

Cut-offs and finite size effects in scale-free networks

M. Boguñá^{1,a}, R. Pastor-Satorras², and A. Vespignani³

¹ Departament de Física Fonamental, Universitat de Barcelona, Avinguda Diagonal 647, 08028 Barcelona, Spain

² Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, 08034 Barcelona, Spain

³ Laboratoire de Physique Théorique (UMR 8627 du CNRS), bâtiment 210, Université de Paris-Sud, 91405 Orsay Cedex, France

Received 18 November 2003

Published online 24 February 2004 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2004

Abstract. We analyze the degree distribution's cut-off in finite size scale-free networks. We show that the cut-off behavior with the number of vertices N is ruled by the topological constraints induced by the connectivity structure of the network. Even in the simple case of uncorrelated networks, we obtain an expression of the structural cut-off that is smaller than the natural cut-off obtained by means of extremal theory arguments. The obtained results are explicitly applied in the case of the configuration model to recover the size scaling of tadpoles and multiple edges.

PACS. 89.75.-k Complex systems – 87.23.Ge Dynamics of social systems – 05.70.Ln Nonequilibrium and irreversible thermodynamics

1 Introduction

Recent years have witnessed an increasing scientific interest for the study of complex networks and the dynamical processes taking place on top of them [1,2]. Indeed, the complex topological properties shown by many real networks have large effects on the behavior of several phenomena characterizing the dynamics and stability of these systems. These effects are particularly intriguing in the case of scale-free (SF) networks, that is, in networks in which the probability $P(k)$ that a vertex is connected to k other vertices (the degree distribution) scales as a power law, $P(k) \sim k^{-\gamma}$ [1,2]. In general, *uncorrelated* SF networks with a degree exponent $\gamma \leq 3$, exhibit the lack of epidemic and percolation threshold, that can be identified in terms of critical phenomena with the absence of any critical point [3–5]. These results have been generalized to several epidemic and percolation models [6–8], even in the presence of correlations [9,10], and in critical phenomena [11,12]. In all these systems, the absence of the critical point finds an explanation in the diverging second moment $\langle k^2 \rangle$ of the degree distribution of SF networks with $\gamma \leq 3$, that implies unbounded degree fluctuations in the limit of infinite network size $N \rightarrow \infty$.

Real networks, however, have always a finite number of vertices N , and it has been pointed out that finite size effects reintroduce a positive threshold in critical processes [6,13]. Indeed, the finite size of real networks introduces a bound in the possible values of the degree, depending on the system size N , which has the effect of restoring a limit in the degree fluctuations, re-inducing in this way a

non-zero critical point. Therefore, the existence of bounds for the maximum degree becomes a relevant element in order to estimate the critical properties of dynamical systems defined in networks with SF topologies.

The presence of bounded SF distributions in complex networks has been observed in several systems [14]. In some cases, the bound or degree cut-off can be explained in terms of a finite capacity of the vertices to collect connections or due to incomplete information [14,15]. In this case, the value of the cut-off is a constant that depends on the physical constraints acting on the systems. A second possibility, the one in which we are interested here, takes place when the cut-off is purely accounted for by the finite size of the network, as usually happens in growing networks, that have grown up to a maximum number of vertices N [16,17].

In this paper we will reconsider the nature of the cut-offs due to finite-size effects in SF networks. We will review how to estimate this cut-off in terms of extreme value theory, and point out how this estimate is affected when one takes into account the topological structure of the network. Our considerations will be illustrated by analyzing an example of uncorrelated network model.

2 Extreme value theory and the natural cut-off

The nature of the degree cut-off in finite-size SF networks has been considered in several instances in network theory. For example, Aiello et al. [18] proposed to define a cut-off k_m as the value of the degree for which we expect to

^a e-mail: mbogunya@ffn.ub.es

observe at most one vertex, that is

$$NP(k_m) \sim 1. \quad (1)$$

For a SF network, this expression provides a dependence of the cut-off with N as

$$k_m(N) \sim N^{1/\gamma}. \quad (2)$$

This definition, however, lacks some mathematical rigor, since it considers the probability of a single point in a probability distribution, which is not completely well-defined in the continuous k limit for large N .

A more physical definition of cut-off was given by Dorogovtsev et al. [19], defining it as the value of the degree k_c above which one expects to find at most one vertex,

$$N \int_{k_c}^{\infty} P(k) dk \sim 1. \quad (3)$$

In this case, we obtain

$$k_c(N) \sim N^{1/(\gamma-1)}, \quad (4)$$

which is known as the *natural* cut-off of the network.

The origin of the natural cut-off, as well as the constraints that must be imposed to validate its accuracy, can be better understood in terms of extreme value theory. Indeed, if we have a random variable distributed according to the probability density $\rho(x)$ and we draw N observations of this quantity, $\{x_i\}$, $i = 1 \dots N$, the maximum value of this sample, $\max\{x_i\}$, will, in turn, be a random variable. Extreme value theory is aimed at finding the statistical properties of this maximum. In the simplest case, the sample is built up of independent events and the distribution function, giving the probability that $\max\{x_i\} < X$, is simply given by

$$\Pi(X) = \left\{ \int_0^X \rho(y) dy \right\}^N. \quad (5)$$

The probability density $\pi(X)$ that $\max\{x_i\}$ is equal to X is therefore the derivative $d\Pi(X)/dX$ and the cut-off is then defined as the average value of the extreme value of the sample, that is,

$$x_c(N) = \int X \rho(X) N \left\{ \int_0^X \rho(y) dy \right\}^{N-1} dX. \quad (6)$$

This cut-off is always an increasing function of the sample size and gives us information about the finite size effects of the process under study. In a scale-free network, the probability density $\rho(x)$ corresponds to the degree distribution, that scales as $P(k) \sim k^{-\gamma}$. By substituting $P(k)$ in equation (6), we recover the natural cut-off scaling given by equation (4). Even though equations (6) and (3) give slightly different exact values for $k_c(N)$, both lead to the correct dependence on the system size N , which is, in most cases, the relevant information.

It is important to recall that this form of the natural cut-off is obtained under the assumption that all the elements of the sample are independently drawn from the probability density $P(k)$. However, in real networks the degrees of the vertices are not simply independently drawn from a probability distribution $P(k)$, but must satisfy some topological constraints due to the network structure. Thus, we must include also the structure of the connections when considering the scaling of the cut-off.

3 Structural properties of networks and the structural cut-off

In order to shed some light on this problem we first need to characterize some structural aspects of networks [20,21]. In what follows we will consider undirected sparse networks, that is, networks with a well-defined thermodynamic limit (or, equivalently, constant average degree $\langle k \rangle$), with N vertices. Let us define N_k as the number of vertices of degree k . This quantity satisfies $\sum_k N_k = N$ which, in the thermodynamic limit ($N \gg 1$), allows to define the degree distribution as $P(k) = N_k/N$. The degree distribution $P(k)$ contains only information about the local properties of vertices, that is, the number of edges that emanate from each vertex. Thus, we also need to specify how different degree classes are connected to each other. To this end, we define the symmetric function $E_{kk'}$, that gives the number of edges between vertices of degree k and k' , for $k \neq k'$, and two times the number of self-connections¹ for $k = k'$. This matrix fulfills the identities

$$\sum_{k'} E_{kk'} = kN_k, \quad (7)$$

$$\sum_{k,k'} E_{kk'} = \langle k \rangle N = 2E, \quad (8)$$

where E is the total number of edges in the network. This last identity allows to define, again in the limit $N \gg 1$, the *joint distribution*

$$P(k, k') = \frac{E_{kk'}}{\langle k \rangle N}, \quad (9)$$

where the symmetric function $(2 - \delta_{k,k'})P(k, k')$ is the probability that a randomly chosen edge connects two vertices of degrees k and k' . It is easy to see that, in fact, the degree distribution $P(k)$ can be derived from the joint distribution $P(k, k')$ as

$$P(k) = \frac{\langle k \rangle}{k} \sum_{k'} P(k, k'). \quad (10)$$

Therefore, the joint distribution conveys all the information at the degree-degree level. In particular, the assortative (disassortative) character of the correlations in the

¹ By self-connections we mean connections between vertices in the same degree class.

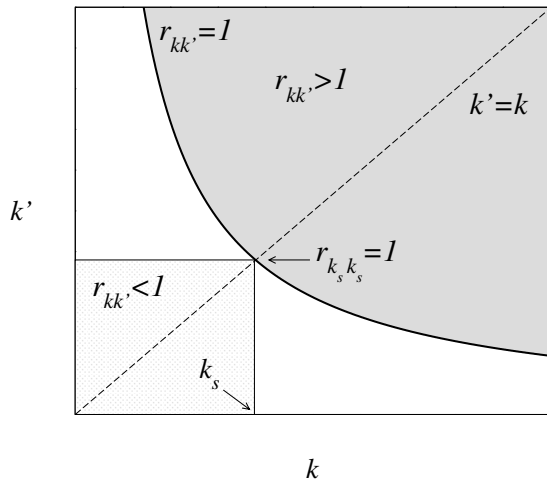


Fig. 1. Geometrical construction of the structural cut-off k_s .

network, that is, the tendency of vertices to connect to vertices of the same (different) degree class, can be quantified by means of the Pearson coefficient r , defined as the correlation coefficient of the joint distribution $P(k, k')$ [22, 23]. It is worth mentioning that all the above quantities are defined outside the context of any specific model, which make them completely general.

To proceed further, let us define $r_{kk'}$ as the ratio between the actual number of edges between vertices of degrees k and k' , $E_{kk'}$, and the maximum value for this quantity, $m_{kk'}$. Assuming that multiple edges are not allowed in the network the maximum number of edges between two degree classes is² $m_{kk'} = \min\{kN_k, k'N_{k'}, N_kN_{k'}\}$ and, consequently, the ratio $r_{kk'}$ can be written as

$$r_{kk'} = \frac{E_{kk'}}{m_{kk'}} = \frac{\langle k \rangle P(k, k')}{\min\{kP(k), k'P(k'), NP(k)P(k')\}}. \quad (11)$$

A key property of this ratio is that it must be smaller than or equal to 1 for any values of k and k' , regardless of the type of network. We can use this simple observation to draw some conclusions over the value of the cut-off imposed by the structure of the network. Let us consider, see Figure 1, the space k - k' in which the joint distribution $P(k, k')$ is defined. The curve $r_{kk'} = 1$ defines the boundary³ separating the region in which the pairs (k, k') take admissible values ($r_{kk'} \leq 1$) from the unphysical region $r_{kk'} > 1$. If we define a *structural* cut-off k_s as the value of the degree delimiting the largest square region of admissible values, we obtain that it is given as the intersection

² While the restriction of not having multiple edges may appear unnecessary under a mathematical point of view, it is usually observed in real networks where redundant edges are not considered as a part of the network. If multiple edges are instead allowed $m_{kk'}$ is simply given by $m_{kk'} = \min\{kN_k, k'N_{k'}\}$.

³ For simplicity, we have assumed that this boundary is given by a smoothly decreasing concave function. The same result applies for convex boundaries. More complex situations can be considered along the same lines of reasoning.

of the curves $r_{kk'} = 1$ and $k' = k$. That is, the structural cut-off can be defined as the solution of the implicit equation

$$r_{k_s k_s} = 1. \quad (12)$$

In the following, we will discuss the implications of the structural cut-off defined in the previous expression.

It is worth noticing that as soon as $k > NP(k')$ and $k' > NP(k)$ the effects of the restriction on the multiple edges are already being felt, turning the expression for $r_{kk'}$ to

$$r_{kk'} = \frac{\langle k \rangle P(k, k')}{NP(k)P(k')}. \quad (13)$$

In the case of interest of SF networks these conditions are fulfilled in the region $k, k' > (\alpha N)^{1/(\gamma+1)}$ (where α is constant depending on the details of the function $P(k)$), well below the natural cut-off. As a consequence, this scaling behavior provides a lower bound for the structural cut-off of the network, in the sense that, whenever the cut-off of the degree distribution falls below this limit, the condition $r_{kk'} < 1$ is always satisfied.

3.1 Uncorrelated networks

Let us first analyze the class of uncorrelated networks. In this case the joint distribution factorizes as

$$P_{nc}(k, k') = \frac{kk'P(k)P(k')}{\langle k \rangle^2} \quad (14)$$

which, in turn, implies that the ratio $r_{kk'}$ takes the simple form [21]

$$r_{kk'} = \frac{kk'}{\langle k \rangle N}. \quad (15)$$

In this case, the structural cut-off needed to preserve the physical condition $r_{kk'} \leq 1$ takes the form

$$k_s(N) \sim (\langle k \rangle N)^{1/2}, \quad (16)$$

independent of the degree distribution, and in particular, of the degree exponent γ in SF networks. This structural cut-off, which has been already discussed in the context of the random network model proposed by Chung and Lu [24, 25], and also by reference [26], coincides with the natural cut-off when the exponent of the degree distribution is $\gamma = 3$ (for instance, the Barabási-Albert network [27]). For $\gamma > 3$, the structural cut-off diverges faster than the natural cut-off, and therefore the latter should be selected as the appropriate one [26]. For $\gamma < 3$, however, the exponent of the natural cut-off is greater than 1/2 and, as a consequence, the cut-off predicted by extreme value theory is diverging faster than the structural one. In other words, this means that uncorrelated SF networks without multiple edges and exponent $\gamma < 3$ must possess a cut-off that behaves as the structural cut-off and is thus smaller than the one predicted by the extreme value theory. If this is not the case, that is, if the actual cut-off is imposed to be larger than the structural cut-off k_s , this

means that the network is not totally uncorrelated and some negative correlations, such as those observed in the Internet [28], must appear in order to fulfill the constraint $r_{kk'} \leq 1$ [25,29].

3.2 Correlated networks

For correlated networks the position of the cut-off will depend, in general, on the nature of the correlations through the specific form of $r_{kk'}$. For assortative networks, that is, networks with positive degree correlations, the cut-off must be even smaller than the uncorrelated one since, in these class of networks, high degree vertices connect preferably to other high degree vertices increasing, thus, the value of $r_{kk'}$ at $k \simeq k'$. This could explain, for instance, the appearance of an abrupt cut-off after a power law regime in many social networks, such as the network of movie actors [30], which have been found to show assortative mixing. In the opposite case of disassortative networks the cut-off can be higher because, in this case, high degree vertices connect preferably to low degree ones and, as a consequence, the ratio $r_{kk'}$ is reduced near the region $k \simeq k'$. In summary, the structural cut-off for SF networks lies somewhere within the interval $[(\alpha N)^{1/(\gamma+1)}, (\beta N)^{1/(\gamma-1)}]$, where α and β are characteristic constants, depending on their correlation structure. In particular, $k_s \sim N^{1/2}$ corresponds to uncorrelated networks whereas greater values of the exponent indicate disassortative correlations and smaller values assortative mixing by degree [16,17].

4 The configuration model

In order to check our previous arguments, we shall consider the configuration model (CM), proposed by Molloy and Reed [31,32] as a practical algorithm to generate uncorrelated random networks with a designed degree distribution. The model first generates a sequence of N degrees, which are independently drawn from the distribution $P(k)$, and then it proceeds by connecting pairs of randomly chosen edge ends. At the same time, however, the model generates a number of multiple edges and *tadpoles*, that is, edges connected to the same vertex at both ends. When the degree distribution has a finite second moment the fraction of multiple edges and tadpoles over the total number of edges vanishes in the thermodynamic limit and, as a consequence, they can be neglected. For SF networks with exponent $\gamma < 3$, the situation is different [29] as it can be shown by applying the reasoning concerning the scaling of the degree distribution cut-off. The problematic degree classes are those connecting vertices of degrees satisfying

$$kk' > \langle k \rangle N. \quad (17)$$

For an uncorrelated SF network with degree exponent $2 < \gamma < 3$, the total number of edges satisfying this condition

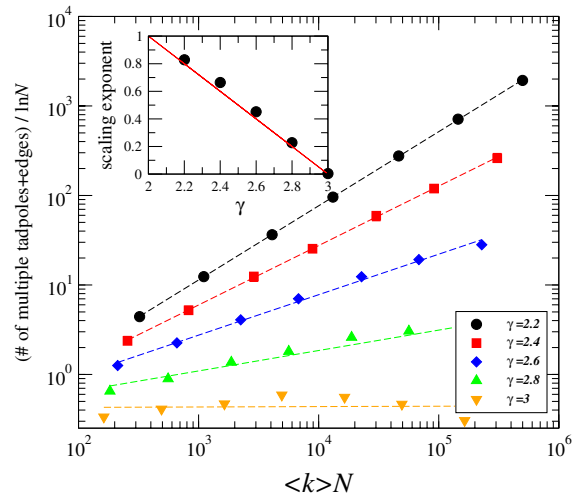


Fig. 2. Number of multiple edges plus tadpoles satisfying condition (17) as a function of the size of the network, $\langle k \rangle N$, for different values of the exponent γ in the configuration model. The inset shows the scaling exponents compared to the theoretical value $3 - \gamma$. Each point corresponds to an average over 10^4 networks.

can be calculated, in the limit $N \gg 1$, as

$$\begin{aligned} \sum_{kk' > \langle k \rangle N} E_{kk'} &\sim N \langle k \rangle \int dk \int_{\langle k \rangle N/k}^{\infty} dk' P_{nc}(k, k') \\ &\sim (\langle k \rangle N)^{3-\gamma} \ln N. \end{aligned} \quad (18)$$

These edges correspond to an unphysical situation and therefore must be balanced by a similar scaling of multiple edges plus tadpoles in order to restore a physical structure for the network connectivity. We thus expect the same scaling law for the number of multiple edges plus tadpoles satisfying condition (17) generated by the CM algorithm. In order to check this *ansatz* we have performed numerical simulations using the CM for exponents $\gamma < 3$ and we have computed the number of multiples edges plus tadpoles as a function of the network size N . The results, shown in Figure 2, are in very good agreement with our *ansatz*.

The weight of these multiple edges with respect to the overall number of edges is small, a fact that could induce us to conclude that these multiple edges are not important. This is not completely true, however, since these multiple edges are not homogeneously distributed among all the degree classes; i.e. they are present in classes with degree larger than the structural cut-off $k_s \sim N^{1/2}$. This implies that, in the thermodynamic limit, all edges connecting vertices with degrees satisfying condition (17) will always contain a finite fraction of multiple edges and tadpoles. It is also interesting to note that by imposing further constraints to avoid this problem usually generates further correlations. For instance, by imposing the restriction that only one edge may connect a pair of vertices will forbid the natural tendency of high degree vertices to connect mutually and favor their linking with small degree vertices, originating the presence of effective disassortative correlations, as discussed in references [25,29].

5 Conclusions

In the present work we have analyzed the behavior of the maximum degree in networks with finite size N . The natural cut-off scaling usually considered is not always the appropriate one. The constraints imposed by the connectivity structure of networks of finite size generate spontaneous correlations that introduce a structural cut-off that in some regimes is diverging slower than the natural one and then determines the actual scaling of the maximum degree. Strikingly, this phenomenon occurs also in random uncorrelated networks with SF degree distribution with $\gamma < 3$, that are usually used as a first approximation to represent many real networked structures. These results might be particularly relevant in the evaluation of the resilience to damage and the spreading of infective agents in SF networks. Indeed, the absence of intrinsic epidemic and damage thresholds makes these processes dominated by finite size effects. In this case, a careful determination of the cut-off behavior with respect to the network's size is determinant in the calculation of the effective thresholds which determines the behavior of these dynamical processes.

This work has been partially funded by the European Commission – Fet Open Project COSIN IST-2001-33555. R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain) and from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Spain). We also thank an anonymous referee for useful comments.

References

1. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford University Press, Oxford, 2003)
2. R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002)
3. R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001)
4. R. Cohen, K. Erez, D. ben Avraham, S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000)
5. D.S. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000)
6. R.M. May, A.L. Lloyd, *Phys. Rev. E* **64**, 066112 (2001)
7. Y. Moreno, R. Pastor-Satorras, A. Vespignani, *Eur. Phys. J. B* **26**, 521 (2002)
8. M.E.J. Newman, *Phys. Rev. E* **64**, 016128 (2002)
9. M. Boguñá, R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **90**, 028701 (2003)
10. A. Vázquez, Y. Moreno, *Phys. Rev. E* **67**, 0015101 (2003)
11. S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, *Phys. Rev. E* **66**, 016104 (2002)
12. M. Leone, A. Vázquez, A. Vespignani, R. Zecchina, *Eur. Phys. J. B* **28**, 191 (2002)
13. R. Pastor-Satorras, A. Vespignani, *Phys. Rev. E* **65**, 035108 (2002)
14. L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, *Proc. Natl. Acad. Sci. USA* **97**, 11149 (2000)
15. S. Mossa, M. Barthélémy, H.E. Stanley, L.A.N. Amaral, *Phys. Rev. Lett.* **88**, 138701 (2002)
16. A.A. Moreira, J.S. Andrade, L.A.N. Amaral, *Phys. Rev. Lett.* **89**, 268703 (2002)
17. P.L. Krapivsky, S. Redner, *J. Phys. A* **35**, 9517 (2002)
18. W. Aiello, F. Chung, L. Lu, *Experimental Math.* **10**, 53 (2001)
19. S.N. Dorogovtsev, J.F.F. Mendes, *Adv. Phys.* **51**, 1079 (2002)
20. M. Boguñá, R. Pastor-Satorras, A. Vespignani, in *Statistical Mechanics of Complex Networks*, Vol. 625 of *Lecture Notes in Physics*, edited by R. Pastor-Satorras, J.M. Rubí, A. Díaz-Guilera (Springer Verlag, Berlin, 2003)
21. M. