Volume 385, issue 1 1 November 2007 ISSN 0378-4371

ELSEVIER

PHYSICA A
STATISTICAL MECHANICS AND ITS APPLICATIONS

Recognized by the European Physical Society

Available online at
ScienceDirect
www.sciencedirect.com

http://www.elsevier.com/locate/physa

Available online at www.sciencedirect.com
ScienceDirect
Physica A 385 (2007) 363–369
PHYSICA A
www.elsevier.com/locate/physa

Interpersonal interactions and human dynamics in a large social network

Andrzej Grabowski

Central Institute for Labour Protection—National Research Institute, 00-701 Warsaw, Poland

Received 27 April 2007; received in revised form 4 June 2007
Available online 9 June 2007

Abstract

We study a large social network consisting of over 10^6 individuals, who form an Internet community and organize themselves in groups of different sizes. On the basis of the users' list of friends and other data registered in the database we investigate the structure and time development of the network. The structure of this friendship network is very similar to the structure of different social networks. However, here a degree distribution exhibiting two scaling regimes, power-law for low connectivity and exponential for large connectivity, was found. The groups size distribution and distribution of number of groups of an individual have power-law form. We found very interesting scaling laws concerning human dynamics. Our research has shown how long people are interested in a single task.

PACS: 89.75.Da; 89.75.Hc; 89.65.Ef

Keywords: Complex networks; Human dynamics; Power-law distributions

1. Introduction

In recent years investigations of complex networks have attracted the physics community's great interest. It was discovered that the structure of various biological, technical, economical, and social systems has the form of complex networks [1]. The short length of the average shortest-path distance, the high value of the clustering coefficient and the scale-free distribution of connectivity are some of the common properties of those networks [1,2]. Social networks, which are an important example of complex networks, have such properties, too.

The advent of modern database technology has greatly advanced the statistical study of networks. The vastness of the available data sets makes this field suitable for the techniques of statistical physics [2]. The study of the statistical properties of social networks, e.g. friendship networks, is still very difficult. It is possible to assess the form of distribution of connectivity as a result of a survey, like in the case of web human sexual contacts [3]. However, other important properties describing the structure of the network are much more

E-mail address: angra@ciop.pl

0378-4371/$ - see front matter
doi:10.1016/j.physa.2007.06.005

difficult to obtain, as a result of the lack of data on the whole network. A survey often provides data on a small sample of the whole network only.

Progress in information technology makes it possible to investigate the structure of social networks of interpersonal interactions maintained over the Internet. Some examples of such networks are e-mail networks [4], blog networks [5] and web-based social networks of artificial communities [6]. The aim of this work is to introduce a data set describing a large social network of an Internet community (*Grono*), which consists of more than 10^6 individuals. The *Grono* (cluster) project was started in Poland in 2004 on the website www.grono.net. During its 36 months of existence, it has grown into a well-known social phenomenon among Internet users in Poland. Membership is strictly invitation only; existing members can invite an unlimited number of friends to the network via email, who, if they choose to do so, join the network by an initial link connecting to the person who invited them. All users can add, by mutual consent, and remove other people from their databases of friends. In this way undirected friendship network is formed. We show that the structure of this network has similar properties to those of other social networks. We have found interesting scaling laws concerning human dynamics [7–12], like lifespan of an individual in the internet community. In order to study the structure of the network and of human activity, we analyzed data containing the list of all friends, creation date and last login date. The network under consideration consisted of a collection of individuals (network nodes) connected among one another by friendly relationships (network links).

2. Results

Basic network measures of the whole network and a giant component (GC) [1] are presented in Table 1. The network consists of $N = 1\,002\,182$ individuals and the GC contains almost all individuals (994 381 individuals); only 7801 individuals do not belong to GC. The value of the clustering coefficient C is two orders of magnitude larger than that of a random graph (RG). The average path length $\langle l \rangle$ in GC is very small and only slightly greater than that in a random graph. A high value of the clustering coefficient and a short average path length $\langle l \rangle$ are characteristic features of social networks [2,13]; they are typical for small-world networks [14].

