PRAMANA — journal of physics © Indian Academy of Sciences Vol. 85, No. 1 July 2015 pp. 173–183

Statistical analysis of the airport network of Pakistan

YASIR TARIQ MOHMAND, AIHU WANG* and HAIBIN CHEN

School of Business Administration, South China University of Technology, Guangzhou 510640, China

*Corresponding author. E-mail: bmawang@scut.edu.cn

MS received 13 July 2013; revised 30 May 2014; accepted 11 June 2014
DOI: 10.1007/s12043-014-0894-3; ***e*Publication:** 26 February 2015

Abstract. Transportation infrastructure plays a vital role in the development of a country's economy and is regarded as one of the most important indicators of its economic growth. In this study, we analyse the Airport Network of Pakistan (ANP), which represents Pakistan's domestic civil aviation infrastructure, as a weighted complex network. We find that ANP is a small-world network and is disassortative in nature. We further analyse the dynamic properties of the network and compare them to their topological counterparts. Although small in size, the ANP does show similar properties as compared to the US, China and especially the Indian airport network.

Keywords. Complex networks; Pakistan; airport network.

PACS Nos. 89.75.−k; 89.20.−a

1. Introduction

The development of a country is usually reflected from the level and quality of its transportation infrastructure. The infrastructure of a country plays a vital role in its economic growth. The transportation infrastructure accomplish not only the movement of the people but also goods and information. As such, it can be regarded as one of the major pillars of an economy. With the expansion of world economy and international trade, the role of transportation infrastructure in the growth of any economy is now clearer than ever before. During the past few years, complex network theory has been applied to better understand the physical and dynamic properties of transportation systems. Although the list of these studies are enormous, some of the related research include the air networks [1–6], railway networks [7–12], public transportation in Singapore [13] and Poland [14], the underground transportation [15,16] and the highway networks [17–19].

As with every country, roadways, railways, and airways are the major means of intercity transport in Pakistan. The target market of the airways compared to roadways and railways is rather small. Recently with the change in government, policies have been put forward to introduce and relax the laws governing licensing of private air companies and as such, this market is expected to grow rapidly, as several new domestic players have showed interest in starting operations in Pakistan. These companies will provide competitive and region-specific services and will target new domestic locations thereby, increasing the air traffic. Understanding these transportation systems is important for reasons of policy, administration and efficiency. In this paper, we study the domestic ANP. Since the network is weighted, we not only study the topological aspects but also the dynamic properties of the network. The rest of the paper is as follows. Sections 2 discusses the network construction methodology, §3 lists the topological and dynamic properties of the network and §4 concludes the paper.

2. Network construction

Data on the airports and the flight schedules were provided by the Civil Aviation Authority (CAA) of Pakistan. The data were transformed and travel for each week was represented as a weighted graph G with N nodes and M edges, an associated adjacency matrix $A = [a_{ij}]$ and a weight matrix $W = [w_{ij}]$ representing the number of weekly flights between airports i and j in a single week. It is important to note that our objective is to study the dynamic properties of the network, not the underlying physical structure of the network. As such, when we say two nodes i and j are connected, $a_{ij} = 1$, we mean that there is at least one flight travelling between these two nodes during the week. Such a representation has already been used to denote the airport network of different countries [1–6].

3. Topological properties

Table 1 provides all computed network statistics, from basic network properties such as the number of nodes and edges to the more complex metrics such as weighted clustering, assortativity and eigenvector centrality.

Table 1. Statistical properties of ANP.

Property	Value
Nodes, N	35
Edges, M	87
Average path length, D	2.1
Average clustering coefficient, C	0.6
Average weighted clustering coefficient, C^{w}	0.63
Diameter, d	4
Average degree	5
Degree range	(1, 25)
Average weight, w	4
Weight range	(1, 28)
Average strength, s	17.4
Strength range	(1, 141)
Assortativity, r	−0.47
Eigenvector centrality	0.14
Weighted eigenvector centrality	0.09

