Statistical analysis of 22 public transport networks in Poland

Julian Sienkiewicz and Janusz A. Hołyst
Faculty of Physics and Center of Excellence for Complex Systems Research, Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland

(Received 8 June 2005; published 20 October 2005)

Public transport systems in 22 Polish cities have been analyzed. The sizes of these networks range from $N=152$ to 2881. Depending on the assumed definition of network topology, the degree distribution can follow a power law or can be described by an exponential function. Distributions of path lengths in all considered networks are given by asymmetric, unimodal functions. Clustering, assortativity, and betweenness are studied. All considered networks exhibit small-world behavior and are hierarchically organized. A transition between dissortative small networks $N \lesssim 500$ and assortative large networks $N \gtrsim 500$ is observed.

DOI: 10.1103/PhysRevE.72.046127 PACS number(s): 89.75.Hc, 02.50.−r, 05.50.+q

I. INTRODUCTION

Since the explosion of the complex network science that has taken place after the works of Watts and Strogatz [1] as well as Barabási and Albert (BA) [2,3] a lot of real-world networks have been examined. The examples are technological networks (Internet, phone call networks), biological systems (food webs, metabolic systems), or social networks (coauthorship, citation networks) [4–7]. Despite this, at the beginning little attention was paid to *transportation networks*—mediums as important and also sharing as much complex structure as those previously listed. However, during the past few years, several *public transport systems* (PTS) have been investigated using various concepts of statistical physics of complex networks [8–21].

Chronogically the first works regarding transportation networks dealt with power grids [1,2,8,9]. One can argue that transformers and transmission lines have little in common with PTS (i.e., underground, buses, and tramways), but they definitely share at least one common feature: embedding in a two-dimensional space. Research done on the electrical grid in the United States—for Southern California [1,2,8,9] and for the whole country [10]—as well as on the GRTN Italian power network [11] revealed single-scale degree distributions [$p(k) \propto \exp(-\alpha k)$ with $\alpha \approx 0.5$], small average connectivity values, and relatively large average path lengths.

All railway and underground systems appear to share well known small-world properties [1]. Moreover, this kind of network possesses several other characteristic features. In fact, Latora and Marichiori have studied in detail a network formed by the Boston subway [12–14]. They have calculated a network *efficiency* defined as a mean value of inverse distances between network nodes. Although the global efficiency is quite large, $E_{\mathrm{glob}}=0.63$, the local efficiency calculated in the subgraphs of neighbors is low, $E_{\mathrm{local}}=0.03$, which indicates a large vulnerability of this network against accidental damages. However, the last parameter increases to $E'_{\mathrm{local}}=0.46$ if the subway network is extended by the existing bus routes network. Taking into account geographical distances between different metro stations, one can consider the network as a weighted graph and one is able to introduce a measure of a network cost. The estimated relative cost of the Boston subway is around 0.2% of the total cost of a fully connected network.

Sen *et al.* [15] have introduced a topology describing the system as a set of train lines, not stops, and they have discovered a clear exponential degree distribution in the Indian railway network. This system has shown a small negative value of assortativity coefficient. Seaton and Hackett [16] have compared real data from the underground systems of Boston (first presented in [14]) and Vienna with the prediction of bipartite graph theory (here, graph of lines and graph of stops) using a generation function formalism. They have found a good correspondence regarding the value of average degree, however other properties such as clustering coefficient or network size have shown differences of 30–50 %.

In the works of Amaral, Barrat, Guimerà, etc. [8,17–19], a survey on the World-Wide Airport Network has been presented. The authors have proposed truncated power-law cumulative degree distribution $P(k) \propto k^{-\alpha} f(k/k_x)$ with the exponent $\alpha=1.0$ and a model of preferential attachment where a new node (flight) is introduced with a probability given by a power law or an exponential function of physical distance between connected nodes. However, only an introduction of geopolitical constraints [19] (i.e., only large cities are allowed to establish international connections) explained the behavior of betweenness as a function of node degree.

Other works on airport networks in India [20] and China [21] have stressed small-world properties of those systems, characterized by small average path lengths ($\langle l \rangle \approx 2$) and large clustering coefficients ($c > 0.6$) with comparison to random graph values. Degree distributions have followed either a power law (India) or a truncated power law (China). In both cases, evidence of strong disassortative degree-degree correlation has been discovered, and it also appears that the airport network of India has a hierarchical structure expressed by a power-law decay of clustering coefficient with an exponent equal to 1.

In the present paper, we have studied a part of data for PTS in 22 Polish cities and we have analyzed their node degrees, path lengths, clustering coefficients, assortativity, and betweenness. Despite large differences in the sizes of the considered networks (number of nodes ranges from $N=152$ to 2881), they share several universal features such as degree

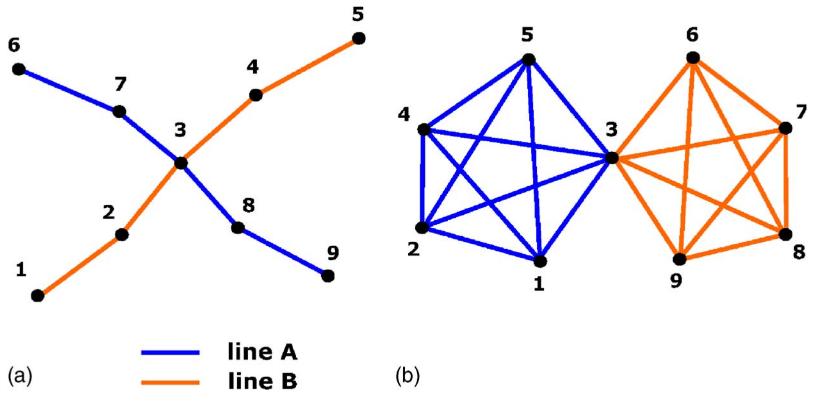

FIG. 1. (Color online) Explanation of the space L (a) and the space P (b).

and path length distributions, logarithmic dependence of distances on node-degrees, or a power-law decay of clustering coefficients for large node degrees. As far as we know, our results are the first comparative survey of several public transport systems in the same country using universal tools of complex networks.

II. THE IDEA OF SPACES L AND P

To analyze various properties of PTS, one should start with a definition of a proper network topology. The idea of the spaces L and P, proposed in a general form in [15] and used also in [16], is presented in Fig. 1. The first topology (space L) consists of nodes representing bus, tramway, or underground stops, and a link between two nodes exists if they are consecutive stops on the route. The node degree k in this topology is just the number of directions (it is usually twice the number of all PTS routes) one can take from a given node while the distance l equals the total number of stops on the path from one node to another.

Although nodes in the space P are the same as in the previous topology, here an edge between two nodes means that there is a direct bus, tramway, or underground route that links them. In other words, if a route A consists of nodes a_i, i.e., $A=\{a_1,a_2,\ldots,a_n\}$, then in the space P the nearest neighbors of the node a_1 are $a_2,a_3,\ldots,a_n$. Consequently, the node degree k in this topology is the total number of nodes reachable using a single route and the distance can be interpreted as a number of transfers (plus one) one has to take to get from one stop to another.

Another idea of mapping a structure embedded in two-dimensional space into another, dimensionless topology has recently been used by Rosvall *et al.* in [22], where a plan of the city roads has been mapped into an "information city network." In the last topology, a road represents a node and an intersection between roads represents an edge, so the network shows information handling that has to be performed to get oriented in the city.

We need to stress that the spaces L and P do not take into account Euclidean distance between nodes. Such an approach is similar to the one used for description of several other types of network systems: Internet [2], power grids [10,11], railway [15], or airport networks [20,21].

III. EXPLORED SYSTEMS

We have analyzed PTS (bus and tramway systems) in 22 Polish cities, located in various state districts as it is depicted in Fig. 2. Table I gathers fundamental parameters of considered cities and data on average path lengths, average degrees, clustering coefficients, as well as assortativity coefficients for corresponding networks.

Numbers of nodes in different networks (i.e., in different cities) range from $N=152$ to 2811 and they are roughly proportional to populations I and surfaces S of corresponding cities (see Fig. 3). One should notice that other surveys exploring the properties of transportation networks have usually dealt with smaller numbers of vertices, such as $N=76$ for the U-Bahn network in Vienna [16], $N=79$ for the Airport Network of India (ANI) [21], $N=124$ in the Boston Underground Transportation System (MBTA) [14], or $N=128$ in the Airport Network of China (ANC) [20]. Only in the cases of the Indian Railway Network (IRN) [15] where $N=579$ and the World-Wide Airport Network (WAN) [19] with 3880 nodes have the sizes of the networks been similar to or larger than for PTS in Poland. Very recently, von Ferber *et al.* [23] have presented a paper on three large PTS: Düsseldorf with $N=1615$, Berlin with $N=2952$, and Paris where $N=4003$.

FIG. 2. Map of examined cities in Poland.

TABLE I. Data gathered on 22 cities in Poland. S stands for the surface occupied by the city (in km^2) [24], I is the city's population in thousands of inhabitants [24], and N is the number of nodes (stops) in the network. $\langle l \rangle$ is the average path length, $\langle k \rangle$ is the average degree value, c is the clustering coefficient, and r is the assortativity coefficient. Indexes L and P stand, consequently, for the space L and the space P. Properties of parameters defined in spaces L and P will be discussed in Secs. IV–VII.

	Basic parameters			Space L				Space P			
City	N	S	I	$\langle l \rangle_L$	$\langle k \rangle_L$	c_L	r_L	$\langle l \rangle_P$	$\langle k \rangle_P$	c_P	r_P
Piła	152	103	77	7.86	2.90	0.143	0.236	1.82	38.68	0.770	0.022
Bełchatów	174	35	65	16.94	2.62	0.126	0.403	1.71	49.92	0.847	−0.204
Jelenia Góra	194	109	93	11.14	2.53	0.109	0.384	2.01	32.94	0.840	0.000
Opole	205	96	129	10.29	3.03	0.161	0.320	1.80	50.19	0.793	−0.108
Toruń	243	116	206	10.24	2.72	0.134	0.068	2.12	35.84	0.780	−0.055
Olsztyn	268	88	173	12.02	3.08	0.111	0.356	1.91	52.91	0.724	0.020
Gorzów Wlkp.	269	77	162	16.41	2.48	0.082	0.401	2.40	38.51	0.816	−0.033
Bydgoszcz	276	174	386	10.48	2.61	0.094	0.147	2.10	33.13	0.799	−0.068
Radom	282	112	232	10.97	2.84	0.089	0.348	1.98	48.14	0.786	−0.067
Zielona Góra	312	58	119	6.83	2.97	0.067	0.237	1.97	44.77	0.741	−0.115
Gdynia	406	136	255	11.41	2.78	0.153	0.307	2.22	52.68	0.772	−0.018
Kielce	414	109	93	16.98	2.68	0.122	0.396	2.05	48.15	0.771	−0.106
Częstochowa	419	160	256	16.82	2.55	0.055	0.220	2.11	57.44	0.776	−0.126
Szczecin	467	301	417	12.34	2.54	0.059	0.042	2.47	34.55	0.794	−0.004
Gdańsk	493	262	458	16.14	2.61	0.132	0.132	2.30	40.52	0.804	−0.058
Wrocław	526	293	637	12.52	2.78	0.147	0.286	2.24	50.83	0.738	0.048
Poznań	532	261	577	14.94	2.72	0.136	0.194	2.47	44.87	0.760	0.160
Białystok	559	90	285	11.93	2.76	0.032	0.004	2.00	62.55	0.682	−0.076
Kraków	940	327	738	21.52	2.52	0.106	0.266	2.71	47.53	0.779	0.212
Łódź	1023	294	800	17.10	2.83	0.065	0.070	2.45	59.79	0.721	0.073
Warszawa	1530	494	1615	19.62	2.88	0.149	0.340	2.42	90.93	0.691	0.093
GOP	2811	1412	2100	19.76	2.83	0.085	0.208	2.90	68.42	0.760	−0.039

IV. DEGREE DISTRIBUTIONS

A. Degree distribution in the space L

Figure 4 shows typical plots for degree distribution in the space L. One can see that there is a slightly better fit to the linear behavior in the log-log description as compared to semi-logarithmic plots. Points $k=1$ are very peculiar since they correspond to the routes' ends. The remaining parts of the degree distributions can be approximately described by a power law,

$$p(k) \sim k^{-\gamma}, \tag{1}$$

although the scaling cannot be seen very clearly and it is limited to less than one decade. Pearson correlation coefficients of the fit to Eq. (1) range from 0.95 to 0.99. Observed characteristic exponents γ are between 2.4 and 4.1 (see Table II), with the majority (15 out of 22) $\gamma > 3$. The values of exponents γ are significantly different from the value $\gamma = 3$ which is characteristic for the Barabási-Albert model of evolving networks with preferential attachment [3], and one can suppose that a corresponding model for transport network evolution should include several other effects. In fact, various models taking into account the effects of fitness, atractiveness, accelerated growth and aging of vertices [26], or deactivation of nodes [27,28] lead to γ from a wide range of values $\gamma \in \langle 2, \infty)$. One should also notice that networks with a characteristic exponent $\gamma > 4$ are considered topologically close to random graphs [25]—the degree distribution is very narrow—and a difference between power law and exponential behavior is very subtle (see the Southern California

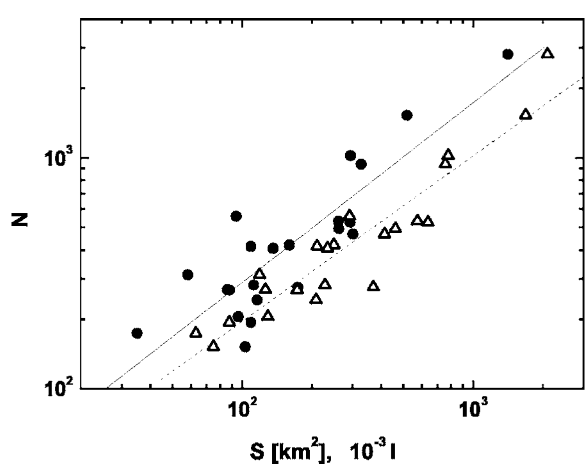

FIG. 3. Log-log plot of the dependence of the number of nodes N on surface S (circles) and population I (triangles).

