PERCOLATION IN REAL ON-LINE NETWORKS*

Andrzej Grabowski[a], Robert A. Kosiński[a,b]

[a]Central Institute for Labour Protection — National Research Institute
Czerniakowska 16, 00-701 Warsaw, Poland
[b]Division of Complex Systems Physics, Faculty of Physics
Warsaw University of Technology
Koszykowa 75, 00-662 Warsaw, Poland

(Received February 10, 2010)

We study bond and site percolation in four real social networks: two Internet society of friends consisting of over 10^6 and 10^7 people, over 10^6 users of music community website and over 5×10^6 users of gamers community server. We study the properties of those systems (*e.g.* the network components size distribution) in function of fraction p of nodes or links that retained in network. We have calculated critical fraction $p_{\rm c}$ at which the percolation transition takes place and giant component emerges.

PACS numbers: 89.75.Hc, 89.75.Fb, 64.60.ah

1. Introduction

In recent years investigations of complex systems have attracted the physics community's great interest, *e.g.* it was discovered that the structure of various biological, technical, economical, and social systems has the form of complex networks sharing common properties [1]. The advent of modern database technology has greatly advanced the statistical study of social systems. The vastness of the available data sets makes this field suitable for the techniques of statistical physics. 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, blog networks [2] and web-based social networks of artificial communities [3]. All users of such systems can add, by mutual consent, other people to their databases of friends.

We study the properties of the diluted network, *i.e.* when a fraction $p_{\rm n}$ of nodes or $p_{\rm b}$ of bonds retained in the network. If p is large enough the network remains connected and exist a giant component. Below a certain

* Presented at the XXII Marian Smoluchowski Symposium on Statistical Physics, Zakopane, Poland, September 12–17, 2009.

threshold $p_{\rm c}$ the giant components disappears. Above the threshold the network is in a percolating phase [4]. Percolation in real-world networks is widely recognized as a key problem of interest [5]. Some examples of possible applications are the robustness of communication networks (*e.g.* the Internet) [6] for random failure, the efficiency of preventive vaccination against the spread of disease [7] or information propagation in networks (*e.g.* rumor spreading) [8].

The aim of this work is to investigate bond and site percolation in four different on-line social networks. Two of them are large social networks of an Internet community, which consists of 10^6 (Grono; `www.grono.net`) and 10^7 (Skyrock; `www.skyrock.com`) individuals. The Skyrock project was started on the website `www.skyrock.com`. During its existence, it has grown into a well-known social phenomenon among (mainly French-speaking) Internet users. In both systems all users can add, by mutual consent, and remove other people from their databases of friends. In this way undirected friendship network is formed.

The third system under investigation is LastFM (`www.last.fm`) — the music community server, and more exactly the part of it known as Audioscrobbler project which was started in year 2002. There is about 10^6 users of this system. Data gathered by the web-service is used to find users with similar music taste. On that basis people with similar music taste and songs they often listen to is present and recommended to users who can see this information on their profile web site via web browser. This way people with similar music taste can meet each other and have the possibility to make friends (mutual consent is required).

The fourth system under investigation is XFire (`www.xfire.com`). It is gamers community program similar to every Internet Chat systems, marked out its integration with almost all popular computer games. People who like to play computer games are using this application to keep in contact with other players even when they do not play any game in that moment or play two different games. For this purpose, they add other people into their friend list (mutual consent is required) and have possibility to see which game their friend plays, how much overall time they played and can always chat with this person when online. X-fire allows to see friends of your friends, so people have greater chance to make new acquaintanceship.

2. Results

Basic network measures of networks and Giant Components (GC) are presented in Table I. In all cases the value of the clustering coefficient C is two orders of magnitude larger than that of a random graph. The average path length $\langle l\rangle$ in GC is very small and comparable to that in a random graph. A high value of the clustering coefficient C and a short average path length $\langle l\rangle$ are characteristic features of social networks [9]; they are typical for small-world networks.

TABLE I

Average properties of the whole network and the giant component (GC) for four systems: Grono, LastFm, XFire, SkyRock. Values in parentheses indicate the uncertainty in the final digit.

Network	N (N_{GC})	$\langle k\rangle$	$\langle l\rangle_{\mathrm{GC}}$	C	k_0	γ	τ	κ
Grono	1002182 (994381)	46.3	4.4	0.2	0	0.83(3)	4.50(5)	138
LastFM	1192118 (816093)	4.2	6.6	0.23	10(1)	4.0(1)	3.61(2)	12
XFire	5241578 (957579)	1.2	6.9	0.17	0	1.92(1)	3.80(1)	18
SkyRock	9823234 (8228056)	16.7	7.5	0.14	0	0.8(1) 0	5.30(1) 2.70(2)	23

The degree distribution of networks is plotted in Fig. 1. The graph shows power-law regime for all networks (the parameters of the power-law distributions were computed with maximum likelihood estimation [10] in other cases least-squares fitting were used). Such a power-law is common in many types of networks [1], also in social networks [9]. The connectivity distribution in each case can be approximated with the power-law $P(k) \sim (k+k_0)^{-\gamma}$ (the values of the parameters are presented in Table I). In the case of Grono for large k ($k > 100$) the degree distribution starts to take an exponential form

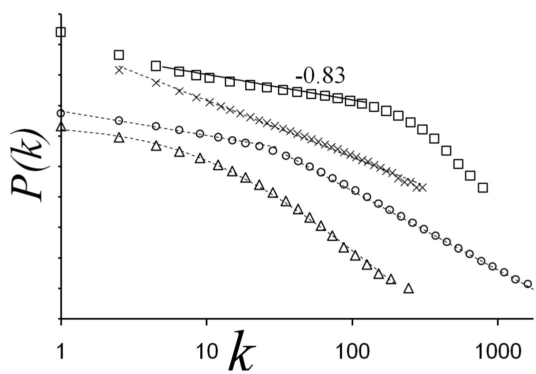

Fig. 1. Degree distribution for four systems: Grono (squares), LastFm (triangles), XFire (crosses) and Skyrock (circles) in a double logarithmic scale. Results can be fit to power-law $P(k) \sim (k+k_0)^{-\gamma}$ (dashed line). The values of the exponents are presented in Table I. The datasets are vertically offset.

$P(k) \sim e^{-0.01k}$. In the case of SkyRock two scaling regimes are observed, for low connectivity with $\gamma = 0.8$ and for large connectivity with $\gamma = 2.7$. To compare the goodness of fit for power-law distribution and more general models, $P_1(k) \sim k^{\beta-1}e^{-\lambda k^{\beta}}$ and $P_2(k) \sim e^{-\lambda k}(x-k_0)^{-\beta}$, the likelihood ratio test was used [10]. These models include cutoffs and saturation, hence can be fitted to all data. The basic idea behind the likelihood ratio test is to compute the likelihood of our data in two competing distributions. The one with the higher likelihood is then the better fit. For low and large values of k the power-law distribution fits better than P_1 and P_2.

For a random network of arbitrary degree distribution, the condition for the existence of a spanning cluster is $\kappa = \frac{\langle k^2\rangle}{2\langle k\rangle} > 1$ [12]. In the networks under investigations the values of the parameter κ are much greater than one (see Table I), thus those networks are in percolative phase. The properties of the structure of the networks under investigation are presented in more details in Ref. [3, 11].

In order to investigate the dynamics of the bond and site percolation in real social networks we introduce two observables: the number of nodes in Giant Component N_{GC} and the number of bonds in GC K_{GC}. Thus, the probability that randomly chosen node belongs to GC equals $S_{\mathrm{N}} = \frac{N_{\mathrm{GC}}}{N}$ and the probability that randomly chosen bond belongs to GC equals $S_{\mathrm{B}} = \frac{K_{\mathrm{GC}}}{K}$, where N is the number of nodes and K is the number of bonds in the network. The value of S_{N} and S_{B} increases with an in increase in a fraction p_{n} of nodes or p_{b} of bonds retained in the network. The relationship between relative number of bonds and nodes in GC and p is shown in Fig. 2. For low values of p the value of order parameter do not change and $S \approx 0$. However for large enough p, near to critical value p_{c}, a rapid increase in S is observed.