3.1 *Basic properties*

The ANP has 35 nodes with 87 edges. The average shortest path length (the minimum number of edges passed through to get from one node to another) between one node to all other nodes of the network is calculated using the following equation:

$$l = \sum_{s,t \epsilon V} \frac{d\,(a, b)}{N\,(N-1)}, \tag{1}$$

where V corresponds to the set of nodes in the network, $d(a, b)$ is the shortest path from a to b and N is the total number of nodes in the network. A small average path length of a single stop ($l = 2.1$) means that there is travel between almost all the airports of Pakistan. The network also features small diameter (maximum path length of a network) $d = 4$, calculated using

$$d = \frac{1}{N\,(N-1)} \sum_{i \neq j} l_{ij}. \tag{2}$$

The observed path lengths are similar to those found in the airport network of US and India [1,2]. Turning our attention to the number of flights, we see that the range of travel between cities varied greatly, $w \, \epsilon \, (1, 25)$. From the data, it is evident that the flow of passengers is directed from nodes of low degree towards nodes of high degree. Karachi, Islamabad, Lahore, Peshawar and Quetta are a few node examples with high degree which also handle most of the flights. The edge weight distribution is plotted in figure 1.

3.2 *Degree and strength distributions*

The degree of a node, a measure of its connectivity, is defined as the fraction of nodes with degree k in the network. In this case, the degree is defined as the number of cities

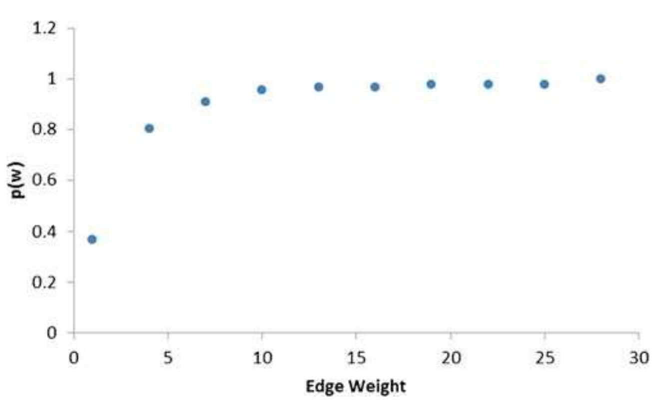

Figure 1. Weight distribution.

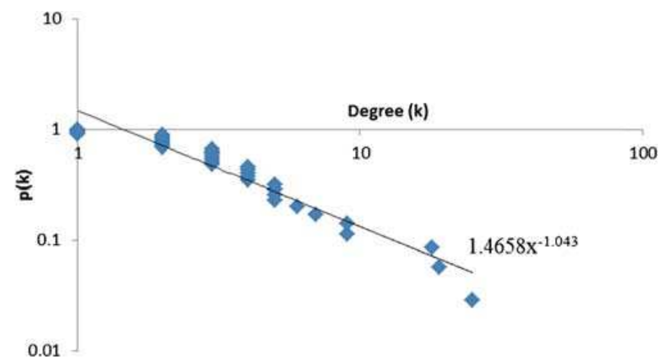

Figure 2. Cumulative degree distribution.

that can be reached from a given city via a single route. For a given node i, the degree can be represented using the equation

$$k_i = \sum_{j}^{N} a_{ij}. \tag{3}$$

Subsequently, a node's strength is simply the sum of the weights on the edges incident upon it and is given by

$$s_i = \sum_{j}^{N} a_{ij} w_{ij}. \tag{4}$$

The network possesses an average degree of 5, indicating high connectivity among the nodes (keeping in view the small number of nodes of ANP). The cumulative degree distribution is plotted in figure 2 which appears to be power law. The strength distribution (figure 3) reveals that although nodes with similar degree exist in the network, the traffic handled by each node is significantly different. We can average the strengths over all nodes with a given degree to get the strength spectrum (figure 4). The spectrum illustrates

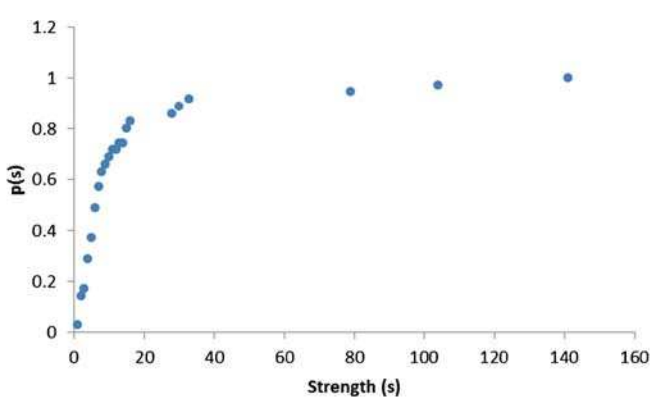

Figure 3. Strength distribution.