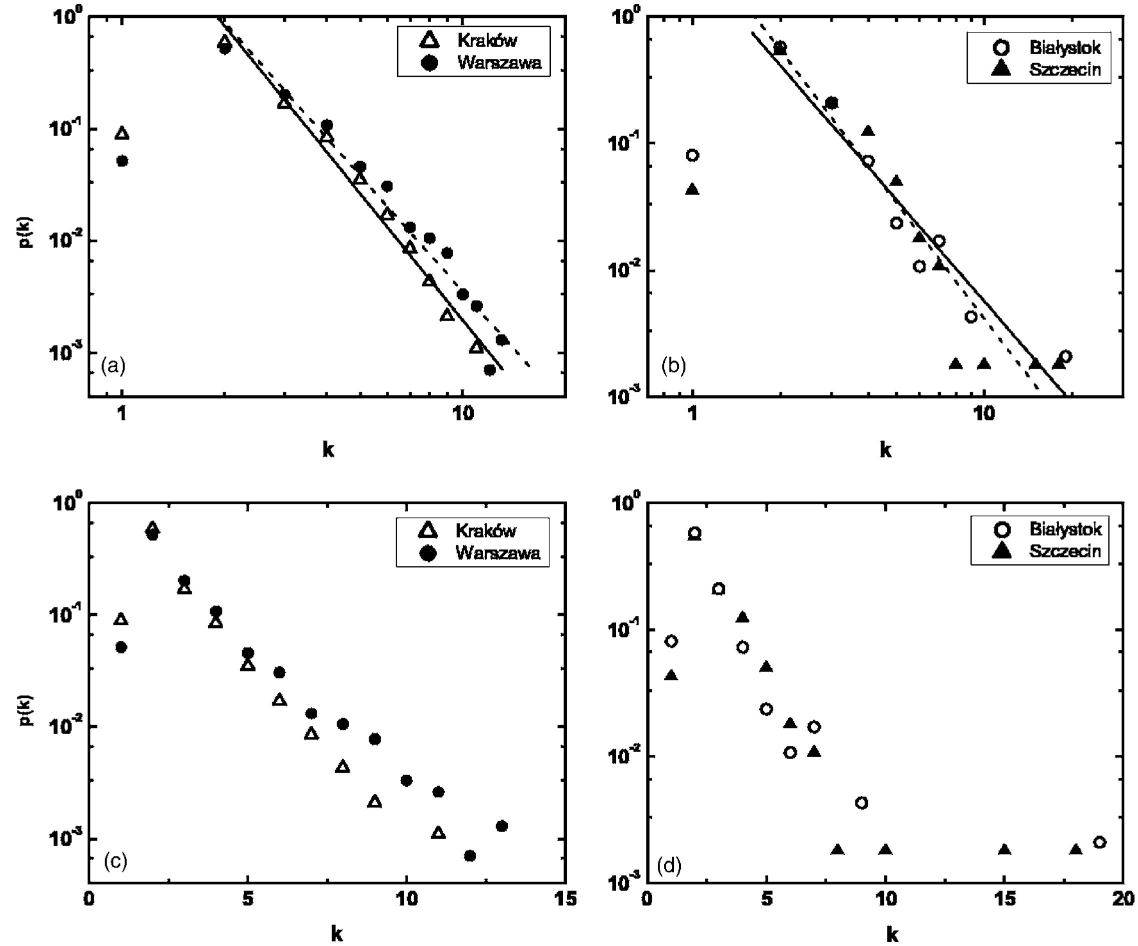

FIG. 4. Degree distributions in the space L for four chosen cities. Plots (a) and (b) show the distributions in log-log scale while plots (c) and (d) show them in the semilog scale.

power grid distribution in [2] presented as a power law with $\gamma \approx 4$ and in [9] depicted as a single-scale cumulative distribution).

Degree distributions obtained for airport networks are also power law (ANC, ANI) or power law with an exponential cutoff (in the case of WAN). For all those systems, the exponent γ is in the range of 2.0–2.2, which differs significantly from the considered PTS in Poland, however one should notice that airport networks are much less dependent on the two-dimensional space, as in the case of PTS. This effect is also seen when analyzing average connectivity ($\langle k \rangle = 5.77$ for ANI, $\langle k \rangle = 9.7$ for WAN, and $\langle k \rangle = 12-14$ for ANC depending on the day of the week the data have been collected).

Let us notice that the number of nodes of degree $k=1$ is smaller as compared to the number of nodes of degree $k=2$ since $k=1$ nodes are the ends of transport routes. The maximal probability observed for nodes with degree $k=2$ means that a typical stop is directly connected to two other stops. Still some nodes (hubs) can have a relatively high degree value (in some cases above 10), but the number of such vertices is very small.

B. Degree distribution in the space *P*

In our opinion, the key structure for the analysis of PTS is *routes* and not single bus/tramway stops. Therefore, we especially take into consideration the degree distribution in the space P.

To smooth large fluctuations, we use here the cumulative distribution $P(k)$ [5] according to the formula

$$P(k) = \int_{k}^{k_{\max}} p(k) dk. \tag{2}$$

The cumulative distributions in the space P for eight chosen cities are shown in Fig. 5. Using the semilog scale, we observe an exponential character of such distributions,

$$P(k) = A e^{-\alpha k}. \tag{3}$$

As is well known [3], the exponential distribution (3) can occur for evolving networks when nodes are attached completely *randomly*. This suggests that a corresponding evolution of public transport in the space P possesses an accidental character that can appear because of a large number of factors responsible for urban development. However, in the next sections we show that other network's parameters such as clustering coefficients or degree-degree correlations calculated for PTS are much larger as compared to corresponding values of randomly evolving networks analyzed in [3].