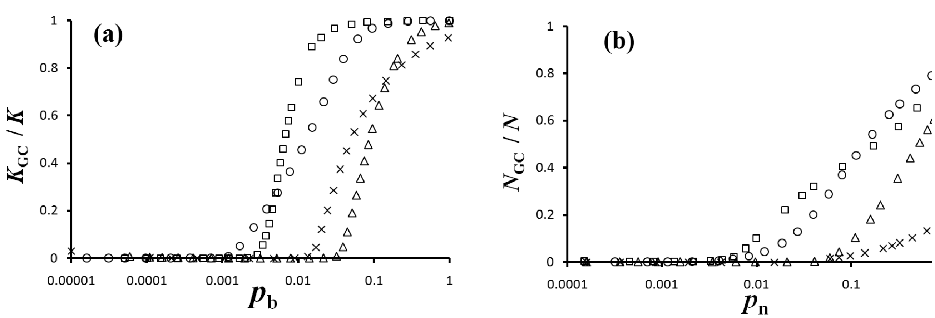

Fig. 2. The relationship between fraction of bonds p_{b} or nodes p_{n} that retained in network and relative number of bonds in Giant Component (a) and relative size of Giant Component (b) for four systems: Grono (squares), LastFm (triangles), XFire (crosses) and Skyrock (circles).

Near criticality, for $p \geq p_{\mathrm{c}}$, the probability of belonging to the spanning cluster (GC) behaves as

$$S \sim (p - p_{\mathrm{c}})^{\beta} \,, \tag{1}$$

where β is the order parameter critical exponent. For infinite-dimensional systems (such as a Cayley tree) and for scale-free networks with $\gamma > 4$ it is known that $\beta = 1$ [4, 13]. According to Ref. [13] the order parameter exponent β for scale-free networks equals $\beta = \frac{1}{\gamma - 3}$ for $3 < \gamma < 4$ and $\beta = \frac{1}{3-\gamma}$ for $2 < \gamma < 3$. The existence of an infinite-order phase transition at $\gamma = 3$ for growing networks of the Albert–Barabási model, has been reported in [15].

The relationship between p and order parameter S for $p \approx p_{\mathrm{c}}$ is shown in Fig. 3. For $p = p_{\mathrm{c}}$ the value of parameter $\kappa \approx 1$. For values of control parameter greater than critical value and for $\frac{p-p_{\mathrm{c}}}{p_{\mathrm{c}}} < 0.5$ the value of order parameter increases approximately linearly with p. Thus, for all networks under investigation we obtain $\beta \approx 1$. The critical values of control parameters equal: $p_{\mathrm{bc}} \approx 0.003$ and $p_{\mathrm{nc}} \approx 0.005$ for Grono, $p_{\mathrm{bc}} \approx 0.035$ and $p_{\mathrm{nc}} \approx 0.05$ for LastFM, $p_{\mathrm{bc}} \approx 0.001$ and $p_{\mathrm{nc}} \approx 0.0025$ for SkyRock, $p_{\mathrm{bc}} \approx 0.015$ and $p_{\mathrm{nc}} \approx 0.03$ for XFire. In all cases $p_{\mathrm{bc}} < p_{\mathrm{nc}}$.

It is surprising that in four different systems the form of distribution of sizes of network components $P(s)$ is very similar. In all cases the results can be approximated with the power-law $P(s) \sim s^{-\tau}$. The values of the exponents are presented in Table I. The relation between parameter γ and the exponent τ for scale-free networks which are close to the percolation threshold is presented in Ref. [13]. It has been shown that the exponent τ bear a strong γ-dependence $\tau = \frac{2\gamma-3}{\gamma-2}$ for $2 < \gamma < 4$ (note that in the case of networks LastFM and XFire the value of the exponent τ seems to be independent of the form of the degree distribution, see Table I). For $\gamma > 4$ usual mean field result $\tau = \frac{5}{2}$ is observed. Related results for growing networks of the Albert–Barabási model are presented in Ref. [14]. In networks under investigation the value of the parameter τ is high, however the fit to power-law relation is much better than to exponential decay. In order to determining whether a power-law is consistent with the empirical data is by computing the p-value for the power-law distribution [10]. In all systems under investigation the p-value is above threshold value (*i.e.* 0.05), hence the power-law model cannot be rejected.

The relationship between fraction of bonds or nodes that retained in network and the value of the exponent τ for networks under investigation is shown in Fig. 4. For the value of p for which the minimum in $\tau(p)$ relationship is observed the number of components in the network reaches maximum.

