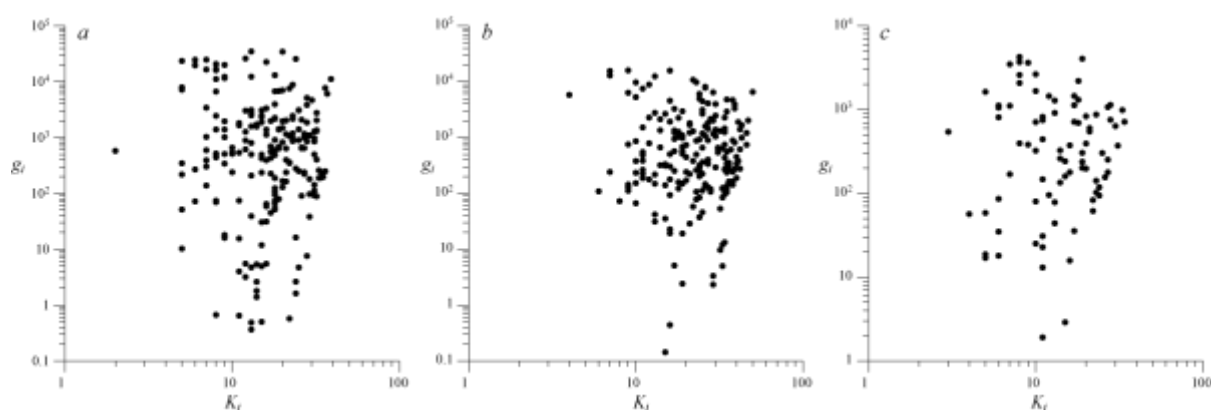


Figure 4. The correlation between the node degree, K_i , and its betweenness centrality, g_i : (a) 1989; (b) 1999, (c) 2009. The plots are presented in double logarithmic scale

To understand the nature of the observed phenomena we present the spatial visualizations of urban networks depicting the betweenness by size of each node. The visualizations are presented in Fig. 5.

In 1989 nodes with high betweenness were located in the south of Kazakhstan and then spreaded to the northwest and north, roughly coinciding with the direction of the main transport routes of Kazakhstan connecting northwestern part of Kazakhstan, Russia and going up to the North, where a large number of cities are also densely concentrated. Moreover, the capital itself, Almaty (the orange dot in Figure 5), does not demonstrate a significant value of betweenness. Characteristics in 1999 and 2009. undergo significant changes related to a restructurization of urban network after the acquisition of independence by Kazakhstan. The changes that have taken place in this regard, first of all, the transfer of the capital and the rapid outflow of the population to big cities with the extinction of small settlements. It can be seen that the changes in the distribution of cities by number across the territory led to the fact that the nodes with the largest betweenness are now located in the north and central Kazakhstan. The southern region with the center in Alma-Ata becomes the periphery of the network and there are very few shortest paths going through it, basically, only those that are associated with this local fragment of the network. It is also pronounced that the role of Astana (the capital) is growing – the number of shortest paths passing through it has almost tripled since 1989. In general, if we consider the transfer of the capital from the point of view of optimizing communications within the urban network, then this was justified step and the place of the new capital was also chosen quite successfully, as indicated by the network analysis of the model.

The observed picture also gives us the answer, why the node betweenness is not correlated with its degree. The spatially-distributed urban network present itself the number of interconnected clusters of cities, which location is determined by complex historical process. This clusters of urbanization have a large number of link inside them, reflecting the strong communication between cities in the same region. At the same time, this dense clusters are interconnected by chain-like node structures, which

mostly repeat the form of the main motorways of the country. Noteworthy, that the long chain of nodes connecting the south and west regions of the country lies almost at the location of the Great Silk Way. This chains of nodes are formed by the cities with rather small populations and, besides, we can observe the process of extinction of this nodes with time. Such dynamics can be considered as a marker of natural or political restructurization of transport paths inside the country, or degrading of transport communication between the regions.