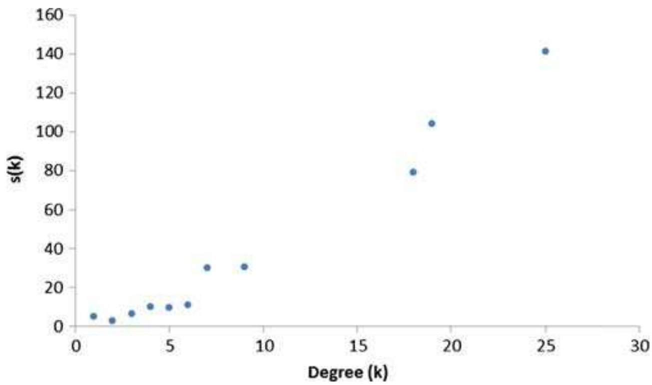

Figure 4. Strength as a function of degree.

a positive relationship between the degree and strength of nodes which means, that the more connected a node is, the more traffic it handles. As nodes are representations of airports in this study, the results of degree and strength distributions provide that the airports with high degree are the ones that handle most of the flights.

3.3 *Topological and dynamic clustering*

The clustering coefficient of a node i is defined as the ratio of the number of links shared by its neighbouring nodes to the maximum number of possible links among them. Simply put, the clustering coefficient is a measure of cohesiveness around a given node i and is defined by the equation

$$C_i = \frac{2E_i}{k_i(k_i-1)} = \frac{2}{k_i(k_i-1)} \sum_{j,h} a_{i,j} a_{i,h} a_{j,h}, \tag{5}$$

where E_i is the number of edges between node i's neighbours and $2/k_i(k_i-1)$ is a normalization factor equal to the maximum number of possible edges among the neighbours. Because of this normalization, C_i is in the interval [0, 1] where 0 and 1 indicate that none or all of the node i's neighbours are linked, respectively. The average clustering coefficient can thus be represented by the following mathematical expression:

$$\langle C \rangle = \frac{1}{N} \sum_{i=0}^{N} C_i. \tag{6}$$

Using the above equation, the average clustering coefficient $\langle C \rangle$ of the network is calculated to be 0.6, indicating that the ANP is a moderately clustered network. This result is substantially higher than the value of an equivalent Erdős–Rényi random graph [20] $\langle C_{ER} \rangle = 0.05$. The clustering coefficient, together with the small average path length (§3.1), indicate that the ANP is a small-world network. The computed clustering coefficient of the network is almost the same as the US network [1] ($C = 0.62$), whereas it is smaller than the Chinese network [6] ($C = 0.73$).

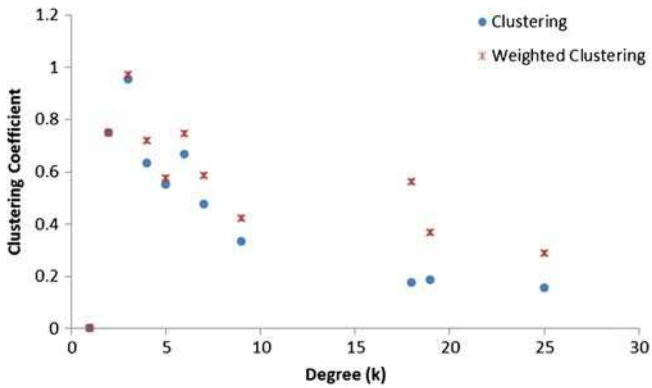

Figure 5. Clustering spectrum.

Unlike the clustering coefficient, the weighted clustering coefficient C_i^{w} also takes into account the weights of edges, i.e.,

$$C_i^{\mathrm{w}} = \frac{1}{s_i\,(k_i - 1)} \sum_{j,h} \frac{w_{i,j} + w_{i,h}}{2} a_{i,j} a_{i,h} a_{j,h}. \tag{7}$$

The average weighted clustering coefficient can thus be represented by the following mathematical expression:

$$\langle C^{\mathrm{w}} \rangle = \frac{1}{N} \sum_{i}^{N} C_i^{\mathrm{w}}. \tag{8}$$

If the weighted clustering coefficient is equal to the clustering coefficient of the network, ($C^{\mathrm{w}} = C$), it means that the weights are completely uncorrelated. However, when the weighted clustering is greater or smaller than the clustering coefficient, the result implies that clustering is formed by edges with larger weight or smaller weights, respectively. In our case, $C^{\mathrm{w}} > C$ implies that the clustering is formed by edges with larger weights (figure 5). Also, we can observe that C falls rapidly in contrast to C^{w}. This indicates that large airports provide air connectivity to far-off unconnected airports.