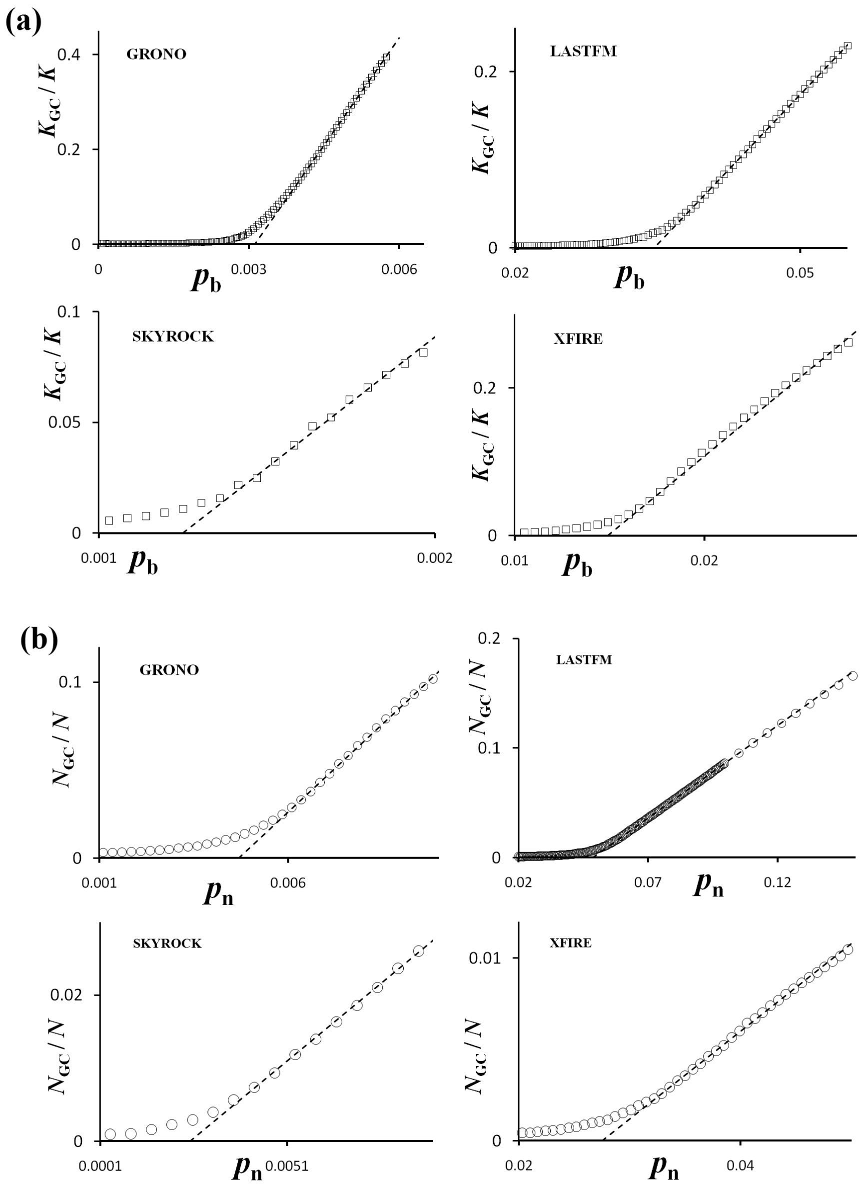

Fig. 3. The relationship between fraction of bonds $p_{\mathrm{b}} \approx p_{\mathrm{bc}}$ or nodes $p_{\mathrm{n}} \approx p_{\mathrm{nc}}$ that retained in network and relative number of bonds in Giant Component (a) and relative size of Giant Component (b).