The degree distribution of the network is plotted in Fig. 1. The graph shows power-law regime $P(k) \sim k^{-\gamma}$ with $\gamma = 0.75$ for low k ($k < 100$). Such a power law is common in many types of networks [1], also in social networks [3,4,6,15]. Thus, such a relation can be a consequence of the fact that in the case of individuals with low k the majority of the links represent genuine pre-existing social acquaintances. However, for large k ($k > 100$) the degree distribution has exponential form $P(k) \sim e^{-0.01k}$ (Fig. 1b). This result is quite different from that presented in Ref. [6], where degree distribution exhibits two power scaling regimes separated by a critical degree. We suggest, that the discrepancy between results is explained by the relatively short age and small size (1.2×10^4 nodes) of the network presented in Ref. [6]. The form of degree distribution indicates the existence of two underlying networks that are simultaneously present in the social network under investigation: the network of links that represent preexisting social contacts for low k and the network of links represent social contacts maintained only via Internet for large k. The form of degree distribution is also different from results observed in other social networks, because in network under investigation two different types of users

Table 1
Average properties of the whole network and the giant component (GC) and comparison with a random graph (RG), Barabási–Albert network (BA) and small-world network model proposed by Strogatz (SW) with the same number of nodes N and the same average connectivity $\langle k \rangle$

	N	C	$\langle l \rangle$	$\langle k \rangle$
Network	1002 182	0.2	–	46.3
GC	994381	0.2	4.4	46.3
RG	994381	0.001	4.0	46.3
BA	994381	0.0005	3.5	46.3
SW	994381	0.7	10	46.3

In the case of SW network the probability of creating of a shortcut between randomly selected pairs of nodes equals $\phi = 0.0001$.

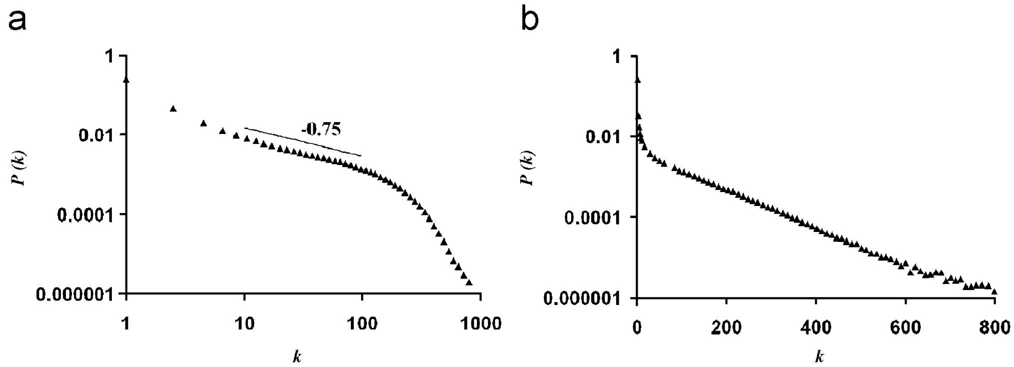

Fig. 1. The degree distribution. In the case of low k results can be approximated with power law $P(k){\sim}k^{-\gamma}$, where $\gamma = 0.75$ (solid line). For large k ($k>100$) the degree distribution has exponential form $P(k){\sim}e^{-ak}$, where $a = 0.01$.

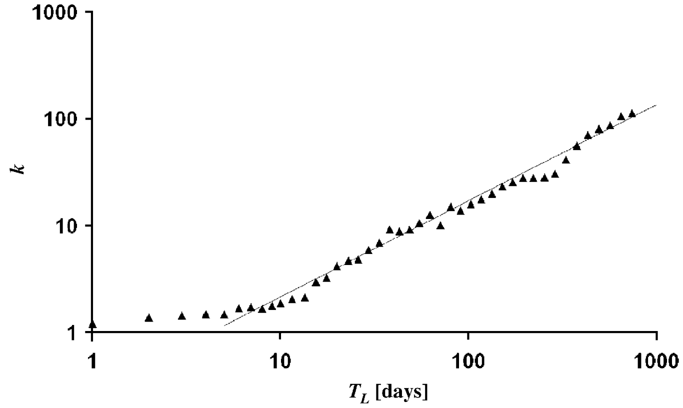

Fig. 2. The relation between lifespan T_L of an individual and its connectivity k (triangles) and fit to power law $k(T_L){\sim}T_L^{0.9}$ (solid line).

behavior are observed. Most of individuals devote their time to maintain contacts with their friends and they are less interested in adding new people to their databases of friends. However, there is a small group of users whose behavior is opposite, they use the web service to make new friends. We suggest, that they aim at collecting new friends and devote little time to maintain contacts with old ones.

