

Percolation in the classical blockmodel

Maksymilian Bujok, Piotr Fronczak, and Agata Fronczak

Faculty of Physics, Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland

Received: April 10, 2014/ Revised version: date

Abstract. Classical blockmodel is known as the simplest among models of networks with community structure. The model can be also seen as an extremely simply example of interconnected networks. For this reason, it is surprising that the percolation transition in the classical blockmodel has not been examined so far, although the phenomenon has been studied in a variety of much more complicated models of interconnected and multiplex networks. In this paper we derive the self-consistent equation for the size the global percolation cluster in the classical blockmodel. We also find the condition for percolation threshold which characterizes the emergence of the giant component. We show that the discussed percolation phenomenon may cause unexpected problems in a simple optimization process of the multilevel network construction. Numerical simulations confirm the correctness of our theoretical derivations.

PACS. 89.75.Fb Structures and organization in complex systems – 64.60.aq Networks – 64.60.ah Percolation

1 Introduction

For over a decade scientists of various disciplines have been showing an increasing interest in the field of complex networks [1,2,3]. It was related to the rapid development of the Internet, which in turn has made available a huge amount of data on the structure and functioning of many real networks, such as social networks, biological networks (e.g. food chains) and data communication networks (e.g. Internet), among others.

Initially, in the awareness of researchers, complex networks functioned as isolated systems. The structural and functional properties of individual networks were investigated. Only recently studying of multiplex network systems has begun in which individual networks may interact with each other [4,5]. An example of such interacting networks is a network that supplies energy to the computer network which in turn controls the energy distribution in the first network.

To understand the functioning of both types of networks (i.e. single networks and systems of interacting networks) the corresponding models are created and different dynamic processes are defined, such as the spread of epidemics and the opinion formation or diffusion processes. In many of these processes, the underlying phenomenon is percolation. For example, using percolation theory it has become possible to understand, why real networks are highly robust to random failures but fragile against attacks [6,7]. Recently, percolation theory was also used in a discussion on structural properties of the important class of multiplex network systems, see e.g. [8,9,10,11].

In this work we study the phenomenon of percolation in the so-called *classical blockmodel* which has a long tra-

dition of research in both social and computer sciences [12,13,14,15,16,17] and recently also in the study of complex networks [18,19,20,21,22]. The classical blockmodel was introduced by Holland, Laskey and Leinhardt in 1983 [12]. This model is interesting from the point of view of recent studies because it can be seen as a model of a single network with community structure [23,24], but also as a simple model of a two-level network which consists of smaller networks (i.e. network of networks) [4,5]. In such a two-level network, nodes may be connected through local edges (at the first level) and global ones (at the second level). This model can also be regarded as a generalization of the classical random graph of Erdős–Rényi (ER) wherein each of the N nodes is assigned to one of K blocks (communities, local area networks) of the same size. The probability of the existence of an edge is different for the nodes belonging to the same block (local level) and different, when nodes belong to different blocks (global level).

Further in the paper we deal with the following issues: In section 2 we recall a simple microscopic formalism described in our earlier work [25], which allowed us to calculate the percolation threshold and the size of the largest connected component in classical random graphs. This formalism, as one of many that have been used to describe the phenomenon of percolation in classical random graphs (see e.g. [26,27,28]), is in our opinion the simplest one. In section 3 we use this formalism to study percolation phase transition in the classical blockmodel. We determine the percolation threshold and the size of the largest connected component (i.e. percolation cluster). These results are compared with the results of numerical simulations. In section 4, we discuss a simple optimization process of the

construction procedure of two-level distribution networks. The procedure is based on the results obtained in section 3. Section 5 is devoted to the summary of results.

2 Percolation in classical random graphs

Classical random graphs were first discussed by Erdős and Rényi in the articles of the mid-twentieth century. In this paper, the term *classical random graph* refers to a certain generalisation of the original ER model: Classical random graph means the graph of N numbered vertices (nodes), where each pair of vertices is connected by an edge with probability p .

