Physica A 390 (2011) 2601–2610

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Topology of the Erasmus student mobility network

A. Derzsi [a], N. Derzsy [a], E. Káptalan [a], Z. Néda [a,b,*]

[a] Department of Theoretical and Computational Physics, Babeş-Bolyai University, Kogâlniceanu street 1, RO-400080, Cluj-Napoca, Romania

[b] Interdisciplinary Computer Modelling Group, Hungarian University Federation of Cluj, Baba Novac street 23/2, RO-400080, Cluj-Napoca, Romania

ARTICLE INFO

Article history:
Received 11 May 2010
Received in revised form 9 February 2011
Available online 11 March 2011

Keywords:
Large social network
Small-world network
Directed and weighted network

ABSTRACT

The collaboration network generated by the Erasmus student mobilities in the year 2003 is analyzed and modeled. Nodes of this bipartite network are European universities and links are the Erasmus mobilities between these universities. This network is a complex directed and weighted graph. The non-directed and non-weighted projection of this network does not exhibit a scale-free nature, but proves to be a small-word type random network with a giant component. The connectivity data indicates an exponential degree distribution, a relatively high clustering coefficient and a small radius. It can be easily modeled by using a simple configuration model and arguing the exponential degree distribution. The weighted and directed version of the network can also be described by means of simple random network models.

1. Introduction

The dynamics of many social, natural and artificial systems are governed by the underlying network structure. Networks are considered to be the backbone of all complex systems ranging from atomic and molecular levels to systems with global extent. Different approaches were proposed in order to describe the organization of such structures. Despite the clear differences in the governing rules of various types of networks, the idea of Refs. [1,2] to assume random connections among the network nodes provided a simple framework for the description of all these systems. In random graphs the nodes are considered to be similar in their networking potential and only chance influences their degree. As a consequence of this, most of the elements have the same number of connections resulting in a characteristic Poisson-type degree distribution [3]. A different approach — closer to real-life networks — introduced by Granovetter [4] emphasized the stabilizing role of weak ties in social networks. According to this view, social networks are composed of many complete subgraphs (formed by strong connection of the elements) and these subgraphs are linked to each other through several external connections of their elements (weak ties). In order to measure how strongly connected the elements of a network are, Watts and Strogatz [5] introduced the clustering coefficient. They found that many real-world networks have higher clustering coefficient as it can be obtained from the random graph model of Erdős and Rényi and presented a new model that combines the clustering property and the idea of random connectivity in a graph [5,6]. Their work brought into focus the small-world problem of sociology [7,8] and generated extensive research on various networks [9–16]. As a result, clustering was observed in all types of the investigated graphs and thus became a general characteristic of complex networks. Small-world type graphs have high clustering coefficient and small value for the average shortest path length, while random graphs exhibit small clustering coefficient and small average node-to-node distance.

The investigation of many real-world networks pointed out the presence of nodes with a high number of connections. This feature could not be explained by the random graph model or by the model proposed by Watts and Strogatz. Further

* Corresponding author at: Department of Theoretical and Computational Physics, Babeş-Bolyai University, Kogâlniceanu street 1, RO-400080, Cluj-Napoca, Romania Tel.: +40 745 310531; fax: +40 264 591906.

E-mail address: zneda@phys.ubbcluj.ro (Z. Néda).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2011.02.041

it was revealed that the degree distribution of these networks follows a power-law, therefore they have been referred to as scale-free networks. Barabási and Albert [17] elaborated the first successful model for these networks by considering a constant growth of the network and preferential attachment of the new elements to existing nodes in the network. This scale-free model was able to reproduce the power-law degree distribution observed in real-world networks and motivated widespread study on the topology and dynamics of large networks [18–21]. The statistical physics community became also interested in this fascinating research field [22].

Previous results have shown that large social networks often present scale-free properties [23]. Well-known examples in such a sense are:

- The bipartite network of actor collaborations [17]: In this network the nodes are the actors, and two actors are linked if they have ever been cast in the same movie.
- The network of sexual relations of people in Refs. [24,25]: The nodes of this network are adult females and males, and the links are sexual relations among them. The degree distribution of this network was mapped by a questionnaire study, sampling 4781 persons with ages between 18–74. The results are less conclusive than other results obtained on complete databases and objective network construction methods.
- The bipartite network of scientific co-authorship in different research fields [26–28]: Nodes of these networks are scientists and the links are co-authorship relations. Two scientists are linked in these graphs if they have ever co-authored a scientific publication. Based on several large electronic databases, large scientific collaboration graphs in the fields of physics, mathematics, biomedical research and neuroscience have been analyzed.
- The bipartite science-citation network [29]: Nodes are again scientists and directional links are defined by citations in scientific papers. Two scientists are connected if one of them has ever cited the other one. Naturally, this network does not give any information about real social contacts. It offers information only about unidirectional knowledge of professionals and their contribution in a given research field.
- Social commerce networks [30]: The nodes of this network are online shops that create referral hyperlinks which connect them to other online shops.
- Online social networks [31,32]: These networks were studied by analyzing the connectivity pattern of online communities.
- Phone calls, instant messaging and E-mail connection networks [33,34]: The nodes of these networks are users of communication devices connected if they have exchanged messages in the studied time-interval and communication system.

In all the examples from above large social systems were investigated and scale-free network characteristics were found. The dynamics of some of them is driven by preferential attachment, and all of them present large and strongly connected components. Sampling of the considered large-scale social networks is however incomplete in most of the cases. Even the most exhaustive studies, that were conducted on the co-authorship networks using electronic data spanning a decade [28], lack a complete picture on the structure of these networks. The main problem that arise in such studies is the time- and field-restricted nature of the databases. Due to this inconvenience a great part of the existing links are not revealed and the missing data can dramatically alter estimates of network-level statistics [35].

The present work maps another large-scale social network, the Erasmus Student Mobility (ESM) network for the year 2003. It also introduces a simple model to reproduce the observed characteristics, topological and weighted features. Student exchange agreements are made by professors through their personal professional contacts. Therefore, it is believed that the Erasmus student mobility network indirectly maps the collaboration network of academic staff working in European universities. The great advantage of this network is that processing the data on the 2003 mobility is enough to draw a complete picture on this collaboration graph. This is possible because each year usually all Erasmus connections are used. The database is complete in such sense, and we believe that the existing collaboration links are realistically revealed.

Before presenting our results, it is worth mentioning however that the Erasmus mobility network has been already used in mapping the international knowledge flows across five European countries [36]. In that study network analysis techniques were used to test the influence of geographical distance as opposed to sectoral and functional distances. The main entities (nodes in the network) were geographical regions in that work. The present study has however a totally different aim and methodology. First, we consider the more complete network, which extends on all involved European countries, and approach the network on its basic constituent level, on the level of the universities involved in the student exchange. Second, we analyze several variants of this network by taking into account or neglecting the weights of the connections (number of mobilities between two universities) and their direction. Finally, we give a simple model that describes in a satisfactory manner the revealed network structures.

2. The Erasmus student mobility network

The Erasmus (European Region Action Scheme for the Mobility of University Students) Programme, established in 1987, represents a part of the initiatives of the European Commission in higher education. The goal of this program is to encourage and support academic mobility of students and teachers in higher education within the European Union or countries of the European Economic Area. Each year the universities (home institutions) that signed the Erasmus partnership collaboration agreement with other universities offer the possibility for some of their selected students or teachers to make a 3–6 months

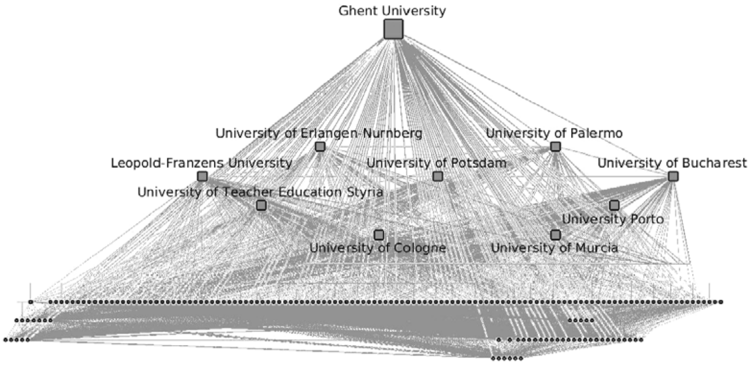

Fig. 1. The k-degree component of the NNESM network for $k = 55$. The 180 nodes of this component are presented in hierarchical arrangement. The most highly connected 10 hubs are indicated by their names.

exchange visit to a partner university (host institution). The students are not supposed to pay tuition fees at the host institute and the period spent abroad is recognized by their home university if objects previously agreed are accomplished. The program is partly financed by the European Union and since its starting it has enjoyed continuous popularity in groups of eligible persons. There are currently over 2 million students from more than 4000 higher education institutions that have already participated in Erasmus mobility. As an example, in 2006 over 150.000 students — almost 1% of the European student population — took part in this program.

The Erasmus student mobility between universities defines a large network: the nodes of the network are the universities involved in the Erasmus student exchange, and the links are student mobility between them. In this network two universities are connected if there is at least one student moving between them. The Erasmus program is based on written collaboration agreements between the universities. These agreements are made through the professional contacts of the professors. The Erasmus network maps thus indirectly these professional contacts, and it is useful for revealing and understanding the scientific information flow or collaboration pathways between university staff in Europe.

The database we studied contains the Erasmus student mobility realized in the year 2003. The codes of the students and the codes of their home and host universities are given in the database. In 2003 there was in total 134 330 students exchange within the Erasmus program and 2333 European universities were involved. From this large database one can construct several networks for the connection topology of the participating universities. Two kinds of directed networks (networks with directed links) can be built: one for the outgoing students and another one for the incoming students. Two undirected networks can be also defined: a restricted one where two universities are connected if there are student exchanges in both directions, and another one which connects universities in case of student exchange in either direction. Additionally to these, for all cases weighted networks can be also constructed. In such networks the strength of each link is proportional with the number of students involved in the given exchange.

Our primary interest in processing the Erasmus exchange network is to reveal the network of professional connections between universities. As a first approach, we assume that there is an active professional connection if at least one Erasmus student travels between two universities. In such manner we are interested mainly in the non-directed network defined by either outgoing or incoming students. From this point of view, the weight of the connection is not of primary importance, since the number of traveling students does not directly measure the strength of the professional connection. The number of students participating in the program is usually influenced by financial and political factors as well. Therefore we will construct first the non-directed and non-weighted network of the IN or OUT connections. In the following we will label this network simply as the NNESM (Non-directed and Non-weighted Erasmus Student Mobility) network.

The NNESM network can be compared with the network presented in Ref. [37] in which physical contact patterns resulting from movements of individuals between specific locations is modeled. The study of this network provides efficient strategies to contain disease outbreaks in urban social networks. In the network of Ref. [37], two locations are connected by a link if there is at least one person moving from one location to the other. As in our NNESM network, the connection of locations is considered without weights of the links. Locations in that case can be taken as a match for universities in our network, while people moving between different locations stand for students in our case.

The graphical representation of the whole NNESM network, due to the large number of connections, is not informative at all. Therefore, to illustrate the structure of this network we plotted in Fig. 1 only the k-degree component of this network in a hierarchical representation, considering $k = 55$. The k-degree component contains those nodes, which has each of them at least k links with the other nodes belonging to this component. A method to construct this component consists in the recursive removal of all nodes from the original NNESM network, until all nodes have at least k connections with the remained ones. Varying the value of k it was found that the largest k value for which nodes still remain in the system is $k = 55$. This is the reason we have plotted this connected component which contains 180 nodes.

Besides the investigation of the NNESM network, we also exploit here some information carried by the direction and intensity of the students' mobility. This is accomplished by analyzing the directed and weighted ESM network, data available

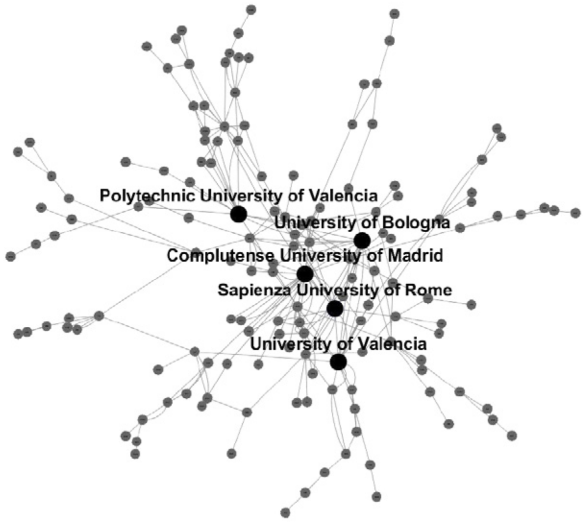

Fig. 2. Part of the weighted ESM network: the figure shows the largest connected component obtained by using links on which there are at least 15 student exchanges in year 2003. This component contains 149 elements, the highlighted nodes are the hubs characterized with the most intense student mobilities.

in the original database. Taking into account the direction of the mobility between pairs of universities, the properties of the network can be determined in terms of IN degree and OUT degree of the nodes. IN degree refers to the number of partner universities that send students to the considered university, and OUT degree refers to the number of universities that accept students from the considered university. Including the number of students traveling on each of these directed links will impose weights for the links. This makes possible the study of the weight distribution and the investigation of the strength of the nodes which represents a measure of centrality. Fig. 2 presents the central portion of the whole weighted network that can be constructed from the student mobility database. In this figure the subgraph formed by those universities that have exchanged at least 15 students between each other in year 2003 is illustrated.

3. Topological properties of the NNESM network

The degree (connection number) distribution of the non-directed and non-weighted ESM network has been studied through a simple degree-rank plot. To construct this plot one first performs a ranking of the nodes according to their degree. The node with the highest degree will have rank number one, and the node with the lowest degree number will have the highest rank. Nodes that have the same number of links will be ordered in a random sequence, and their rank will be assigned increasingly in this order. The degree-rank plot is obtained then by plotting for each node its rank number r as a function of the degree k. It is easy to realize that the degree-rank function $r(k)$ yields the cumulative degree distribution function $F(k)$, up to a $1/N$ proportionality factor (N the number of nodes in the network):

$$F(k) = \frac{1}{N} r(k). \tag{1}$$

Using the cumulative degree distribution instead of the simple degree distribution $p(k)$ which describes the probability of finding a node with a given degree k, results in a much smoother curve. Taking into account the discrete nature of the degree, and the fact that possible values of the degree are natural numbers, the relation between $F(k)$ and $p(k)$ is:

$$F(k) = \sum_{x=k}^{k_{\max}} p(x) \approx \int_k^{\infty} p(x)\mathrm{d}x \tag{2}$$

where $k_{\max}$ is the highest degree number found in the network. The degree-rank plot obtained for the NNESM network is given in Fig. 3. Using a semi-log plot, one finds a fair linear trend for the highly connected universities, suggesting an exponential tail for the cumulative distribution:

$$F(k) = \exp\left(-\frac{k}{\lambda}\right). \tag{3}$$

In the case when the exponential distribution extends on the whole $[0, \infty)$ interval, the λ constant gives the mean degree of the nodes ($\lambda = \langle k \rangle$). This is not the case however for the considered NNESM network. It results from the exponential nature of the cumulative distribution that the tail of the degree distribution is also exponential:

$$p(k) = -\frac{\partial F(k)}{\partial k} = \frac{1}{\lambda} \exp\left(-\frac{k}{\lambda}\right). \tag{4}$$

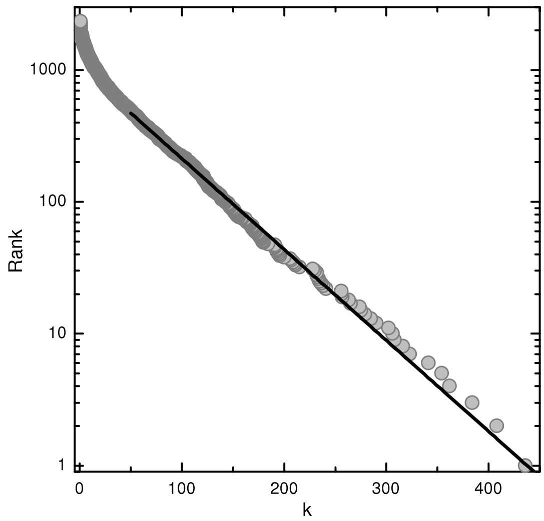

Fig. 3. Degree-rank plot on semi-log axis for the NNESM network (filled circles). The thick continuous line fits the exponential tail generated by the highly connected universities (more than 50 connections).

Recent studies have shown that large-scale social networks are usually scale-free ones [17,24,26–29,33]. The NNESM network is thus a peculiar example where this observation is not true, and instead of a power-law type tail of the degree distribution an exponential one is obtained. We will see in the following sections that the exponential nature of the tail of the degree distribution has its roots in the exponential tail of the size distribution–function of the universities.

The global clustering coefficient C_g of the NNESM network is 0.183. This coefficient is defined as $C_g = (W_c)/(W_c + W_o)$, where W_c denotes the number of closed triplets and W_o is the number of open triplets in the net. A triplet is formed by any three nodes in the network that are connected by either two (open triplet) or three (closed triplet) links. The global clustering coefficient obtained for the whole NNESM network is quite large in comparison with an Erdős–Rényi type random network with the same number of nodes ($N = 2333$) and links ($L = 37\,492$): $C_{E-R} = (2L)/[N(N-1)] = 0.0138$. The local clustering coefficient C_l of the whole NNESM network is 0.292. This coefficient is given by the average value of the C_i local clustering coefficients of the nodes, defined as $C_i = (2W_i)/[k_i(k_i - 1)]$, where W_i is the number of linked neighbors of node i and k_i is the degree of node i. The relatively high clustering coefficients and the non-Poissonian degree distribution suggest that the NNESM network is not an Erdős–Rényi type simple random graph. High clustering coefficients are typical for the case of social networks originating from the existence of strong transitivity in human relationships. Social networks usually exhibit the validity of the idea that the friend of one's friend is likely to be one's friend, thus ensuring the presence of a high number of closed triplets in the system. High values of the clustering coefficients — both for global and local coefficients — can be also found for instance in networks of co-authorship in the field of physics, mathematics and biology [26,38,39], in the network of film actors [5] as well as in the network of company directors [40].

The ESM network is not a one-component graph. It exhibits several connected components, one of them being a giant component (containing 99% of all the nodes) characterized by an average distance of 2.91 between its nodes.

The dynamics of the network realized by the Erasmus program can be investigated by following the network generated in consecutive years. Unfortunately this interesting study cannot be realized here since data for other years is not available to us. The present study has to limit itself on exploiting the information carried by the data for only one year. The investigation of the presence of selective linking (assortative mixing) [41,42] — characteristic of many social networks — could yield further insight into the topology of the NNESM network. For the NNESM network several types of assortative mixing can be studied. Selective linking as a function of the nation or study language of the participating universities is not informative at all, since these types of links are strictly regulated by the Erasmus program principles and it is not a result of self-organization. Selective linking that is relevant for the NNESM network would be for example a closed interconnectivity of the highly connected universities. If this happened, it would mean that universities with a large number of connections prefer to connect with universities that have also a large number of links.

In Fig. 4(a) the mean degree of the neighbors of a node is plotted as a function of the degree of the node (filled symbols), revealing no evidence for a clear increasing trend. The figure also shows the averages (continuous line) taken on a moving interval with fixed length on the horizontal axes. The plot suggests a rather random attachment rule in establishing the connections in the NNESM network, despite the fact that the obtained degree distribution is not characteristic for a simple Erdős–Rényi type random network. Fig. 4(b) shows the distribution of the local clustering coefficients for the nodes of the NNESM network as a function of the node degree. Averages of the local clustering coefficients taken on a moving horizontal interval indicates again no clear trend. These results suggest that such kind of assortative linking is not present.

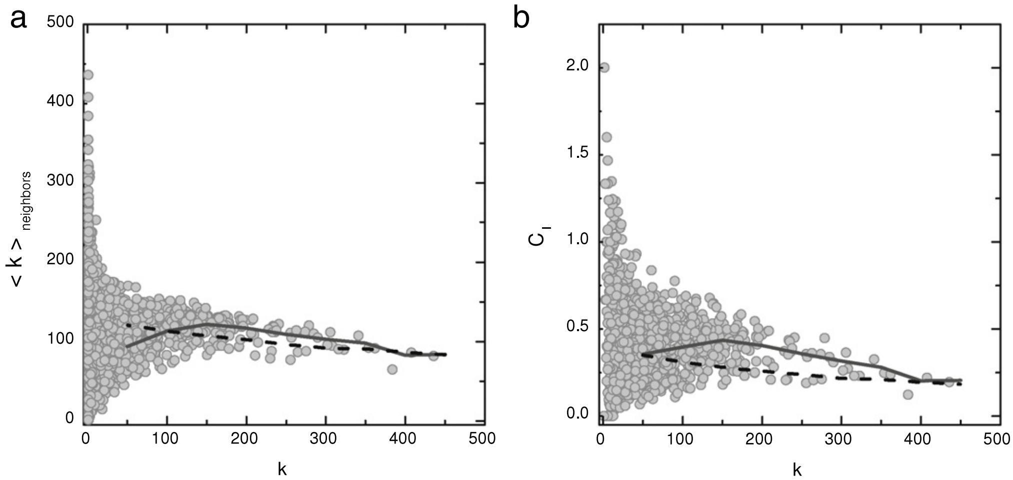

Fig. 4. Degree of a node neighbors (a) and the local clustering coefficients of the nodes (b) plotted as a function of the degree of a node (filled circles). The continuous line on both panels indicate averages of the points taken with a moving average of length 50 on the horizontal axes. Averages calculated for the network obtained from the configuration model appear with dashed black lines.

In order to further support the idea that such assortative mixing is not a dominant factor in describing the NNESM network, we investigated a simple configuration model [43] where a network with the same degree distribution as the NNESM network is considered. Within this model we start from the degree sequence of the NNESM network with free-endpoint links. The links are randomly allocated by the following algorithm: we randomly choose two free-end links belonging to different nodes and connect them — at this stage multiple connections between nodes are allowed — until all free-end-links are assigned. This preliminary link allocation procedure is followed by random local rewiring of nodes until all multiple connections are eliminated. The topology of such random networks can be analyzed and the results can be compared with the ones obtained for the NNESM network.

Networks created by the procedure described above are composed of a few connected components, similarly with the case of the NNESM network, exhibiting also a giant component. The average global and local clustering coefficient of the model networks generated by the configuration model is 0.132 and 0.269 respectively. In comparison with this, the NNESM network is characterized by a global clustering coefficient of 0.183 and a local clustering coefficient of 0.292. The nodes in the largest component of the model network have an average distance of 2.75 between them and this value for the NNESM network is 2.91. These results suggest that the simple configuration model reproduces well the NNESM network structure. One can also study in the model networks the average degree of the neighbors and the local clustering coefficient as a function of the degree of the node. Results in such sense are plotted with dashed lines in Fig. 4(a) and (b). Comparison with the averages calculated for the NNESM network (continuous lines in Fig. 4(a) and (b)) suggests again, that the simple configuration model — without any assortative mixing — describes acceptably the NNESM network.

In the remaining part of this section we explore the origin of the exponential degree distribution, since the configuration model uses this as a starting point and does not offer an explanation for this. Our main assumption is that the number of Erasmus links a university has is proportional with its size. The reason for this assumption is straightforward: the number of ESM connections a university has is proportional with the number of professional connections of the professors, which is inevitably proportional with the number of professors. The number of professors of a university reflects however the total number of the students of the university. According to this, we state that

$$k_i = cn_i \sim n_i, \tag{5}$$

where k_i is the degree in the NNESM network of a university with rank $r = i$, n_i is the size of this university (quantified by the total number of students) and c is a proportionality constant. Unfortunately we do not have in our database any information on the sizes of the universities in the ESM network, in such a manner we cannot rigorously prove our conjecture. There is no freely available statistics on the number of students of the Europen universities. In order to get information on the student numbers for the involved universities, the only possibility is to visit their homepages and dig there for the relevant information. This is a rather tedious job, since many universities do not publish such information for year 2003, and even if it exists, one has to find it on their homepage. In order to support the conjecture with some data we have done a random sampling. First we arranged the universities in random order, and following this order we looked up the relevant information for them until results for 150 universities were found. In Fig. 5 the degree of these universities in the NNESM network is plotted as a function of the number of students in year 2003. As expected, the data has quite a large scatter, however an average of these points considered on fixed length intervals of student number (length of the average was chosen as 5000) suggests the assumed linear trend.

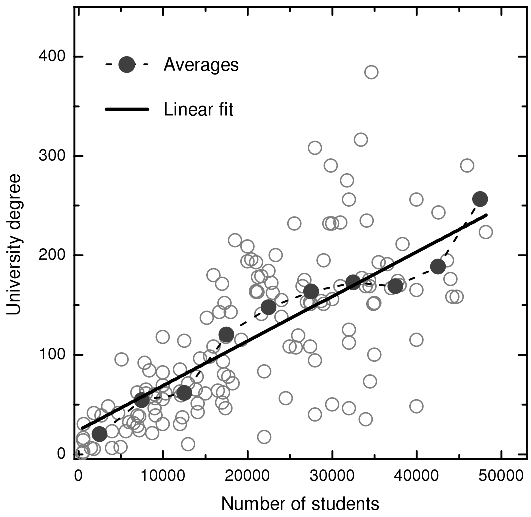

Fig. 5. Degree of the universities in the NNESM network as a function of their sizes. Open circles are results for random sampling on 150 universities, and filled circles present the average for an interval of length 5000 on the horizontal axes. The continuous line is a linear fit of the data points.

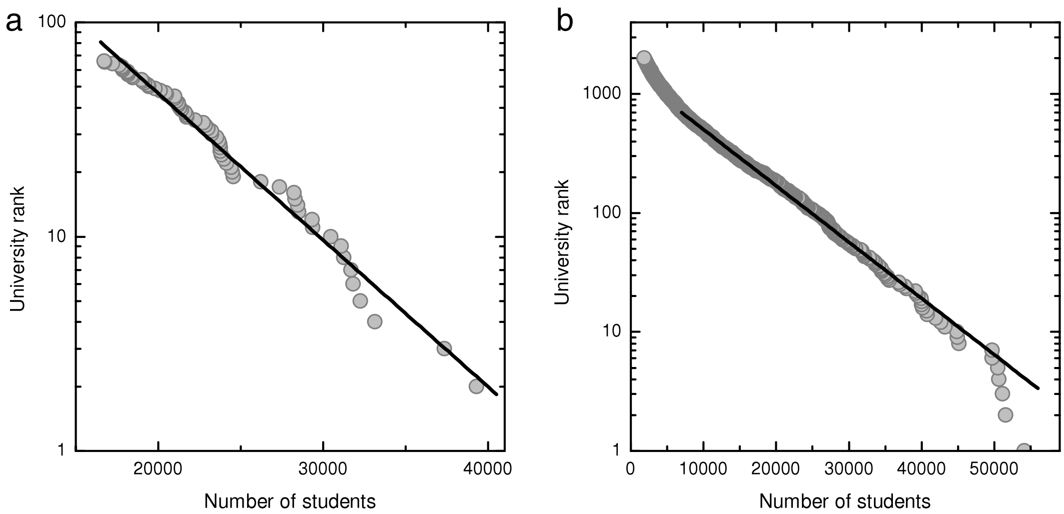

Fig. 6. University size-distribution by ranking plots: rank-size distribution of the largest UK (a) and US (b) universities. The thick continuous line on both figures suggests the exponential trend in the distribution.

We can bring now further arguments in supporting this conjecture. If this assumption holds, it means that the tail of the size-rank plot of the universities $r(n)$ should also follow an exponential trend. Following (1), (3) and (5):

$$r(k) = N \exp\left(-\frac{k}{\lambda}\right) = N \exp\left(-\frac{cn}{\lambda}\right) \longrightarrow r(n) = N \exp\left(-\frac{n}{\langle n \rangle}\right). \tag{6}$$

Although a complete size-distribution for the top European universities is not available, such data can be obtained for other geographical regions. For example the sizes of the largest universities from UK and US are freely accessible on the Internet [44,45]. Plotting their cumulative size-distribution by a ranking plot (similarly as in Fig. 3) a clear exponential trend is observable (Fig. 6).

The exponential nature of the size-distribution of the universities is seemingly a general feature for any geographical region and one might assume that it is valid for the European universities as well. This exponential size-distribution is in fact present in many other group-sizes like cattle-farm sizes [46], fish-schools [47–49] and other animal groups. The reason behind the generality of this group-size distribution is believed to be the maximum entropy principle [50]. The exponential distribution is the maximum entropy distribution among all continuous distributions supported in the $[0, \infty)$ interval that have a fixed mean. Assuming that for a given geographical region the number of universities and the number of students is fixed, the mean university size gets also fixed. If the maximum entropy principle holds, the distribution which realizes this for the considered student population and universities ensemble, is the exponential one.

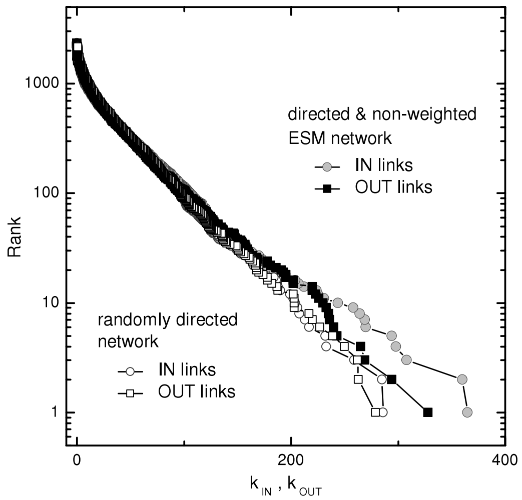

Fig. 7. Rank of universities as a function of their IN and OUT degree for the directed ESM network (filled symbols) and for the randomly directed configuration network (open symbols).

Thus, we accept the exponential tail for the size-distribution of the universities in a given geographical region and assume that the number of ESM links a university has is proportional with the size of the university. These two conjectures explain the obtained exponential tail for the degree distribution of the NNESM network.

4. The directed and weighted ESM network

In order to get a better insight into the development of the observed distribution of weights in the directed and weighted network constructed from the ESM database, we aimed also to model this more complex network. We start again with the simplest possible assumptions. By using the network obtained from the configuration model, we construct a randomly directed network. The undirected links of this initial network are randomly changed to symmetric or asymmetric type of connections. A simple analysis on the original ESM database shows that roughly 2/3 of the links between universities are unidirectional ones and 1/3 of the links are bidirectional. In agreement with this finding, the links obtained in the configuration model are set with 2/3 probability as asymmetric (choosing randomly its direction) and with 1/3 probability as symmetric ones. The network we obtain in such manner has a similar IN and OUT degree distribution as the one we construct from the ESM database. A comparison in such sense is plotted in Fig. 7. Both for the real and the model network the IN and OUT distribution of the links are quite similar. Please note however, that although each IN connection is evidently also an OUT connection, this does not necessarily mean that the degree distribution of the IN and OUT connections should be the same.

We generated thus a randomly directed network by using the undirected configuration network. The next step is to assign weights to these directed links, which means that we have to partition the number of students participating in the Erasmus mobility program on the previously obtained directed links. One can assume that the number of students leaving any university (S_{OUT}) should change monotonically as a function of the number of OUT links a university has (k_{OUT}). This assumption can be immediately verified in the original ESM database by plotting the student outflow from universities as a function of the number of OUT connections a university has. Results in such sense are shown with filled symbols in Fig. 8. The log–log plot suggests a simple scaling with exponent 1.17. One can use thus this scaling as a starting point to partition the total number of students on their home universities (universities with OUT links). Partitioning of the students on the OUT links rises again the question of selective or preferential linking. A first possibility would be to choose randomly and in an uncorrelated manner the available k_{OUT} links. This means that there is no preferential linking. A second possibility would be to assume that links that are already more populated with students will be chosen more probably by the students left, meaning a correlation and a kind of preferential linking. Of course, many other functional relations can be chosen to describe this preferential linking, but the simplest one is the linear proportionality.

Comparison between the obtained weight distribution after the allocation of students on the OUT links and the results measured for the real ESM network will decide which assumption is more realistic. First, the random, uncorrelated allocation of students is realized. Results for the obtained weight distribution is plotted in Fig. 9 with filled triangles. This figure also shows with open circles the weight distribution of the original ESM network. Next, we considered the correlated (preferential) placement of students on the outgoing links by a simple linear rule. First each OUT link belonging to the chosen university is occupied by one student. The students left are then one-by-one randomly assigned to one of the possible OUT links so that links that have already a bigger number of assigned students are considered with a higher probability. The simplest assumption is that the probability for one link to be chosen is linearly proportional with the number of students

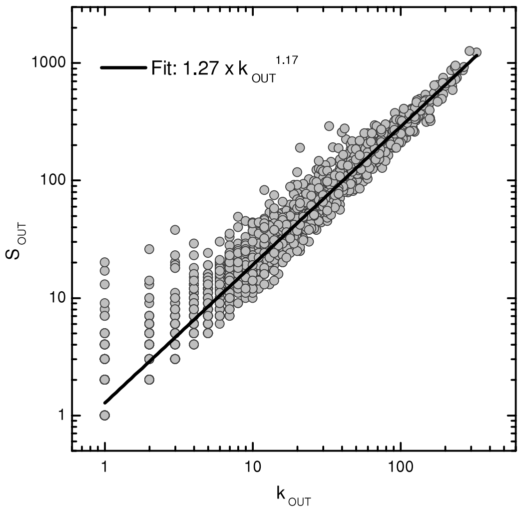

Fig. 8. Scaling of the total number of students sent out from a given university as a function of the number of OUT connections the university has. The continuous line shows a power-law fit with exponent 1.17.

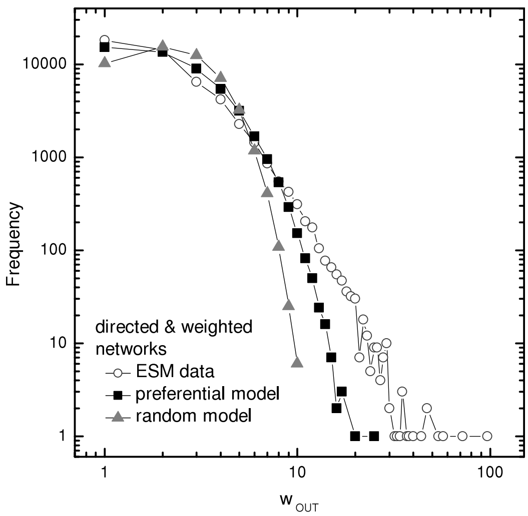

Fig. 9. Comparison of weight distributions for the directed and weighted networks: ESM network (open circles), the directed version of the network obtained by the configuration model with random (filled triangles) and preferential (filled squares) generation of weights.

already placed on that link. Results obtained for the weight distribution in this case is plotted in Fig. 9 with filled squares. Comparison with the weight distribution of the non-selective partition and the results for the real ESM networks suggests that preferential linking is indeed important for this allocation process. Results are improved relative to the ones offered by the non-correlated random partition.

5. Conclusion

In order to approach any complex system the network view is helpful. Nowadays it is believed that large social networks tend to have a scale-free topology and many times preferential attachment is considered to be a key ingredient in modeling their evolution. We have studied here a large social network, the ESM network for the year 2003, which is believed to reflect well the professional connections of the staff in European universities. Universities are the nodes of this network, and Erasmus student mobility between the universities are the links. Contrarily with the expectations, it is found that the non-directed and non-weighted version of this large network (having more than 2000 nodes) is not a scale-free one, and the tail of its degree distribution is fitted well by an exponential function. This exponential degree distribution can be explained by assuming that the universities have exponential size-distributions and their degree and sizes are linearly proportional. Empirical data for some randomly selected European universities and the top UK and US universities confirmed the exponential nature of this size distribution and a simple maximum entropy assumption leads also to this conclusion.

The directed and non-weighted structure of the ESM network could be understood by randomly directing the links in the network obtained by the configuration model and considering with 1/3 probability bidirectional links. The weight distribution of the ESM network is approximated by considering that (1) the number of outgoing students from one university scales with an exponent 1.17 as a function of the number of OUT connections of the given university and (2) there is a simple linear preferential (selective) linking rule in occupying the links: links where more students are already assigned are more probably chosen. Our analysis revealed also the most important hubs of the ESM network.

The considered network is thus a nice example of a complex large-scale social network which is not scale-free and where simple random connection models are helpful in understanding their topology. Although correlations proved not to be important for understanding the structure of the non-weighted ESM network, for describing the weights of the links a correlated partitioning is necessary.

Acknowledgements

The authors wish to thank Dr. László Letenyei for providing the data set containing information on the Erasmus scholarships. Financial support was provided by grant IDEI/PN2 nr. 2369 and from programs co-financed by the Sectoral Operational Programme Human Resources Development, Contract POSDRU 6/1.5/S/3 Doctoral Studies: "Through science towards society".

References

[1] P. Erdős, A. Rényi, Publ. Math. Debrecen 6 (1959) 290–297.
[2] P. Erdős, A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci. 5 (1960) 17–61.
[3] B. Bollobás, Random Graphs, Academic Press, London, 1985.
[4] M.S. Granovetter, Am. J. Sociol. 78 (1973) 1360–1380.
[5] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440–442.
[6] D.J. Watts, Princeton University Press, Princeton, NJ, 1999.
[7] S. Milgram, Psychol. Today 1 (1967) 60–67.
[8] S. Schnettler, Soc. Networks 31 (2009) 165–178.
[9] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, Nature 407 (2000) 651–654.
[10] H. Jeong, S. Mason, A.-L. Barabási, Z.N. Oltvai, Nature 411 (2001) 41–42.
[11] R. Albert, A.-L. Barabási, Rev. Modern Phys. 74 (2002) 47–97.
[12] A.-L. Barabási, Linked: The New Science of Networks, Perseus, Cambridge, MA, 2002.
[13] A.-L. Barabási, Z.N. Oltvai, Nat. Rev. Genet. 5 (2004) 101–114.
[14] S.N. Dorogovtsev, J.F.F. Mendes, Evolution of Networks: From Biological Nets to the Internet and WWW, Oxford University Press, Oxford, 2003.
[15] M.E.J. Newman, J. Park, Phys. Rev. E 68 (2003) 036122.
[16] S. Bortoluzzi, C. Romualdi, A. Bisognin, G.A. Danieli, Physiol. Genomics 15 (2003) 223–227.
[17] A.-L. Barabási, R. Albert, Science 286 (1999) 509–512.
[18] R. Albert, H. Jeong, A.-L. Barabási, Nature 406 (2000) 378–382.
[19] Z. Dezső, A.-L. Barabási, Phys. Rev. E 65 (2002) 055103(R).
[20] A.-L. Barabási, E. Bonabeau, Sci. Am. 288 (2003) 50–59.
[21] M.E.J. Newman, A.-L. Barabási, D.J. Watts, The Structure and Dynamics of Networks, Princeton University Press, Princeton, NJ, 2006.
[22] P. Bonacich, Soc. Networks 26 (2004) 285–288.
[23] C. Castellano, S. Fortunato, V. Loreto, Rev. Modern Phys. 81 (2009) 591–646.
[24] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Aberg, Nature 411 (2001) 907–908.
[25] B.F. de Blassio, A. Svensson, F. Liljeros, Proc. Natl. Acad. Sci. USA 104 (2007) 10762–10767.
[26] M.E.J. Newman, Proc. Natl. Acad. Sci. USA 98 (2001) 404–409.
[27] M.E.J. Newman, Proc. Natl. Acad. Sci. USA 101 (2004) 5200–5205.
[28] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Physica A 311 (2002) 590–614.
[29] S. Redner, Eur. Phys. J. B 4 (1998) 131–134.
[30] A.T. Stephen, O. Toubia, Soc. Networks 31 (2009) 262–270.
[31] G. Csányi, B. Szendrői, Phys. Rev. E 69 (2004) 036131.
[32] H.B. Hu, X.F. Wang, Phys. Lett. A 373 (2009) 1105–1110.
[33] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, Proc. Natl. Acad. Sci. USA 104 (2007) 7332–7336.
[34] C.A. Hidalgo, C. Rodriguez-Sickert, Physica A 387 (2008) 3017–3024.
[35] G. Kossinets, Soc. Networks 28 (2006) 247–268.
[36] M.A. Maggioni, T.E. Uberti, Ann. Regional Sci. 43 (2009) 691–720.
[37] S. Eubank, H. Guclu, V.S.A. Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, N. Wang, Nature 429 (2004) 180–184.
[38] M.E.J. Newman, Phys. Rev. E 64 (2001) 016131.
[39] J.W. Grossman, P.D.F. Ion, Congr. Numer. 108 (1995) 129–131.
[40] G.F. Davis, M. Joo, W.E. Baker, Strateg. Organ. 1 (2003) 301–326.
[41] M.E.J. Newman, Phys. Rev. Lett. 89 (2002) 208701.
[42] M.E.J. Newman, Phys. Rev. E 67 (2003) 026126.
[43] M.E.J. Newman, SIAM Rev. 45 (2003) 167–256.
[44] Higher Education Statistics Agency (HESA), 2009. All students by institution, mode of study, level of study and domicile 2007/08. http://www.hesa.ac.uk (accessed 01.02.10).
[45] GlobalScholar, 2008. CollegeFinder. http://www.globalscholar.com/collegefinder/results.aspx (accessed 01.02.10).
[46] C.T.K. Ching, Am. J. Agric. Econ. 55 (1973) 500–502.
[47] J.J. Anderson, Fish. Bull. 79 (1981) 315–323.
[48] E. Bonabeau, L. Dagorn, Phys. Rev. E 51 (1995) R5220–R5223.
[49] H.-S. Niwa, J. Theoret. Biol. 224 (2003) 451–457.
[50] E.T. Jaynes, Phys. Rev. 106 (1957) 620–630.