In the case of IRN [15], the degree distribution in the space P has also maintained the single-scale character $P(k)$

TABLE II. Coefficients γ and α with their fitting errors $\Delta\gamma$ and $\Delta\alpha$. Fitting to the scaling relation (1) has been performed at whole ranges of degrees k while fitting to Eq. (3) has been performed at approximately half of the available ranges to exclude large fluctuations occurring for higher degrees (see Fig. 5).

City	γ	$\Delta\gamma$	α	$\Delta\alpha$
Piła	2.86	0.17	0.0310	0.0006
Bełchatów	2.8	0.4	0.030	0.002
Jelenia Góra	3.0	0.3	0.038	0.001
Opole	2.29	0.23	0.0244	0.0004
Toruń	3.1	0.4	0.0331	0.0006
Olsztyn	2.95	0.21	0.0226	0.0004
Gorzów Wlkp.	3.6	0.3	0.0499	0.0009
Bydgoszcz	2.8	0.3	0.0384	0.0004
Radom	3.1	0.3	0.0219	0.0004
Zielona Góra	2.68	0.20	0.0286	0.0003
Gdynia	3.04	0.2	0.0207	0.0003
Kielce	3.00	0.15	0.0263	0.0004
Częstochowa	4.1	0.4	0.0264	0.0004
Szczecin	2.7	0.3	0.0459	0.0006
Gdańsk	3.0	0.3	0.0304	0.0006
Wrocław	3.1	0.4	0.0225	0.0002
Poznań	3.6	0.3	0.0276	0.0003
Białystok	3.0	0.4	0.0211	0.0002
Kraków	3.77	0.18	0.0202	0.0002
Łódź	3.9	0.3	0.0251	0.0001
Warszawa	3.44	0.22	0.0127	0.0001
GOP	3.46	0.15	0.0177	0.0002

$\sim e^{-\alpha k}$ with the characteristic exponent $\alpha=0.0085$. The values of average connectivity in the studies of MBTA ($\langle k\rangle=27.60$) and the U-Bahn in Vienna ($\langle k\rangle=20.66$) are smaller than for the considered systems in Poland, however one should notice that the sizes of the networks in MBTA and Vienna are also smaller.

C. Average degree and average square degree

Taking into account the normalization condition $P(k_{\min})=1$, we get the following equations for the average degree and the average square degree:

$$\langle k\rangle = \frac{k_{\min}e^{-\alpha k_{\min}} - k_{\max}e^{-\alpha k_{\max}}}{e^{-\alpha k_{\min}} - e^{-\alpha k_{\max}}} + \frac{1}{\alpha}, \quad (4)$$

$$\langle k^2\rangle = \frac{k_{\min}^2 e^{-\alpha k_{\min}} - k_{\max}^2 e^{-\alpha k_{\max}}}{e^{-\alpha k_{\min}} - e^{-\alpha k_{\max}}} + \frac{2(k_{\min}e^{-\alpha k_{\min}} - k_{\max}e^{-\alpha k_{\max}})}{\alpha(e^{-\alpha k_{\min}} - e^{-\alpha k_{\max}})} + \frac{2}{\alpha^2}. \quad (5)$$

Dropping all terms proportional to $e^{-\alpha k_{\max}}$, we receive simplified equations for $\langle k\rangle$ and $\langle k^2\rangle$,

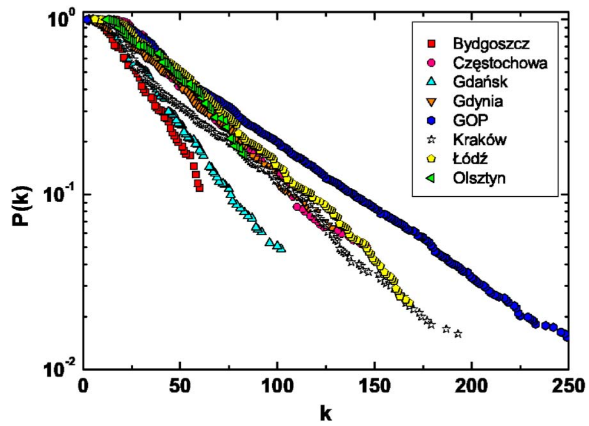

FIG. 5. (Color online) $P(k)$ distribution in the space P for eight chosen cities.

$$\langle k\rangle \approx k_{\min} + \frac{1}{\alpha}, \quad (6)$$

$$\langle k^2\rangle \approx k_{\min}^2 + \frac{2k_{\min}}{\alpha} + \frac{2}{\alpha^2}. \quad (7)$$

Since values of $k_{\min}$ range between 3 and 16 and they are independent of network sizes N as well as observed exponents α, we have approximated $k_{\min}$ in Eqs. (6) and (7) by an average value (mean arithmetical value) for considered networks, $\langle k_{\min}\rangle\approx 8.5$. In Figs. 6 and 7, we present a comparison between the real data and values calculated directly from Eqs. (6) and (7).

V. PATH LENGTH PROPERTIES

A. Path length distributions

Plots presenting path length distributions $p(l)$ in spaces L and P are shown at Figs. 8 and 9, respectively. The data fit well to the asymmetric, unimodal functions. In fact, for all systems a fitting by the Lavenberg-Marquardt method has been made using the following trial function:

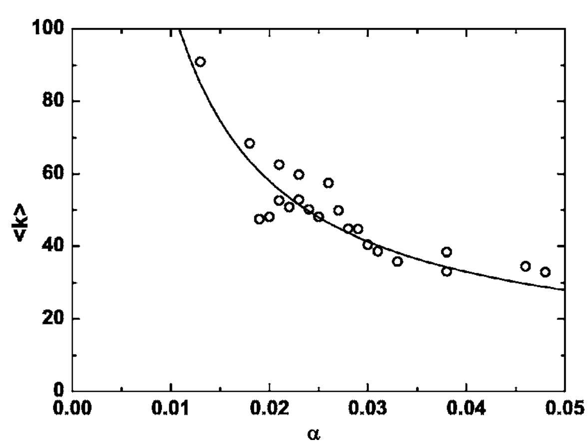

FIG. 6. $\langle k\rangle$ as a function of α. Circles are real data values, while the line corresponds to Eq. (6).

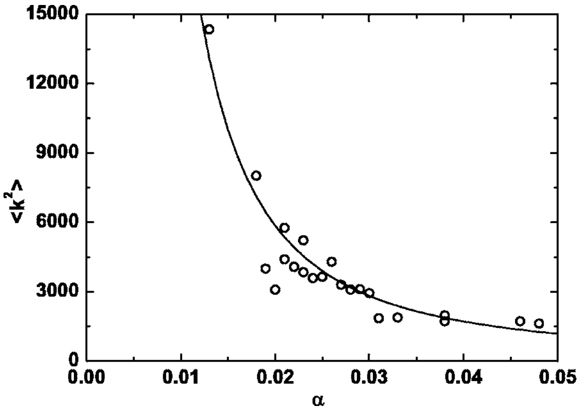

FIG. 7. $\langle k^2\rangle$ as a function of α. Circles are real data values, while the line corresponds to Eq. (7).

$$p(l) = Ale^{-Bl^2+Cl}, \quad (8)$$

where A, B, and C are fitting coefficients.