In this section, our goal is to remind one of the method [25], that allows to determine the size, N_G , of the largest cluster (i.e. the number of nodes belonging to the cluster) in classical random graphs. From previous works on this subject, we know that for $N \gg 1$, when the average node degree is less than one, $\langle k \rangle < 1$, the relative size of the largest cluster, $S = N_G/N$, is equal to zero. Only for $\langle k \rangle > 1$, the parameter S becomes greater than zero and increases, reaching the value of $S = 1$ when all nodes belong to the same cluster. The critical value of $\langle k \rangle = 1$ is called the percolation threshold and the largest connected component for $\langle k \rangle > 1$ is called the percolation cluster. In percolation theory, the relative size of the percolation cluster, S , which expresses the probability that a randomly selected node of the graph belongs to that cluster, acts as an order parameter of percolation phase transition.

To determine the value of parameter S , let us consider a randomly selected node i belonging to the ER graph. From the design procedure it follows that in the classical random graph all nodes and all edges are, from a statistical point of view, the same. This means that probability S , that the randomly chosen vertex i belongs to the largest cluster is equal to probability that at least one of $N - 1$ other vertices of the graph belongs to such a cluster and it is connected to the node i . Thus, if $A_{\{i,j\}}$ corresponds to the event that the edge $\{i,j\}$ (between the considered node i and a node j) belongs to the largest cluster, then the probability that the node i also belongs to this cluster becomes:

$$S = P\left(\bigcup_{j=1}^{N-1} A_{\{i,j\}}\right). \quad (1)$$

Since the events $A_{\{i,j\}}$ are independent, then using the known theorem on the sum of independent events [25,29], i.e.

$$P\left(\bigcup_{j=1}^{N-1} A_{\{i,j\}}\right) = 1 - \exp\left[-\sum_{j=1}^{N-1} P(A_{\{i,j\}})\right], \quad (2)$$

Eq. (1) can be written as:

$$S = 1 - \exp\left[-\sum_{j=1}^{N-1} P(A_{\{i,j\}})\right]. \quad (3)$$

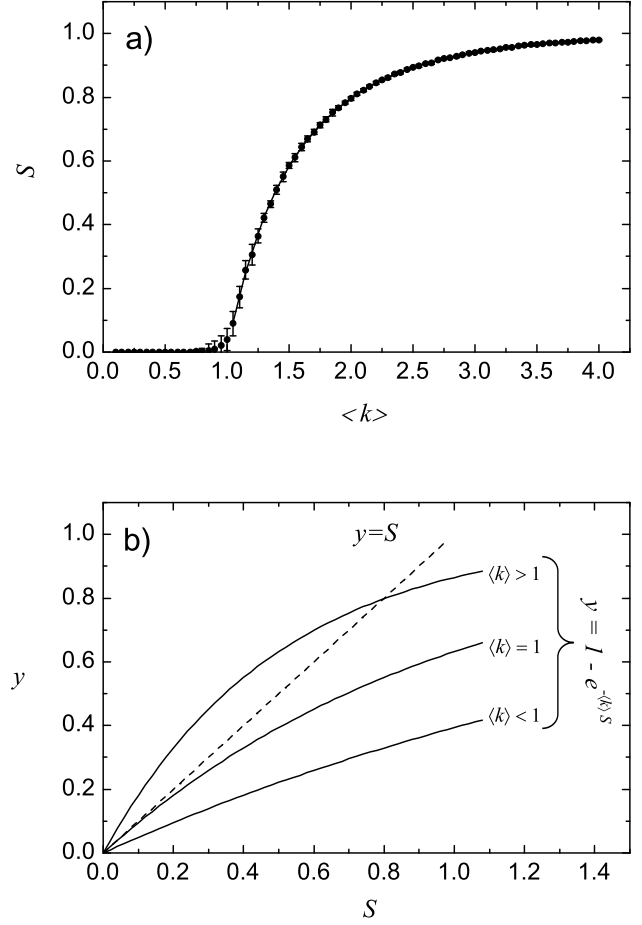


Fig. 1. Percolation in classical random graphs. a) Size of the giant component S versus the average degree $\langle k \rangle$ (the scattered points are results of numerical simulations and the solid line is theoretical solution obtained from Eq. (5)). b) Graphical solution of Eq. (5) for the size of the giant component (detailed description is given in the text).