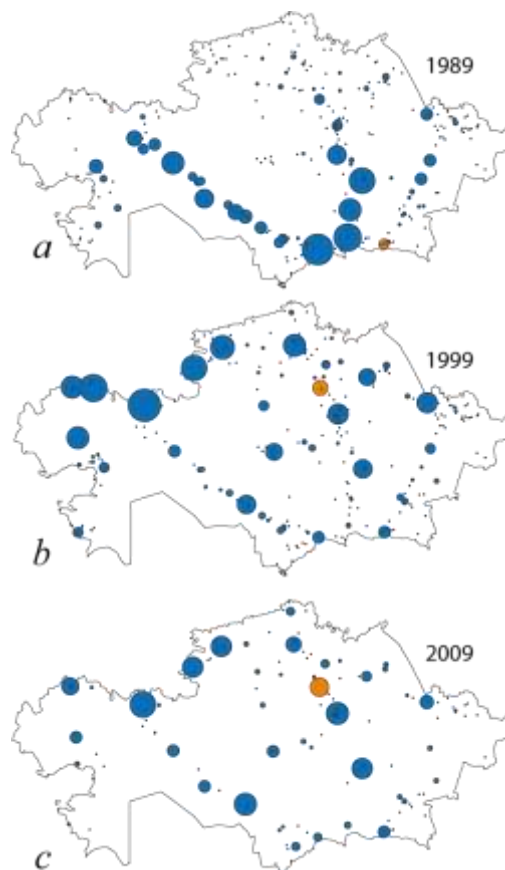


Figure 5. The betweenness centrality of the given networks. The size of the node reflects the value of betweenness. The capital is marked by orange color

4. Conclusion

Summarizing the results, we should note, that the information presented in this paper can be important for optimizing the structure of city networks, traffic flows, demographic and internal migration policies. Further development of mathematical models of such socio-geographic systems will allow to more accurately determine the geographic and demographic characteristics of a network of urban settlements of a particular region or country, and to develop recommendations to government based on rigorous mathematical approaches to improve the efficiency of such systems.

Acknowledgements

This work has been supported by the Ministry of Education and Science of Russian Federation (Grants No. 3.861.2017/4.6 and 3.4593.2017/6.7) and the Council for Grants of the President of the Russian Federation by the supporting program of leading scientific schools (project NSh-2737.2018.2).

References

- [1] Strano E, Shai S, Dobson S and Barthélemy M 2015 *J. R. Soc. Interface* **12** 20150651
- [2] Molinero C, Murcio R and Arcaute E 2017 *Sci. Rep.* **7** 4312

- [3] Nicosia V and Latora V 2015 *Phys. Rev. E* **92** 03280
- [4] Pagani G A, and Aiello M 2013 *Physica A: Statistical Mechanics and its Applications* **392** 11 2688–2700
- [5] Wang J, Mo H, Wang F and Jin F 2011 *Journal of Transport Geography* **19** 4 712–21
- [6] Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A and Batty M 2015 *J R Soc Interface* **12** 20140745
- [7] Capello R 2000 *Urban Studies* **37** 1925–45
- [8] Leo Y, Fleury E, Alvarez-Hamelin J I, Sarraute C and Karsai M 2016 *J R Soc Interface* **13** 20160598
- [9] Zhen F, Wang X, Yin J and Zhang M 2013 *Chinese geographical science* **23** 274–85
- [10] Haynes K E and Fotheringham A S 1984 *Gravity and spatial interaction models* Vol 2 (Beverly Hills, CA: Sage)
- [11] Makse H A, Havlin S and Stanley H E 1995 *Nature* **608** 3776550
- [12] Barabási A L, and Albert R 1999 *Science* **286** 5439 509–12
- [13] Ebel H, Mielsch L I and Bornholdt S 2002 *Phys. Rev. E* **66** 3 035103
- [14] Barthélemy M 2004 *The European physical journal B* **38** 2 163–68
- [15] Crucitti P, Latora V and Porta S 2006 *Phys. Rev. E* **73** 3 036125
- [16] Puzis R, Altshuler Y, Elovici Y, Bekhor S, Shiftan Y and Pentland A 2013 *Journal of Intelligent Transportation Systems* **17** 1 91–105
- [17] Pastor-Satorras R and Vespignani A 2001 *Physical review letters* **86** 3200