Insets in Figs. 8 and 9 present a comparison between the experimental results of $\langle l\rangle$ and corresponding mean values obtained from Eq. (8). One can observe a very good agreement between averages from Eq. (8) and experimental data.

The agreement is not surprising in the case of Fig. 9 since the number of fitted data points to curve (8) is quite small, but it is more prominent for Fig. 8.

Ranges of distances in the space L are much broader as compared to corresponding ranges in the space P, which is a natural effect of topology differences. It follows that the average distance in the space P is much smaller ($\langle l\rangle < 3$) than in the space L. The characteristic length 3 in the space P means that in order to travel between two different points, one needs on average no more than two transfers. Other PTS also share this property. Depending on the system size, the following results have been obtained: $\langle l\rangle = 1.81$ (MBTA), $\langle l\rangle = 1.86$ (Vienna), and $\langle l\rangle = 2.16$ (IRN). In the case of the space L, the network MBTA with its average shortest path length $\langle l\rangle = 15.55$ is placing itself among the values acquired for PTS in Poland. The average path length in airport networks is very small: $\langle l\rangle = 2.07$ for ANC, $\langle l\rangle = 2.26$ for ANI, and $\langle l\rangle = 4.37$ for WAN. However, because flights are usually *direct* (i.e., there are no stops between two cities), one sees immediately that the idea of the space L does not apply to airport networks—they already have an intrinsic topology similar to the space P. The average shortest path lengths $\langle l\rangle$ in those systems should be relevant to values obtained for other networks after a transformation to the space P.

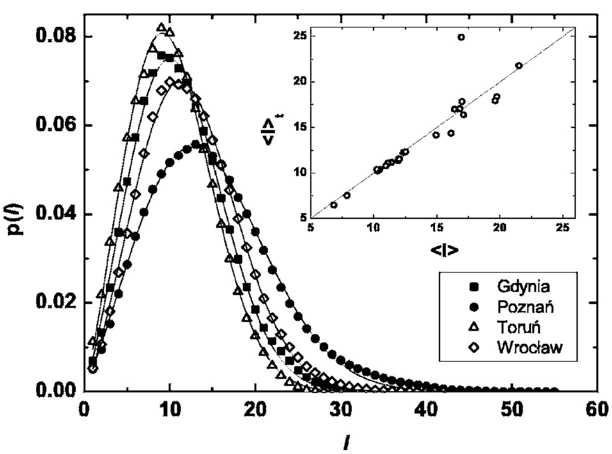

FIG. 8. Fitted path length distribution in the space L.

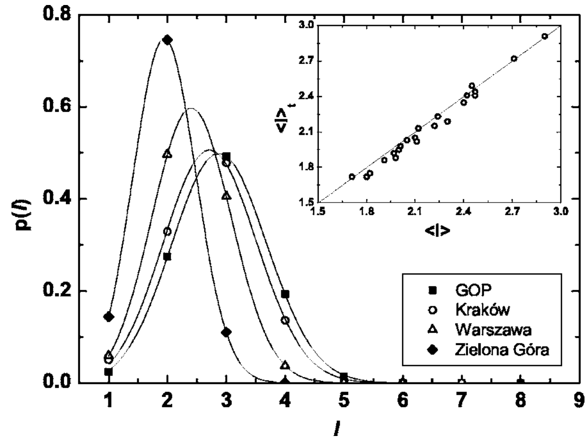

FIG. 9. Fitted path length distribution in the space P.

The shape of path length distribution can be explained in the following way: because transport networks tend to have an inhomogeneous structure, it is obvious that distances between nodes lying on the suburban routes are quite large and such a behavior gives the effect of observed long tails in the distribution. On the other hand, the shortest distances between stops not belonging to suburban routes are more random and they follow the Gaussian distribution. A combined distribution has an asymmetric shape with a long tail for large paths.

We need to stress that internode distances calculated in the space L are much smaller as compared to the number of network nodes (see Table I). Simultaneously clustering coefficients c_L are in the range $\langle 0.03, 0.15\rangle$. Such a behavior is typical for small-world networks [1] and the effect has been also observed in other transport networks [8,14–16,20,21]. The small-world property is even more visible in the space P where average distances are between $\langle 1.80, 2.90\rangle$ and the clustering coefficient c_P ranges from 0.682 to 0.847, which is similar to MBTA ($c=0.93$), Vienna ($c=0.95$), or IRN ($c=0.69$).

B. Path length as a function of product k_ik_j

In [29], an analytical estimation of average path length $\langle l\rangle$ in random graphs has been found. It has been shown that $\langle l\rangle$ can be expressed as a function of the degree distribution. In fact, the mean value for shortest path length between i and j can be written as [29]

$$l_{ij}(k_i,k_j) = \frac{-\ln k_ik_j + \ln(\langle k^2\rangle - \langle k\rangle) + \ln N - \gamma}{\ln(\langle k^2\rangle/\langle k\rangle - 1)} + \frac{1}{2}, \quad (9)$$

where $\gamma = 0.5772$ is the Euler constant.

Since PTS are not random graphs and large degree-degree correlation in such networks exist, we have assumed that Eq.

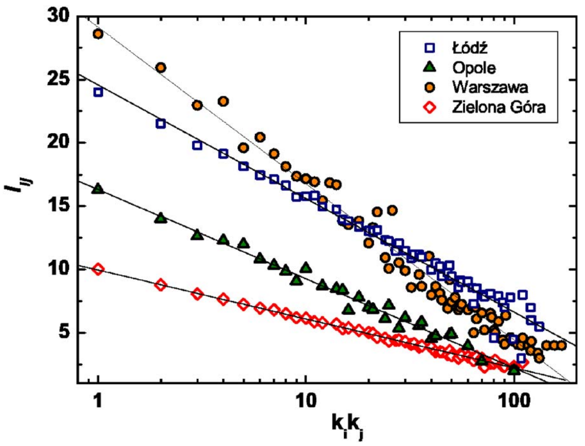

FIG. 10. (Color online) Dependence of l_{ij} on k_ik_j in the space *L*.