Eq. (3) can be further simplified, using the fact that $P(A_{\{i,j\}})$ is equal to the product of probability p which means that the edge $\{i,j\}$ exists and probability S that j belongs to the largest connected component:

$$P(A_{\{i,j\}}) = pS. \quad (4)$$

Therefore, substituting the expression (4) to equation (3) we obtain the well-known self-consistency equation for the size of percolation cluster:

$$S = 1 - e^{-\langle k \rangle S}, \quad (5)$$

where $\langle k \rangle = p(N-1) \simeq pN$ is the average degree of nodes in in the classical random graph.

For the first time, the equation (5) was given by Erdős and Rényi in 1959 [26]. It gives the relative size of the giant component for any given value of the mean degree $\langle k \rangle$. However, although this equation is very simple it does not have a simple solution for the size of the giant component as a function of $\langle k \rangle$ in closed form. The numerical solution of Eq. (5) for S as compared with results of Monte

Carlo simulations is shown in Fig. 1a. Fortunately, graphical solution of this equation, which is given in Fig. 1b, in a very suggestive way illustrates the percolation transition in classical random graphs (see also Fig. 12.1 in [3] and Fig. 3 in [25]). The three solid curves shown in Fig. 1b represent the right hand side of Eq. (5), i.e. the function $y(S) = 1 - e^{-\langle k \rangle S}$, for different values of $\langle k \rangle$. The dashed line in the figure is the linear function $y(S) = S$. Where the line and the curve cross, the corresponding value of S is a solution to Eq. (5).

As the figure shows, depending on the value of $\langle k \rangle$ there may be either one solution for S or two. For small $\langle k \rangle$ (bottom curve in Fig. 1b) the only solution is $S = 0$, which means that the percolation cluster does not exist. For sufficiently large $\langle k \rangle$ (top curve), the equation (5) has two solutions, $S = 0$ and $S \neq 0$, the first of which is unstable. It means, that the percolation cluster of size $S \neq 0$ appears in a graph. The middle curve in the figure corresponds to the transition point between the two regimes. The characteristic point (the percolation threshold) is where the gradient of the curve $y(S) = 1 - e^{-\langle k \rangle S}$ and the slope of the dashed line $y(S) = S$ match at $S = 0$. Therefore, the point can be determined from the first derivative of the right-hand side of Eq. (5):

$$\left. \frac{d(1 - e^{-\langle k \rangle S})}{dS} \right|_{S=0} = 1. \quad (6)$$

From the above equation one immediately finds that the percolation threshold in classical random graphs is given by:

$$\langle k \rangle = 1. \quad (7)$$

3 Percolation in the classical blockmodel

In the classical blockmodel (see Fig. 2), each of N nodes is assigned to one of K blocks of the same size R , i.e. $N = KR$. In the model, blocks (modules) are ER graphs. The probability p that there is an edge between the nodes belonging to the same block is usually different from the probability q that there is an edge between the nodes belonging to different blocks. Further in this paper, the edges connecting the nodes that belong to the same block will be called *local* connections while the edges connecting nodes from different blocks will be referred to as *global* ones. Similarly, when we talk about percolation inside the blocks and at the level of local connections we will use the notion of the *local percolation cluster*. The *global percolation cluster* will be called the largest connected component of the whole graph. Such a cluster will be built with the nodes belonging to different blocks, which are connected through both local or global edges.

In this section our aim is to find the expression for percolation threshold and to calculate the relative size of the global percolation cluster in the classical blockmodel. In order to do it we will use a method similar to that of the previous section which allowed us to describe the percolation transition in classical random graphs.