The time development of the connectivity of an individual, i.e. the relation between lifespan of an individual and its connectivity, is shown in Fig. 2. The lifespan of an individual T_L is defined as the number of days since the time of an individual was added (invited to the network) to the date of last logging. The relation $T_L(k)$ can be approximated by power-law relation $T_L(k){\sim}k^{0.9}$. It should be noted that similar value of the exponent (0.8) was found in other social network [15].

Each individual can invite to the network unlimited number of friends and the distribution of number of invited friends k_I is shown in Fig. 3. It is visible that the invite network degree distribution is qualitatively different from the total degree distribution. This is so, because the invitation links represent only a part of pre-existing social acquaintance without contacts maintained only via Internet. The number of invited friends k_I increases with total degree k and can be approximated by power-law relation $k_I{\sim}k^{0.4}$.

The average connectivity of the nearest neighbors k_{NN} of a node with k connections is shown in Fig. 4. It can be seen that the greater the k, the greater the k_{NN}. Hence, the network under investigation is assortative mixed by degree; such a correlation is observed in many social networks [16]. In social networks it is entirely possible, and is often assumed in sociological literature, that similar people attract one another. On the other hand individuals organize in groups of different sizes in order to chat together. Since it is highly probable that members of a group are connected to other members, the positive correlations between degrees may at least in part reflect the fact that the members of a large (small) group are connected to the other members of the same large (small) group. The relation $k_{NN}(k)$ can be approximated by power-law relation $k_{NN}(k){\sim}k^{0.2}$. It should be noted that similar value of the exponent (0.18) was found in other social network [15].

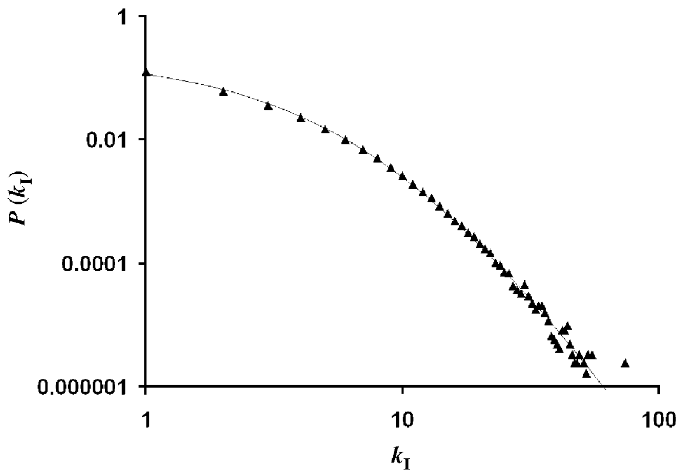

Fig. 3. The degree distribution of the invitation tree (triangles) and fit to power law $P(k_I) \sim (k_I + 9)^{-6}$ (solid line).

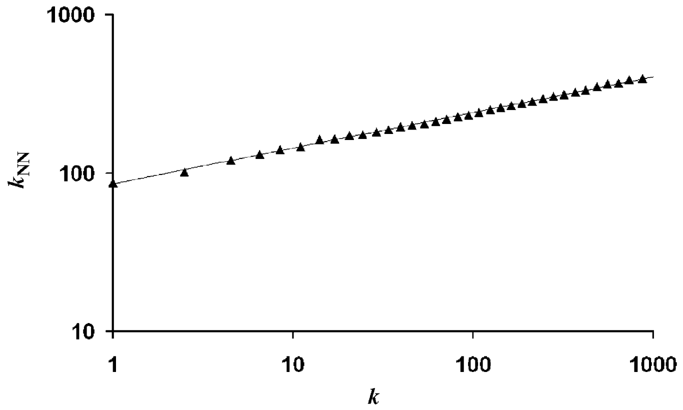

Fig. 4. The relation between the average connectivity of the nearest neighbors k_{NN} of a node and its connectivity k (triangles) and fit to power law $k_{NN}(k) \sim k^{0.2}$ (solid line).

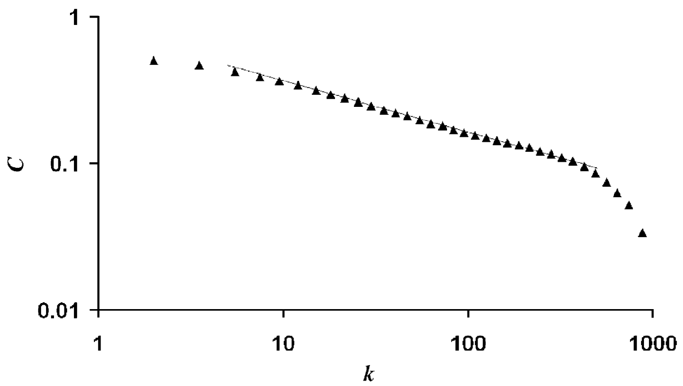

Fig. 5. The relation between the clustering coefficient of a node $C(k)$ and its connectivity k (triangles) and fit to power law $C(k) \sim k^{-0.35}$ (solid line).

The behavior of the clustering coefficient C is an interesting problem. Fig. 5 plots the correlation between the local clustering coefficient $C(k)$ and the node degree k, showing the existence of a power law $C(k) \sim k^{-\alpha}$ with $\alpha = 0.35$. Similar value of the exponent α has been observed in other social networks ($\alpha = 0.33$ [6] and $\alpha = 0.44$ [15]). The power-law relation $C(k)$ is similar to the relationship observed in hierarchical networks

[17]. Such power laws hint at the presence of a hierarchical architecture: when small groups organize themselves into increasingly larger groups in a hierarchical manner, local clustering decreases on different scales according to such a power law. This may be connected with the fact that individuals can freely make acquaintances, without barriers connected with spatial distance between individuals. The influence of spatial distance between nodes in Euclidean growing scale-free networks on $C(k)$ relation is described in Ref. [18].

The website contains additional services such as discussion forums, hence users can organize themselves into groups of different sizes S. Each group is called *grono* (cluster), and members of that group chat together on a single topic (music, movies, health, politics, etc.). Each individual can join to unlimited number of groups. Hence, the size of a group increases in time. It occurs that the relation between the size of a group S and its lasting time t_G has an exponential form $S(t_G)\sim\exp(0.005t_G)$. The distribution of S has a power-law form and is shown in Fig. 6. These results indicate that the increase in the size of a group (in one time unit) and the probability that an individual join to a group are proportional to the size of this group. Because an individual can be a member of many groups it is interesting to plot the distribution of number of groups of an individual n_G (see Fig. 7). Most of individuals belong to small number of groups, $n_G<10$. For greater values of n_G, the probability decreases more abruptly and the distribution can be fit to a power law. The number n_G is positive correlated with degree of an individual, and this relation can be approximated by power law $n_G\sim k^{0.5}$.

On-line communities offer a great opportunity to investigate human dynamics, because much information about individuals is registered in databases. To analyze how long people are interested in a single task, we studied creation date and last login date registered in the database. Individuals can lose interest in using the website after some time. We consider an individual as active when it regularly uses web services, and we consider an individual as inactive when it does not login for more than one month. The distribution of probability P_L that an inactive individual has the lifespan T_L is shown in Fig. 8. This distribution can be approximated with the power law $N_L(T_L)\sim T_L^{-\beta}$ with $\beta=0.6$. Thus, the probability that a human will devote

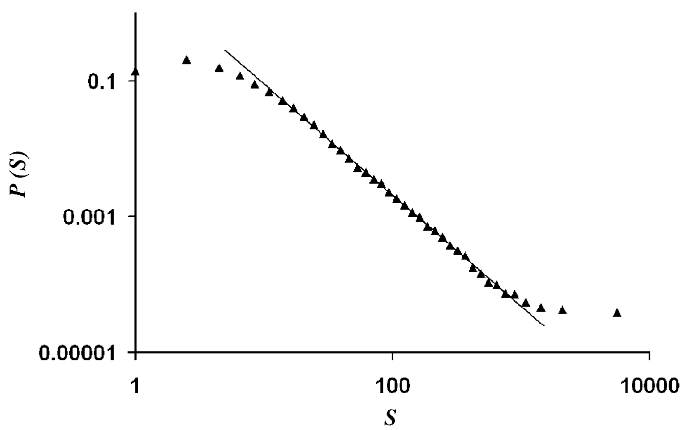

Fig. 6. Groups size distribution (triangles) and fit to power law $P(S)\sim S^{-1.6}$ (solid line). The number of groups $N_G=70\,100$.

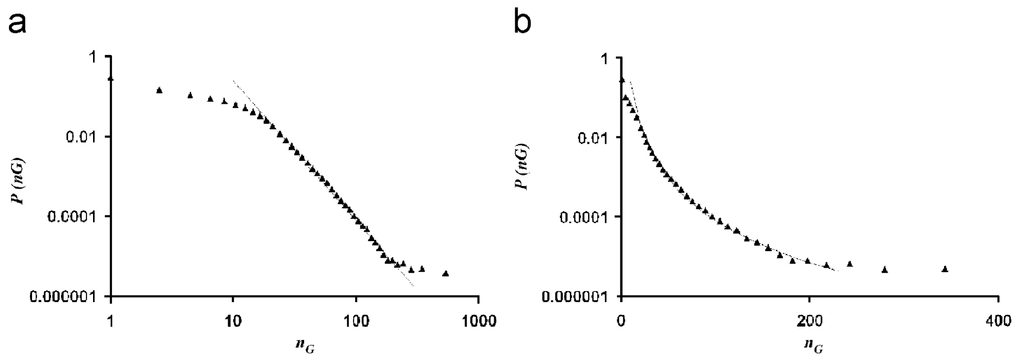

Fig. 7. Distribution of number of groups of an individual (triangles) and fit to power law $P(n_G)\sim n_G^{-3.5}$ (solid line).

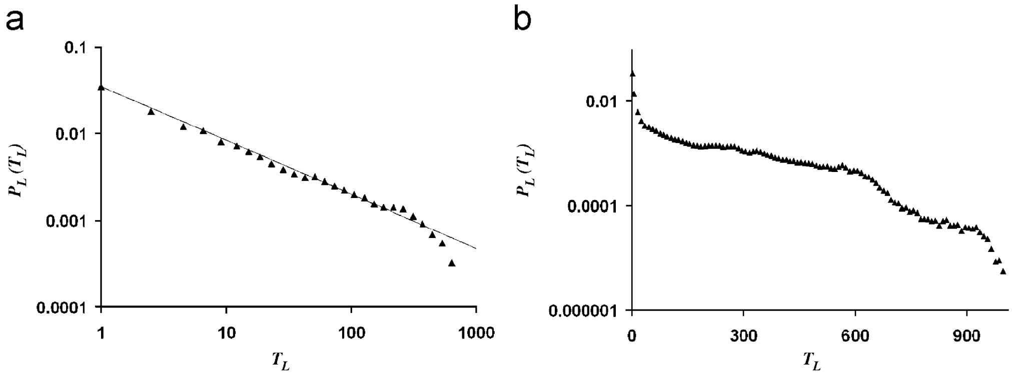

Fig. 8. The distribution of probability P_{L} that activity of an individual in the community last T_{L} days (triangles) and fit to power law $P_{\mathrm{L}}(T_{\mathrm{L}}){\sim}T_{\mathrm{L}}^{-0.6}$ (solid line). Average time T_{L} equals 200 days.

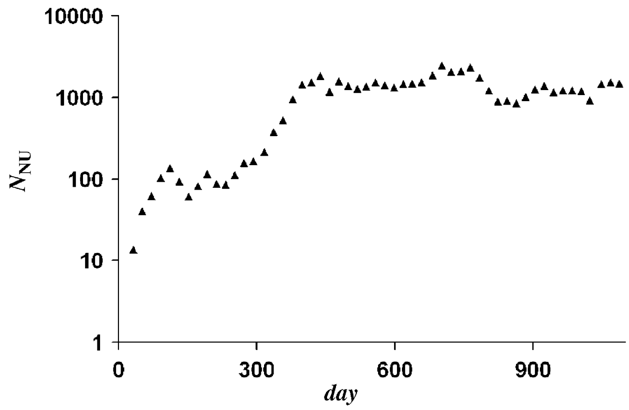

Fig. 9. The number N_{NU} of new users added to the system per day. The maximal value of N_{NU} is $N_{\mathrm{NU}}^{\mathrm{MAX}} = 5 \times 10^3$.

the time t to a single activity has a fat-tailed distribution. Many individuals lose interest in using the website very shortly, so they cannot make new acquaintances. Therefore, abrupt decrease in $P(k)$ for very low k is visible (see Fig. 1).

For large values of T_{L} an abrupt decrease in P_{L} is visible (see Fig. 8b). This is so because in the first year the web service was much less popular. The number of new users N_{NU} added per day was less than 10^2 (see Fig. 9). After the first year an abrupt increase in the number N_{NU} is visible, because *Grono* project has grown into a well-known social phenomenon among Internet users in Poland. In the last two years the number N_{NU} is greater than 10^3 and is one order of magnitude greater than in the first year. Therefore there is a decrease in the number of individuals having large values of T_{L}.

It should be stressed that we have found such a distribution of lifespan (with value of the exponent $\beta = 1.0$) in other social network [15]. It indicates that such a scaling law can be a typical relation and the value of the exponent depends on type of human activity (e.g. entertainment, interpersonal interactions). Similar relations concerning human dynamics have also been observed elsewhere [19] and can be a consequence of a decision-based queuing process. The model of such a process was recently proposed by Barabási [7,20,21]. It indicates that scale-free distributions are common in human dynamics.

3. Conclusions

In conclusion, we have shown that a friendship network maintained in Internet community has similar properties (e.g. large clustering, a low value of the average path length, assortative mixing by degree, hierarchical structure) to other social networks. However, here a degree distribution exhibiting two scaling regimes, power law for low connectivity, and exponential for large connectivity, was found. The time development of the connectivity of an individual was calculated. Similar value of the exponent in the power-law relation between life-span of an individual and its connectivity was found in other social network. Individuals organize themselves into groups of different sizes. The groups size distribution and distribution of number of groups in which an individual is a member have power-law form. On the basis of creation date and last login date of each individual recorded on the server, we have presented results concerning human dynamics. The power-law form of distributions $N_L(T_L)$, $k(T_L)$ and other authors' [7,15] results indicate that such a scaling law is common in human dynamics and should be taken into account in models of the evolution of social networks and of human activity.

Acknowledgments

We wish to thank Izabela Krawczyk for help in the preparation of this article. A.G. acknowledges financial support from the Foundation for Polish Science (FNP 2007).

References

[1] R. Albert, A-L. Barabási, Rev. Mod. Phys. 74 (2002) 47–97.
[2] S.N. Dorogovtsev, J.F.F. Mendes, Evolution of Networks, Oxford University Press, Oxford, 2004.
[3] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Aberg, The web of human sexual contacts, Nature 411 (2001) 907–908.
[4] H. Ebel, L.I. Mielsch, S. Bornholdt, Scale free topology of e-mail networks, Phys. Rev. E 66 (2002) 035103(R).
[5] W. Bachnik, S. Szymczak, P. Leszczynski, R. Podsiadlo, E. Rymszewicz, L. Kurylo, D. Makowiec, B. Bykowska, Quantitive and sociological analysis of blog networks, Acta Phys. Pol. 36 (2005) 3179–3191.
[6] G. Csanyi, B. Szendroi, Structure of a large social network, Phys. Rev. E 69 (2004) 036131.
[7] A-L. Barabási, The origin of bursts and heavy tails in human dynamics, Nature 435 (2005) 207.
[8] A. Vazquez, Impact of memory on human dynamics, Phys. A 373 (2006) 747–752.
[9] A. Hellervik, G.J. Rodgers, A power law distribution in patients' lengths of stay in hospital, Physica A 379 (2007) 235–240.
[10] G. Ghoshal, P. Holme, Attractiveness and activity in internet communities, Physica A 364 (2006) 603–609.
[11] A. Namazi, A. Schadschneider, Statistical properties of online auctions, Int. J. Mod. Phys. C 17 (2006) 1485–1493.
[12] U. Hardera, M. Paczuski, Correlated dynamics in human printing behavior, Physica A 361 (2006) 329–336.
[13] M.E.J. Newman, J. Park, Why social networks are different from other types of networks, Phys. Rev. E 68 (2003) 036122.
[14] S.H. Strogatz, Exploring complex networks, Nature (London) 410 (2001) 268.
[15] A. Grabowski, N. Kruszewska, Experimental study of the structure of a social network and human dynamics in a virtual society, Int. J. Mod. Phys. C 18 (2007), to appear.
[16] M.E.J. Newman, Mixing patterns in networks, Phys. Rev. E 67 (2003) 026126.
[17] E. Ravasz, A-L. Barabási, Hierarchical organization in complex networks, Phys. Rev. E 67 (2003) 026112.
[18] P. Sen, S.S. Manna, Phys. Rev. E 68 (2003) 026104.
[19] T. Henderson, S. Nhatti, Modelling user behavior in networked games, Proceedings of the 9th ACM International Conference on Multimedia, 2001, pp. 212–220.
[20] J.G. Oliveira, A-L. Barabási, Darwin and Einstein correspondence patterns, Nature 437 (2005) 1251.
[21] A. Vázquez, Exact results for the Barabási model of human dynamics, Phys. Rev. Lett. 95 (2005) 248701.