3.4 *Degree–degree correlation*

Another important topological characteristic of a network that is examined is the degree–degree correlation between connected nodes. A given network is said to be assortative if the high-degree nodes have a tendency to connect to other high-degree nodes. Similarly, in disassortative networks, low-degree nodes tend to connect to high-degree nodes. Newman introduced a summary statistic for assortativity (r) in 2002 [21], defined as the Pearson correlation coefficient of the degrees at either end of an edge. Mathematically, this expression can be represented by the following equation:

$$r = \frac{1}{\sigma_q^2} \sum_{jq} jk \left(e_{jk} - q_j q_k \right), \tag{9}$$

where

$$q_k = \sum_j e_{jk} \quad \text{and} \quad \sigma_q^2 = \sum_k k^2 q_k - \sum_k k q_k^2. \tag{10}$$

This statistic lies in between the range of [−1, 1], where −1 indicates a completely disassortative network and 1 indicates a completely assortative network. Examples of r greater than 0 include the Pakistan highway network [18] and ship transport network of China [22], whereas examples of r less than 0 include the Indian railway network [8,10], public transportation of Singapore [13] and Poland [14] and so on. For ANP, the observed topological disassortativity is −0.47 (similar to Indian air network). A closer inspection of the degree correlations can be done using another measure, the average degree of nearest neighbour, $K_{\rm nn}(k)$, for nodes of degree k.

$$k_{\rm nn}, i = \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j. \tag{11}$$

If $K_{\rm nn}(k)$ increases with k, the network is assortative. If $k_{\rm nn}(k)$ decreases with k, the network is disassortative. The weighted version is given by

$$k_{{\rm nn},i}^{\rm w} = \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} w_{ij} k_j, \tag{12}$$

where $k_{{\rm nn},i}^{\rm w} \approx k_{{\rm nn},i}$ implies that the edge weights are uncorrelated with the degree of i's neighbours. If the resultant weighted neighbour degree is greater than simple neighbour degree ($k_{{\rm nn},i}^{\rm w} > k_{{\rm nn},i}$), then heavily weighted edges connect to neighbours with larger degree, while the opposite occurs when $k_{{\rm nn},i}^{\rm w} < k_{{\rm nn},i}$. The average degree of nearest neighbour is represented in figure 6, where $K_{\rm nn}(k)$ decreases with k and $k_{{\rm nn},i}^{\rm w} > k_{{\rm nn},i}$ means the network is disassortative and heavily weighted edges connect to neighbours with larger degree.

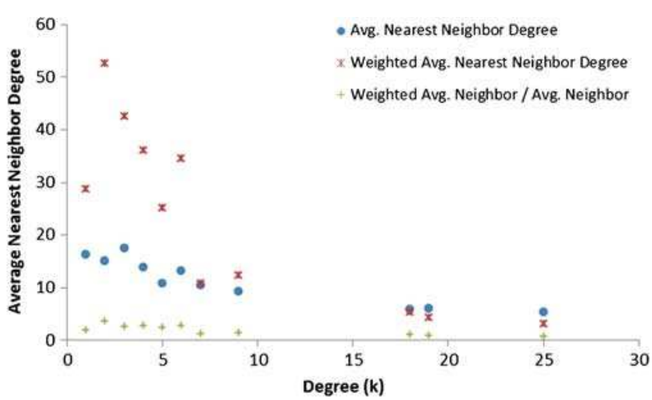

Figure 6. Average degree of nearest neighbours of nodes with degree k.

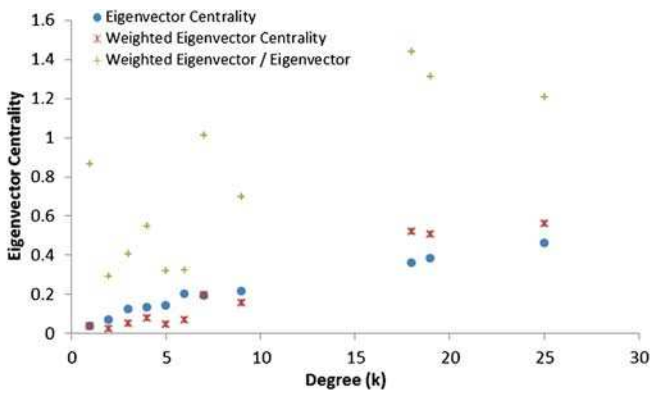

Figure 7. Eigenvector centrality.