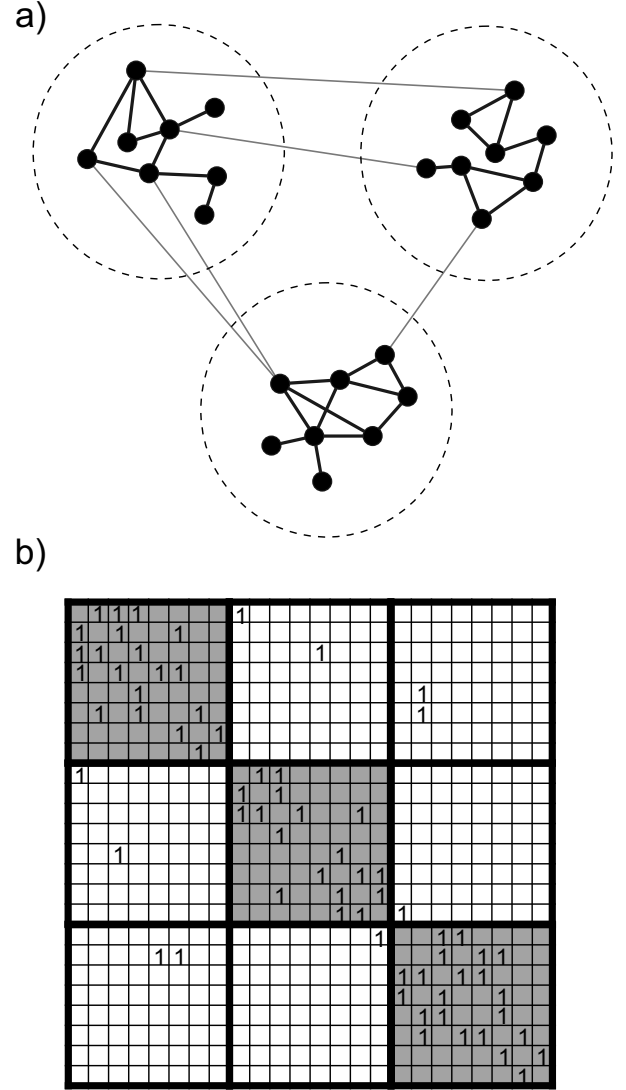


Fig. 2. Classical blockmodel. a) A small network with block structure of the type considered in this paper. In this case, there are three blocks $K = 3$ of size $R = 8$, denoted by the dashed circles, which have dense local connections. Global connections between the blocks are sparse. b) The adjacency matrix of the graph shown in Fig. 1a. Gray areas along the diagonal represent adjacency matrices for the nodes belonging to the same blocks. Non zero matrix elements occurring outside the gray areas represent global connections.

Thus, let G be the probability that a randomly chosen node i in the classical blockmodel belongs to the global percolation cluster. When we take into account only local connections of this node, there are only two possible cases: a) the node i belongs to the local percolation cluster of size S , and b) the node belongs to one of the smaller local clusters of size s , wherein the probability that a randomly chosen vertex belongs to a component of size s is given by (see Eq. (12.50) in [3]):

$$\pi_s = \frac{e^{-\langle k \rangle s} (\langle k \rangle s)^{s-1}}{s!}, \quad (8)$$

where

$$\sum_s \pi_s = 1 - S, \quad (9)$$

and

$$\langle k \rangle = pR \quad (10)$$

is the mean *local degree* of nodes in the classical blockmodel¹. In both the cases, the probability that the node i belongs to the global percolation cluster is equal to the probability that node i itself or at least one of the nodes (e.g., node j) which belong to the same local cluster (percolation or not) are connected by a global edge to a node belonging to the global percolation cluster.

Therefore, let B_{jl} be the event that the global edge $\{j, l\}$ between the nodes j and l belongs to the global percolation cluster (the node j should be thought of as a node belonging to the same block as well as to the same local cluster as the node i). It follows that the parameter G , which is equal to the probability that a randomly selected node is a global percolation cluster is equal to:

$$G = SP \left(\bigcup_{j=1}^{SR} \bigcup_{l=1}^{N-R} B_{\{j,l\}} \right) + \sum_s \pi_s P \left(\bigcup_{j=1}^s \bigcup_{l=1}^{N-R} B_{\{j,l\}} \right), \quad (11)$$

where the summation over j is carried out over all the nodes belonging to the same local cluster to which the node i belongs while the summation over l is carried out over the nodes belonging to other modules.