(9) is only partially valid and we have written it in a more general form [30–33],

$$\langle l_{ij}\rangle = A - B \log k_i k_j. \tag{10}$$

To check the validity of Eq. (10), we have calculated the values of the average path length between l_{ij} as a function of their degree product k_ik_j for all systems in the space *L*. The results are shown in Fig. 10, which confirms the conjunction (10). A similar agreement has been received for the majority of investigated PTS. Equation (10) can be justified using a simple model of random graphs and a generating function formalism [34] or a branching tree approach [30–33]. In fact, the scaling relation (10) can also be observed for several other real-world networks [30–33].

It is useless to examine the relation (10) in the space *P* because corresponding sets l_{ij} consist usually of three points only.

VI. CLUSTERING COEFFICIENT

We have studied clustering coefficients c_i defined as a probability that two randomly chosen neighbors of node *i* possess a common link.

The clustering coefficient of the whole network seems to depend weakly on parameters of the space *L* and of the space *P*. In the first case, its behavior with regard to network size can be treated as fluctuations, when in the second one it is possible to observe a small decrease of *c* along with the networks size (see Table I). We shall discuss only properties of the clustering coefficients in the space *P* since the data in the space *L* are meaningless.

It has been shown in [15] that the clustering coefficient in IRN in the space *P* decays linearly with the logarithm of degree for large *k* and is almost constant (and close to unity) for small *k*. In the considered PTS, we have found that this dependency can be described by a power law (see Fig. 11),

$$c(k) \sim k^{-\beta}. \tag{11}$$

Such a behavior has been observed in many real systems with hierarchical structures [37,38]. In fact, one can expect

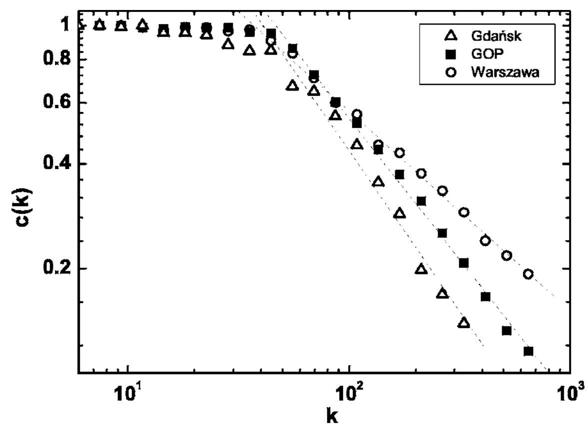

FIG. 11. $c(k)$ for Gdańsk (triangles), GOP (squares), and Warszawa (circles). Dashed lines are fits to Eq. (11) with the following exponents: Gdańsk, $\beta=0.93\pm0.05$; GOP, $\beta=0.81\pm0.02$; and Warszawa, $\beta=0.57\pm0.01$. All data are logarithmically binned with the power of 1.25.

that PTS should consist of densely connected modules linked by longer paths.

The observed values of exponents β are in the range $\beta \in \langle 0.54, 0.93\rangle$. This can be explained using a simple example of a *star* network: suppose that the city transport network is a star consisting of *n* routes with *L* stops each. Node *i*, at which all *n* routes cross, is a vertex that has the highest degree in the network. We do not allow any other crossings among those *n* routes in the whole system. It follows that the degree of node *i* is $k_i=n(L-1)$ and the total number of links among the nearest neighbors of *i* is $E_i=n(L-1)(L-2)/2$. In other words, the value of the clustering coefficient for the node with the maximum degree is

$$c(k_{\max}) = \frac{2E_i}{k_i(k_i-1)} = \frac{L-2}{n(L-1)-1}, \tag{12}$$

where $k_{\max}=n(L-1)$. It is obvious that the minimal degree in the network is $k_{\min}=L-1$ and this corresponds to the value $c(k_{\min})=1$. Using these two points and assuming that we have a power-law behavior, we can express β as

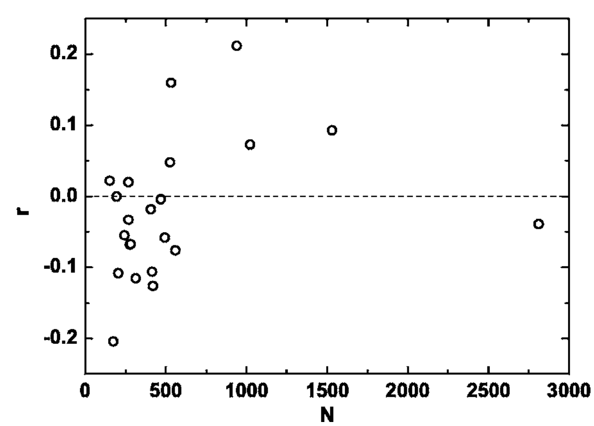

FIG. 12. The assortativity coefficient *r* in the space *P* as a function of *N*.

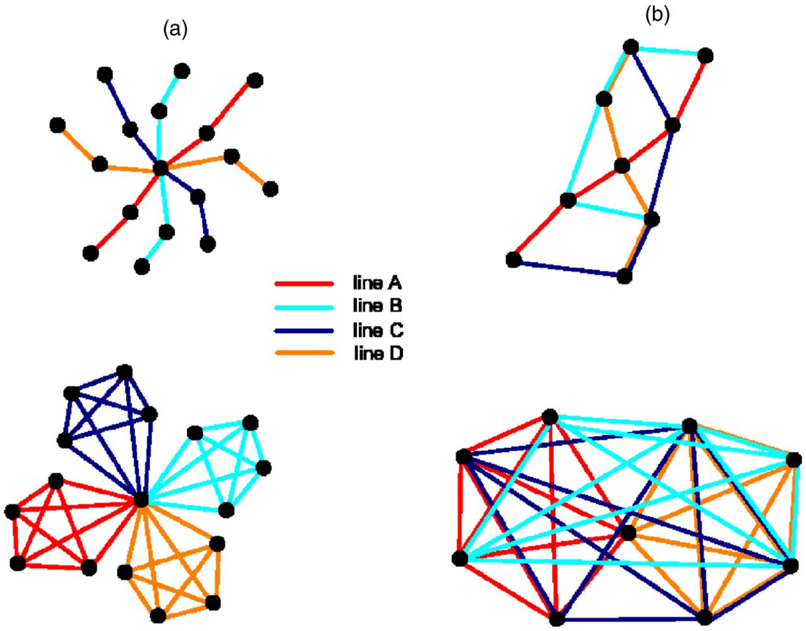

FIG. 13. (Color online) Crossing of four routes of five stops each. (a) In the *star* example, there is only one *hub* and the assortativity coefficient is equal to $r=-0.25$ according to Eq. (15). In case (b), a few hubs exist due to a multiple crossing of routes and $r=-0.19$. Upper diagrams, the space L; lower diagrams, the space P.

$$\beta = -\frac{\ln c(k_{\max}) - \ln c(k_{\min})}{\ln k_{\max} - \ln k_{\min}} = -\frac{\ln \dfrac{L-2}{n(L-1)-1}}{\ln n}. \quad (13)$$

Because $n(L-1) \gg 1$ and $L-1 \approx L-2$, we have $\beta \approx 1$.

In real systems, the value of the clustering coefficient of the highest degree node is larger than in Eq. (12) due to multiple crossings of routes in the whole network, which leads to a decrease of the exponent β (see Fig. 11). This decrease is also connected to the presence of degree-degree correlations (see the next section).

VII. DEGREE-DEGREE CORRELATIONS

To analyze degree-degree correlations in PTS, we have used the assortativity coefficient r, proposed by Newman [35], that corresponds to the Pearson correlation coefficient [36] of the node degrees at the end points of the link,

$$r = \frac{\sum_i j_i k_i - \frac{1}{M}\sum_i j_i \sum_i k_i}{\sqrt{\sum_i j_i^2 - \frac{1}{M}(\sum_i j_i)^2}\sqrt{\sum_i k_i^2 - \frac{1}{M}(\sum_i k_i)^2}}, \quad (14)$$

where M is the number of pairs of nodes (twice the number of edges), j_i, k_i are the degrees of vertices at both ends of the ith pair, and the index i goes over all pairs of nodes in the network.

Values of the assortativity coefficient r in the space L are independent of the network size and are always positive (see Table I), which can be explained in the following way: there is a little number of nodes characterized by high values of degrees k and they are usually linked among themselves. The majority of the remaining links connect nodes of degree $k=2$ or $k=1$, because $k=2$ is an overwhelming degree in networks.

Similar calculations performed for the space P lead to completely different results (Fig. 12). For small networks, the correlation parameter r is negative and it grows with N, becoming positive for $N \gtrsim 500$. The dependence can be explained as follows: small towns are described by star structures and there are only a few *doubled routes*, so in this space a lot of links between vertices of small and large k exist. Using the previous example of a star network and taking into account that the degree of the central node is equal to $k_c = n(L-1)$, the degree of any other node is $k_o = L-1$. After some algebra we obtain the following expression for the assortativity coefficient of such a star network:

$$r = -\frac{1}{L-1}. \quad (15)$$

Let us notice that the coefficient r is independent of the number of crossing routes and is always a negative number.

On the contrary, in the large cities there are lots of connections between nodes characterized by large k (transport hubs) and there is a large number of routes crossing in more

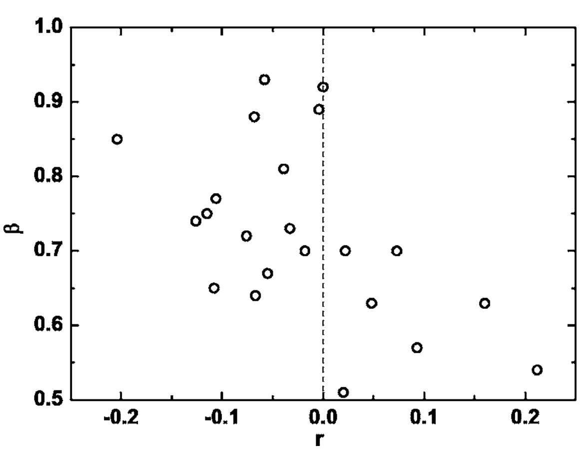

FIG. 14. β coefficient [see Eq. (11)] as a function of the assortativity coefficient r in the space P.

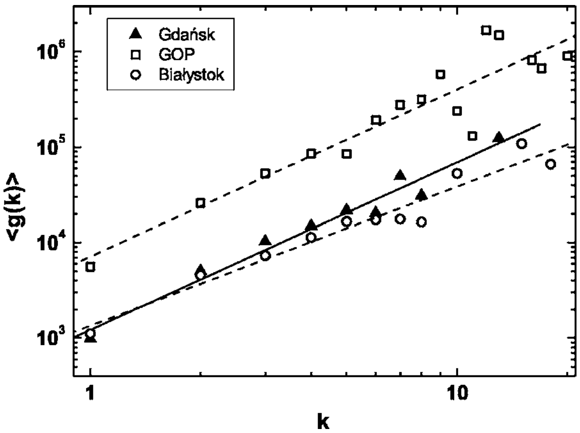

FIG. 15. The average betweenness $\langle g \rangle$ as a function of k in the space L for three chosen cities.

than one point (see Fig. 13). It follows that the coefficient r can be positive for such networks. A strange behavior for the largest network (GOP) can be explained as an effect of its peculiar structure: the system is a conglomerate of many towns rather than a single city. Thus, the value of r is lowered by single links between the subsets of this network.

In Fig. 14, we show coefficients β as a function of r in the space P. One can see that in general, positive values of the assortativity coefficient correspond to lower values of β, being an effect of the existence of several links between hubs in the networks.

The reported values of assortativity coefficients in other transport networks have been negative ($r=-0.402$ for ANI [20] and $r=-0.033$ for IRN [15]), and since these systems are of the size $N<600$ they are in agreement with our results.

VIII. BETWEENNESS

The last property of PTS examined in this work is betweenness [39], which is the quantity describing the "importance" of a specific node according to the following equation [40]:

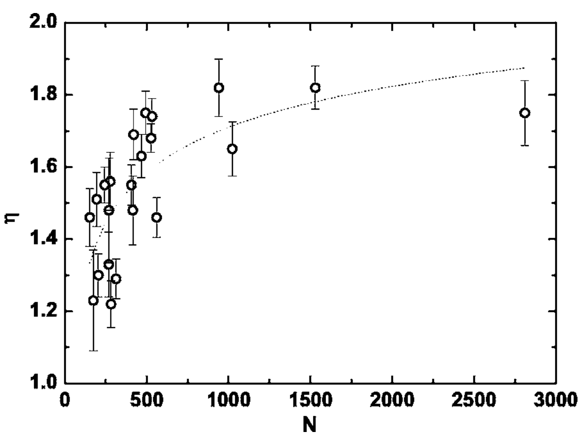

FIG. 16. η coefficient as a function of network size N.

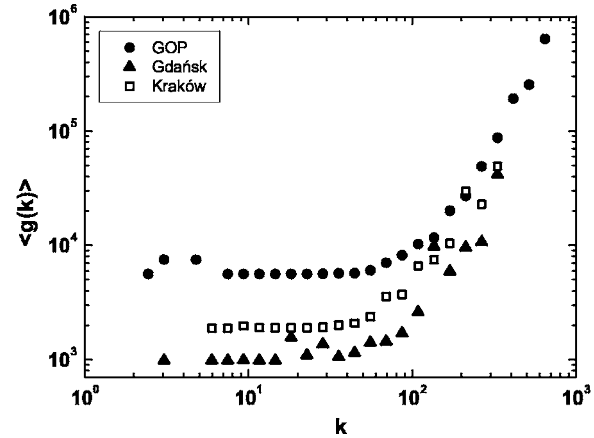

FIG. 17. The average betweenness $\langle g \rangle$ as a function of k in the space P for three chosen cities.

$$g(i) = \sum_{j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}, \tag{16}$$

where σ_{jk} is a number of the shortest paths between nodes j and k, while $\sigma_{jk}(i)$ is a number of these paths that go through the node i.

A. Betweenness in the space *L*

Figure 15 shows the dependence of the average betweenness $\langle g \rangle$ on node degree calculated using the algorithm proposed in [41] (see also [42]). The data in Fig. 15 fit well to the scaling relation

$$g \sim k^{\eta} \tag{17}$$

observed in Internet Autonomous Systems [43], coauthorship networks [44], and the BA model or Erdős-Rényi random graphs [40].

The coefficient η is plotted in Fig. 16 as a function of network size. One can see that η is getting closer to 2 for large networks. Since it has been shown that there is $\eta=2$ for random graphs [40] with Poisson degree distribution, it can be suggested that large PTS are more random than small ones. Such an interpretation can also be made from Table II, where larger values of the exponent γ are observed for large cities.

B. Betweenness in the space *P*

The betweenness as a function of node degree k in the space P is shown in Fig. 17. One can see large differences between Figs. 15 and 17. In the space P there is a saturation of $\langle g \rangle$ for small k, which is a result of the existence of the suburban routes, while the scale-free behavior occurs only for larger k. The saturation value observed in the limit of small k is given by $\langle g(k_{\min}) \rangle = 2(N-1)$ and the length of the saturation line increases with the mean value of a single route's length observed in a city.

IX. CONCLUSIONS

In this study, we have collected and analyzed data for public transport networks in 22 cities that make up over 25%

of the population in Poland. The sizes of these networks range from $N=152$ to 2881. Using the concept of different network topologies, we show that in the space L, where distances are measured in numbers of passed bus/tramway stops, the degree distributions are approximately given by power laws with $\gamma=2.4-4.1$, while in the space P, where distances are measured in numbers of transfers, the degree distribution is exponential with characteristic exponents $\alpha=0.013-0.050$. Distributions of paths in both topologies are approximately given by a function $p(l)=Ale^{-Bl^2+Cl}$. Small-world behavior is observed in both topologies, but it is much more pronounced in space P where the hierarchical structure of the network is also deduced from the behavior of $c(k)$. The assortativity coefficient measured in the space L remains positive for the whole range of N while in the space P it changes from negative values for small networks to positive values for large systems. In the space L, distances between two stops are linear functions of the logarithm of their degree products.

Many of our results are similar to features observed in other works regarding transportation networks: underground, railway, or airline systems [8,12–21,23]. All such networks tend to share small-world properties and show strong degree-degree correlations that reveal the complex nature of those structures.

ACKNOWLEDGMENTS

The work was supported by the EU Grant *Measuring and Modelling Complex Networks Across Domains—MMCOMNET* (Grant No. FP6-2003-NEST-Path-012999), by the State Committee for Scientific Research in Poland (Grant No. 1P03B04727), and by a special Grant of Warsaw University of Technology.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] A.-L. Barabási, R. Albert, and H. Joeng, Physica A **272**, 173 (1999).
[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[5] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[6] J. F. F. Mendes, S. N. Dorogovtsev, and A. F. Ioffe, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[7] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004).
[8] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[9] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[10] R. Albert, I. Albert, and G. L. Nakarado, Phys. Rev. E **69**, 025103(R) (2004).
[11] P. Crucitti, V. Latora, and M. Marchiori, Physica A **338**, 92 (2004).
[12] M. Marchiori and V. Latora, Physica A **285**, 539 (2000).
[13] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).
[14] V. Latora and M. Marchiori, Physica A **314**, 109 (2002).
[15] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, Phys. Rev. E **67**, 036106 (2003).
[16] K. A. Seaton and L. M. Hackett, Physica A **339**, 635 (2004).
[17] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, Proc. Natl. Acad. Sci. U.S.A. **102**, 7794 (2005).
[18] R. Guimerà and L. A. N. Amaral, Eur. Phys. J. B **38**, 381 (2004).
[19] A. Barrat, M. Barthélémy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[20] W. Li and X. Cai, Phys. Rev. E **69**, 046106 (2004).
[21] G. Bagler, e-print cond-mat/0409773.
[22] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).
[23] C. von Ferber, Yu. Holovatch, and V. Palchykov, Condens. Matter Phys. **8**, 225 (2005).
[24] Data on population and city surfaces have been taken from the official site of the Polish National Central Statistical Office (http://www.stat.gov.pl/bdrpuban/ambdr.html). One should mention here that S and I for GOP (Upper-Silesian Industry Area) are the sum of the values for several towns GOP consists of.
[25] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **91**, 168701 (2003).
[26] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 4, 1079 (2002).
[27] A. Vázquez, M. Boguñá, Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **67**, 046111 (2003).
[28] K. Klemm and V. M. Eguíluz, Phys. Rev. E **65**, 036123 (2002).
[29] A. Fronczak, P. Fronczak, and J. A. Hołyst, Phys. Rev. E **68**, 046126 (2003).
[30] J. A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, and K. Suchecki, Physica A **351**, 167 (2005).
[31] J. A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, and K. Suchecki, Phys. Rev. E **72**, 026108 (2005).
[32] J. A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, K. Suchecki, and P. Wójcicki, in *Science of Complex Networks: From Biology to the Internet and WWW; CNET 2004*, edited by J. F. F. Mendes, J. G. Oliveira, F. V. Abreu, A. Povolotsky, and S. N. Dorogovtsev, AIP Conf. Proc. No. 776 (AIP, New York, 2005), p. 69.
[33] J. Sienkiewicz and J. A. Hołyst, Acta Phys. Pol. B **36**, 1771 (2005).
[34] A. E. Motter, T. Nishikawa, and Y.-C. Lai, Phys. Rev. E **66**, 065103(R) (2002).
[35] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[36] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[37] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[38] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).
[39] L. C. Freeman, Sociometry **40**, 35 (1977).

[40] M. Barthélémy, Eur. Phys. J. B **38**, 163 (2004).
[41] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
[42] An algorithm for the calculation of the betweenness centrality that parallels the one presented in [41] had been published in U. Brandes, J. Math. Sociol. **25**, 163 (2001) and before in 2000, see http://www.inf.uni-konstanz.de/brandes/publications
[43] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[44] K.-I. Goh, E. Oh, B. Kahng, and D. Kim, Phys. Rev. E **67**, 017101 (2003).