3.5 *Eigenvector centrality*

In the previous sections, we described the degree and strength centralities which are the simplest measures of centrality. Apart from them, there also exist other centrality measures like betweenness and closeness centralities; however, we shall not be using them as these are defined only for unweighted networks. Instead, we shall be using the eigenvector centrality measure. The basic theme behind this centrality measure is that the quality of an edge should matter, i.e., an edge to a highly central node should matter more than an edge to a node with low centrality. As with all other measures, we average the eigenvector and weighted eigenvector over all nodes with a certain degree k (figure 7).

Both the weighted and unweighted eigenvectors appear to increase slightly with k because nodes with low degree are connected mainly to nodes with high traffic and as a result have great importance. This is supported by the disassortiveness of the weighted assortativity spectrum (figure 6). For example, lowest degree nodes like Skardu and Gilgit are connected to Islamabad having high traffic flow.

4. Robustness and community analysis

Technically, the term robustness refers to the endurance of the network's behaviour to conditions of uncertainty [23]. In simpler words, the robustness of a network is the ability of the network to maintain its function when nodes or edges in the network suffer from random or intentional attack [24]. Although two different approaches exist in earlier studies to analyse the robustness of a network, in this paper we use the static robustness (the other being dynamic robustness).

The average network efficiency of the network is taken to analyse the robustness of the airport network of Pakistan. The strength of connectivity can be reflected by E [25]. The connected efficiency ε_{ij} between a node pair (i, j) is assumed to be the inverse of the path length d_{ij} between the two nodes, and E is the mean of ε_{ij} values of all node pairs [24]. Mathematically,

$$E(G) = \frac{\sum_{i \neq j \in G} \varepsilon_{ij}}{n\,(n-1)} = \frac{1}{n\,(n-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}},$$

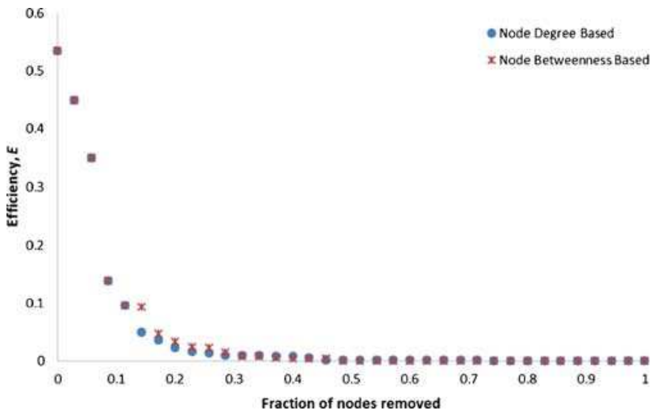

Figure 8. Robustness analysis.

where n is the number of nodes in the network. The larger the E, the better is the connectivity of the network.

The robustness of the airport network is represented in figure 8. As is evident from the figure, the airport network of Pakistan is highly dependent on the high degree nodes. On failure of the high degree nodes, the efficiency of the network drops significantly, until and unless the efficiency becomes almost equal to zero for the low degree nodes.

Another parameter studied is the community analysis of the network. The communities have been identified by following the procedure put forth by Newman, whereby modularity is used to identify the communities in a network. Modularity according to Newman, is

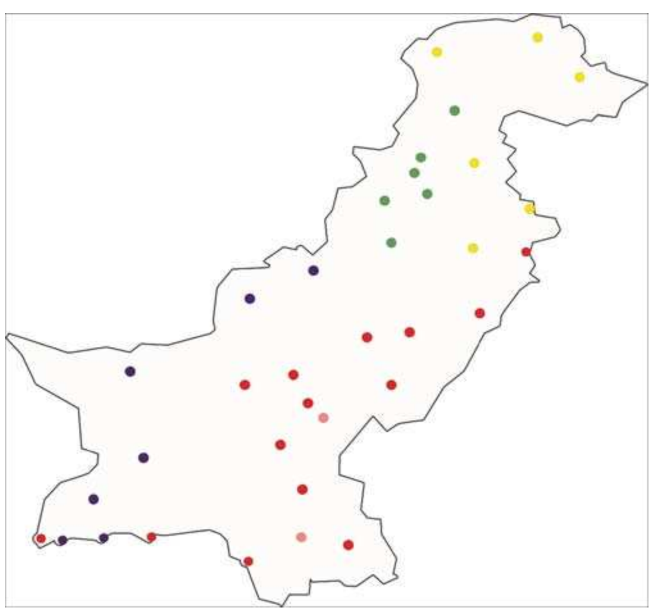

Figure 9. Community analysis.

the number of edges falling within groups minus the expected number of edges in an equal sized network of randomly placed edges and hence, positive value of modularity indicates the presence of a community structure in the network. In the case of the ANP, community structure of the network is performed by optimizing for positive modularity. The resultant output is presented in figure 9. The high degree nodes of the ANP form a separate community as these handle most of the traffic of the network and as such fall in the same community.

5. Conclusion and future work

Transportation networks, be it land, air or sea, represent the development level of a country and can rightly be described as the backbone of economic development. Along with other tools, complex network methodologies have been extensively used to analyse transportation networks. As an addition to the theory and application of complex networks, the weighted domestic airport network of Pakistan is analysed using complex network theory. The ANP is a moderately clustered network, where the degree distribution is a power law with a few nodes having the highest degree connectivity. The small-world properties and disassortative mixing of the network are evident from the calculated results. Although the network is small compared to Chinese, Indian and the US airport networks, in terms of topological properties, the ANP shows some resemblance to these networks, especially to its close neighbour, India.

With the addition of several low-cost airline services, the ANP is expected to grow, both in terms of coverage and frequency of flights. It will be interesting to study this network with regard to passenger or cargo flow as it will further provide an understanding of the dynamic features of the airport network. Also, other research methods can be applied to the airlines operating in Pakistan by comparing the results to their counterparts. Such a study will point out the limitations and hopefully leave room for improvements in the current aviation industry of Pakistan.

Acknowledgements

This research is supported by 2011 Founded Project of National Natural Science Foundation of China (71171084), 2011, Research Fund for the Doctoral Program of Higher Education of China (20110172110010) and The Fundamental Research Funds for the Central Universities (2012, x2gsD2117850).

References

[1] C Li-Ping *et al*, *Chin. Phys. Lett.* **20**, 1393 (2003)
[2] G Bagler, *Physica A* **387**, 2972 (2008)
[3] R Guimerá and L A N Amaral, *Euro. Phys. J. B* **38**, 381 (2004)
[4] R Guimera, S Mossa, A Turtschi and L A N Amaral, *Proc. Natl. Acad. Sci. USA* **102**, 7794 (2005)
[5] L Hong-Kun and Z Tao, *Acta Phys. Sin.* **56**, 106 (2007)
[6] W Li and X Cai, *Phys. Rev. E* **69**, 046106 (2004)

[7] X Cai and L Guo, *Int. J. Mod. Phys. C* **19**, 1909 (2008)
[8] S Ghosh *et al*, *Acta Phys. Polon. B: Proc. Suppl.* **4**, 123 (2011)
[9] W Li and X Cai, *Physica A* **382**, 693 (2007)
[10] P Sen *et al*, *Phys. Rev. E* **67**, 036106.1 (2003)
[11] Y-L Wang *et al*, *Physica A* **388**, 2949 (2009)
[12] Y T Mohmand and A Wang, *Disc. Dyn. Nat. Soc.* **2014** (2014)
[13] H Soh *et al*, *Physica A* **389**, 5852 (2010)
[14] J Sienkiewicz and J A Hołyst, *Phys. Rev. E* **72**, 046127 (2005)
[15] V Latora and M Marchiori, *Physica A* **314**, 109 (2002)
[16] V Latora and M Marchiori, *Phys. Rev. E* **71**, 015103 (2005)
[17] A Erath, M Löchl and K Axhausen, *Net. Spat. Econ.* **9**, 379 (2009)
[18] Y T Mohmand and A Wang, *Disc. Dyn. Nat. Soc.* **2013**, 5 (2013)
[19] P R Villas Boas, F A Rodrigues and L da F Costa, *Phys. Lett. A* **374**, 22 (2009)
[20] P Erdős and A Renyi, *Acta Math. Hung.* **12**, 261 (1961)
[21] M E J Newman, *Phys. Rev. Lett.* **89**, 208701 (2002)
[22] X Xu, J Hu and F Liu, *Chaos* **17**, 023129.1 (2007)
[23] R Albert, H Jeong and A-L Barabasi, *Nature* **406**, 378 (2000)
[24] Z Zou, Y Xiao and J Gao, *Kybernetes* **42**, 383 (2013)
[25] R Albert and A-L Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002)