Then, using the theorem on the sum of independent events, cf. Eq. (2), and proceeding similarly as in Sec. 2, the expression (11) can be simplified to:

$$G = S \left(1 - \exp \left[- \sum_{j=1}^{SR} \sum_{l=1}^{N-R} P(B_{\{j,l\}}) \right] \right) + \sum_s \pi_s \left(1 - \exp \left[- \sum_{j=1}^s \sum_{l=1}^{N-R} P(B_{\{j,l\}}) \right] \right), \quad (12)$$

where

$$P(B_{\{j,l\}}) = qG, \quad (13)$$

is the probability that a node l at the end of the global edge $\{j, l\}$ belongs to the global percolation cluster. Finally, substituting Eq. (13) to (12), we get the self-consistent equation for G :

$$G = S \left(1 - e^{-qSR(N-R)G} \right) + \sum_s \pi_s \left(1 - e^{-qs(N-R)G} \right) = f(G). \quad (14)$$

The right hand side of Eq. (14), i.e. the function $f(G)$, is continuous and monotonically increasing for G in the range from 0 to 1. In fact, the behavior of this function is very similar to the behavior of the right hand side of

¹ cf. the so-called *internal* and *external* node degrees in blockmodels, as they were defined in [22];

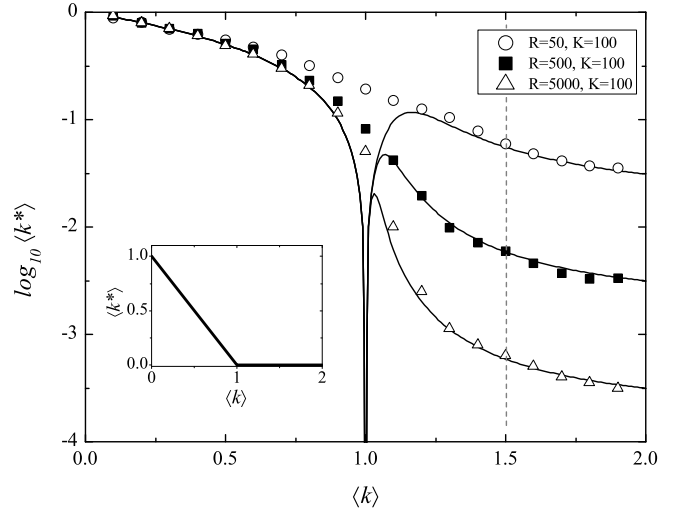


Fig. 3. Phase diagram for percolation transition in the classical blockmodel. Points placed in the figure represent results of numerical simulations for three different network sizes and correspond to values of the parameters $\langle k \rangle$ and $\langle k^* \rangle$ for which the largest cluster (averaged over 10 network realizations) starts to form. Solid lines correspond to Eq. (15).

Eq. (5) which was discussed in Sec. 2. In particular, it is easy to see that the value of $G = 0$ (which means lack of the global percolation cluster) is always the solution of Eq. (14). Furthermore, for certain values of p, q, R, K parameters, a non-zero solution, $G \neq 0$, of this equation appears which characterizes the emergence of the global percolation cluster.

Thus, similarly as in classical random graphs, the condition for the emergence of the global percolation cluster in the classical blockmodel may be obtained by comparing the first derivatives on both sides of Eq. (14) at $G = 0$. After simple algebra the condition for the global percolation transition, i.e. the point at which a sample-spanning global cluster first appears, takes the following form:

$$\langle k^* \rangle = \frac{1}{S^2 R + (1 - S)\langle s \rangle}, \quad (15)$$

where

$$\langle k^* \rangle = q(N - R), \quad (16)$$

is the mean number of global connections attached to a node, i.e. the average *global degree* of nodes in the classical blockmodel, whereas $\langle s \rangle$ is the mean size of the local component to which a randomly chosen node belongs (see Eq. (12.34) in [3]),

$$\langle s \rangle = \frac{\sum_s s \pi_s}{\sum_s \pi_s} = \frac{1}{(1 - S)(1 - \langle k \rangle) + \langle k \rangle S}. \quad (17)$$

Figure 3 shows the results of numerical simulations for percolation threshold in the classical blockmodel against the theoretical curves predicted by Eq. (15). Notable is that while the theoretical and numerical results agree for $\langle k \rangle < 1$ and $\langle k \rangle > 1$, for $\langle k \rangle$ close to 1 the theoretical

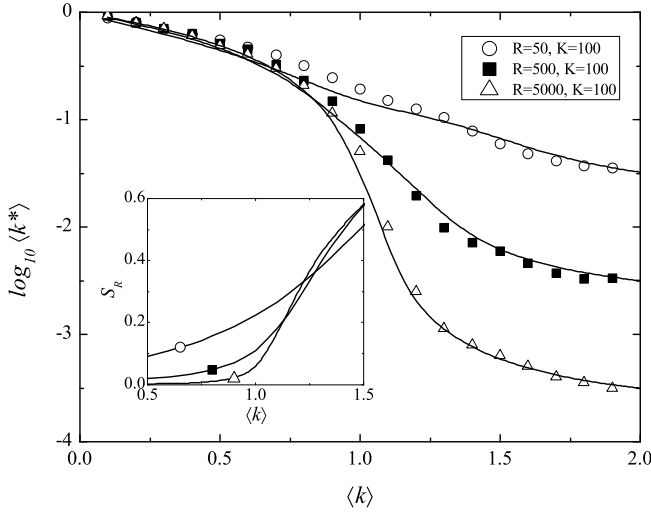


Fig. 4. Percolation threshold in the finite-size classical blockmodel. Symbols correspond to the same numerical results as shown in Fig. 3. Solid lines represent Eq. (15), in which the parameter S characterizing the size of the giant component in infinite networks was replaced by the parameter S_R , which applies to finite networks. In the inset, the numerical values of S_R for different network sizes (as indicated by the respective symbols) are shown.

curves diverge, which is not in agreement with numerical data. This discrepancy is due to the fact that Eq. (17) for $\langle s \rangle$ diverges at $\langle k \rangle = 1$. At this point, $S = 0$ and the denominator of Eq. (17) vanishes. However, this is only true for infinite graphs. In finite ER graphs, the size of the largest cluster $S_R > 0$ for $\langle k \rangle = 1$ (see inset in Fig. 4). If we replace S in Eq. (15) with values of S_R obtained from numerical simulations we get rid of the divergence problem (see Fig. 4).

In Fig. 3, the dashed line indicates the fixed value of the parameter $\langle k \rangle = 1.5$ for which the relative size of the global percolation cluster, G , is examined against the theoretical prediction of the self-consistent mean-field version of Eq. (14):

$$G \simeq S \left(1 - e^{-SR\langle k^* \rangle G} \right) + (1 - S) \left(1 - e^{-\langle s \rangle \langle k^* \rangle G} \right). \quad (18)$$

Figure 5 shows comparison between the theoretical curves given by Eq. (18) and the numerically obtained values of G as a function of $\langle k^* \rangle$ for three different network sizes. The numerical simulations provide an independent test for correctness of our theoretical calculations.

4 Network optimization procedure

In this section we discuss a possible application of the discussed percolation phenomenon, namely a network optimization process which has been widely studied in recent years [30,31,32]. Such an optimization is of common interest in many different areas, among them electrical engineering, telecommunication, road construction and trade

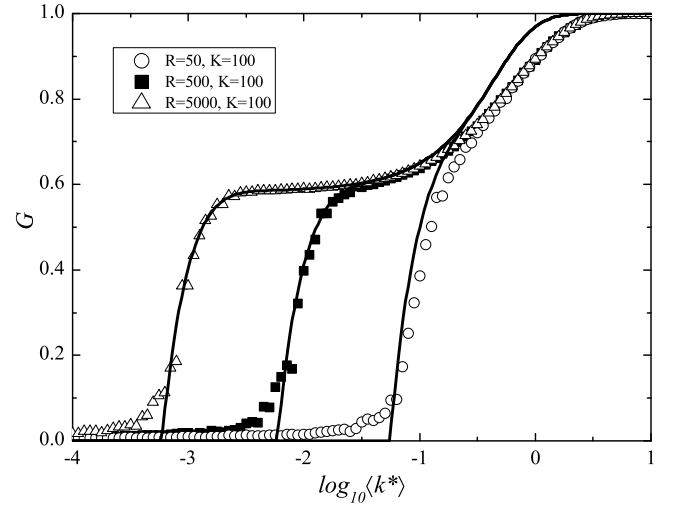


Fig. 5. Percolation in the classical blockmodel. Size of the global percolation cluster G versus the average global degree $\langle k^* \rangle$ for the fixed value of $\langle k \rangle = 1.5$. The solid lines represent theoretical solution of Eq. (18) for G .

logistics. In the rest of this section we will concentrate on the economic optimization process which typically aims to reduce costs while increasing revenues.

Thus, let us consider a business entity such as a network operator or a network manager who earns profit from connecting local or regional networks. Its revenue is related to the number in interconnected customers, i.e. it is proportional to the size of the giant connected component G , while the costs arise from constructing and maintaining links between regional networks, i.e. they depend on the number of links $L = \langle k^* \rangle N$ and this dependence is, according to the so-called economies-of-scale effects, sub-linear. Such a sublinear behavior arises from the fact that the decrease in unit cost of a product or service results from large-scale operations and the fixed costs are spread out over more units of output [33]. In the following, for the purpose of better demonstration, we assume that the cost scales logarithmically with L . Taking above into consideration, one has to optimize the following equation:

$$C = (1 - \lambda) \log(L) - \lambda G, \quad (19)$$

where λ is a parameter controlling a ratio between expenses and income parts. The last equation states, that the network manager has to find an optimal link density considering two contradictory demands: an expensive to maintain, densely connected network which integrates all the potential customers or inexpensive, sparse network which brings little income.

The balance (or cost) function, Eq. (19), for $\lambda = 0.97$ is presented in Fig. 6. During the first phase of the network construction (for $L < 40$), the growing costs represent an initial investment a business owner needs to start up a firm. In the second phase one can see the two well separated minima of the cost function. They mean that the network provider who tries to operate in accordance with the economic rule (19) has to remember that the expand-

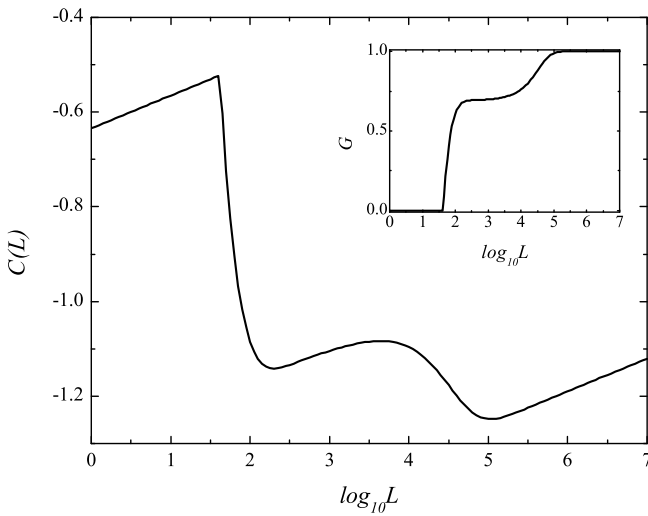


Fig. 6. Cost function C versus the number of global connections L for the network characterized by $R = 2000$, $K = 20$ and $\lambda = 0.97$. Inset: size of the global percolation cluster G , Eq. (18), versus L for the same network.

ing the business can lead to a temporal increase of costs and can be discouraging since one has to pass over the cost barrier.

5 Summary

The classical blockmodel is the simplest among models of networks with community structure. Majority of scientists, who are engaged in the research of complex networks, knows this model mainly due to the fact that over the last decade, it was often used as a benchmark graph for testing community detection algorithms [18, 19, 20]. However, in our opinion, this model deserves attention because it is also an extremely simple example of interconnected networks. For this reason, it is surprising that the percolation transition in this model has not been examined so far, although the phenomenon has been studied in a variety of much more complicated models of interconnected and multiplex networks [8, 9, 10, 11].

In summary, in this paper we study percolation in the classical blockmodel. At the beginning, in Sec. 2, we present a general formalism and apply it to the analytical calculation of the structural properties of classical random graphs. Next, in Sec. 3, we use this formalism to obtain the self-consistent equation for the size, G , of the global percolation cluster in the classical blockmodel. The formalism allows to calculate the position of the phase transition (i.e. percolation threshold) at which the sample-spanning global cluster first appears. The carried out numerical simulations confirm the correctness of our theoretical predictions. Finally, in Sec. 4 we show, how our theoretical derivations may help to understand the cost optimization procedure in distribution networks which have a modular structure.