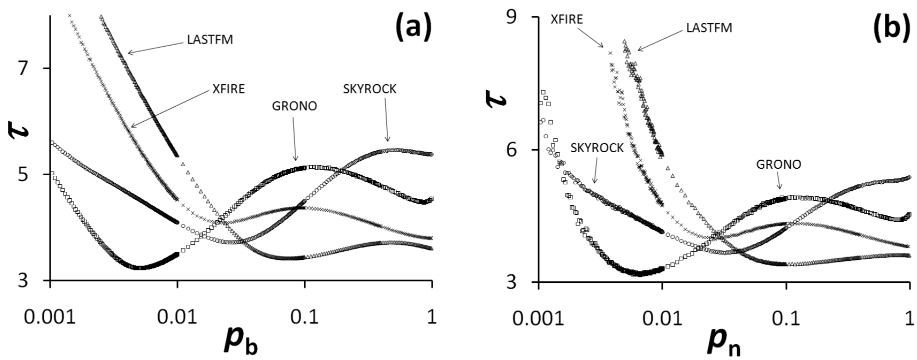

Fig. 4. The relationship between fraction of bonds $p_{\rm b}$ (a) or nodes $p_{\rm n}$ (b) that retained in network and the value of the exponent τ. The size of error bars is smaller than the size of marks. The uncertainty of the exponent τ is smaller than 1%.

3. Conclusions

Percolation in real-world networks is widely recognized as a key problem of interest. In this work we have investigated site and bond percolation in four real social networks. We have calculated approximated value of fraction $p_{\rm n}$ of nodes or $p_{\rm b}$ of bonds retained in the network at which giant component emerges. Note that in order to calculate the critical value of probability $p_{\rm c}$ with greater accuracy, the simulations for greater networks should be performed. However in our case it is impossible, because we investigate the percolation phenomenon in real networks and we cannot change the size of the system.

We have calculated the value of critical exponents β and τ. The relationship between the fraction of bonds or nodes retained in the network and value of exponent β was presented. The properties of the real, on-line social networks are different than uncorrelated scale-free networks. Contrary to results presented in [13] the values of critical exponents β and τ seems to be independent of the form of degree distribution. This result indicates that in the case of real social networks the internal structure of network (*e.g.* community structure) has much stronger influence on percolation phenomenon (*e.g.* the value of critical exponents) than the degree distribution. However it should be noted that the meaning of critical exponents and the percolation threshold is not very precise as result of finite size of systems under investigation.

Social network services nowadays become a new, very important medium of exchange of information between users. Some examples of possible applications of percolation theory in the case of on-line social networks are the information propagation in networks (*e.g.* rumor spreading, in such a case the values of parameters $p_{\rm b}$ and $p_{\rm n}$ correspond to the value of parameter that describes how interesting the rumor is) and viral marketing.

The paper was supported by the Polish Ministry of Science and Higher Education, Grant No. 496/N-COST/2009/0. A.G. wishes to thank Natalia Kruszewska for her help and useful discussions.

REFERENCES

[1] R. Albert, A-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).

[2] W. Bachnik, S. Szymczak, P. Leszczynski, R. Podsiadło, E. Rymszewicz, L. Kuryło, D. Makowiec, B. Bykowska, *Acta Phys. Pol. B* **36**, 3179 (2005).

[3] A. Grabowski, N. Kruszewska, R.A. Kosiński, *Eur. Phys. J.* **B66**, 107 (2008).

[4] D. Stauffer, A. Aharony, *Introduction to Percolation Theory*, 2nd ed., Taylor and Francis, London 1994.

[5] R. Albert, H. Jeong, A.L. Barabási, *Nature* **406**, 378 (2000).

[6] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).

[7] A. Grabowski, M. Rosińska, *Acta Phys. Pol. B* **37**, 1521 (2006).

[8] K. Malarz, Z. Szvetelszky, B. Szekf, K. Kułakowski, *Acta Phys. Pol. B* **37**, 3049 (2006).

[9] M.E.J. Newman, J. Park, *Phys. Rev.* **E68**, 036122 (2003).

[10] A. Clauset, C.R. Shalizi, M.E.J. Newman, *SIAM Review* **51**, 661 (2009) [`arXiv:0706.1062v2 [physics.data-an]`].

[11] A. Grabowski, R. Kosiński, *Acta Phys. Pol. B* **39**, 1291 (2008).

[12] M. Molloy, B. Reed, *Random Structures and Algorithms* **6**, 161 (1995).

[13] R. Cohen, D. ben-Avraham, S. Havlin, *Phys. Rev.* **E66**, 036113 (2002).

[14] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, *Phys. Rev.* **E64**, 066110 (2001).

[15] D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman, S.H. Strogatz, *Phys. Rev.* **E64**, 041902 (2001).