6 Acknowledgments

This work was supported by the Foundation for Polish Science (grant no. POMOST/2012-5/5) and by the European Union within European Regional Development Fund (Innovative Economy).

References

1. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks. From Biological Nets to the Internet and WWW*, Oxford University Press, 2003.
2. A. Barrat, A.V.M. Barthélemy, *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge, 2008.
3. M.E.J. Newman, *Networks. An Introduction*, Oxford University Press, Oxford, 2010.
4. S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, S. Havlin, *Nature* **464**, 08932 (2010).
5. J. Gao, S.V. Buldyrev, H.E. Stanley, S. Havlin, *Nature Phys.* **8**, 40 (2012).
6. R. Cohen, K. Erez, D. ben-Avraham, *Phys. Rev. Lett.* **85**, 4626 (2000).
7. R. Cohen, K. Erez, D. ben-Avraham, *Phys. Rev. Lett.* **86**, 3682 (2001).
8. R. Parshani, S.V. Buldyrev, S. Havlin, *Phys. Rev. Lett.* **105**, 048701 (2010).
9. J. Gao, S.V. Buldyrev, S. Havlin, H.E. Stanley, *Phys. Rev. Lett.* **107**, 195701 (2011).
10. S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, M. Paczuski, *Europhys. Lett.* **97**, 16006 (2012).
11. G. Bianconi, S.N. Dorogovtsev, arXiv:1402.0218 [cond-mat.stat-mech].
12. P.W. Holland, K.B. Laskey, S. Leinhardt, *Soc. Networks* **5**, 109 (1983).
13. K. Faust, S. Wasserman, *Soc. Networks* **14**, 5 (1992).
14. C.J. Anderson, S. Wasserman, K. Faust, *Soc. Networks* **14**, 137 (1992).
15. T.A. Snijders, K. Nowicki, *J. Classification* **14**, 75 (1997).
16. E.M. Airoldi, D.M. Blei, S.E. Fienberg, X.P. Xing, *J. Mach. Learn. Res.* **9**, 1981 (2008).
17. A. Goldenberg, A.X. Zheng, S.E. Feinberg, E.M. Airoldi, *Found. Trends Mach. Learn.* **2**, 1 (2009).
18. M. Girvan, M.E.J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).
19. M.E.J. Newman, M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
20. A. Lancichinetti, S. Fortunato, F. Radicchi, *Phys. Rev. E*, **78**, 046110 (2008).
21. B. Karrer, M.E.J. Newman, *Phys. Rev. E* **83**, 016107 (2011).
22. P. Fronczak, A. Fronczak, M. Bujok, *Phys. Rev. E* **88**, 032810 (2013).
23. S. Fortunato, *Phys. Rep.* **486**, 75 (2010).
24. M.E.J. Newman, *Nature Phys.* **8**, 25 (2012).
25. A. Fronczak, P. Fronczak, J.A. Hołyst, *AIP Conf. Proc.* **776**, 52 (2005). arXiv:cond-mat/0502663 [cond-mat.dis-nn].
26. P. Erdős, A. Rényi, *Publicationes Mathematicae* **6**, 290 (1959).
27. M. Molloy, B. Reed, *Comb. Prob. and Comput.* **7**, 295 (1998).

- 28. D.S. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
- 29. A. Fronczak, P. Fronczak, J.A. Holyst, Phys. Rev. E **70**, 056110 (2004).
- 30. M. T. Gastner and M. E. J. Newman, Eur. Phys. J. B **49**, 247 (2006).
- 31. R. Ferrer i Cancho and R. Solé, *Statistical Physics of Complex Networks, Lecture Notes in Physics*, Springer, Berlin, 2003.
- 32. F. Schweitzer, *Brownian Agents and Active Particles*, Springer, Berlin, 2003.
- 33. G. M. Gelles and D. W. Mitchell, J. Econ. Educ. **27**(3), 259 (1996).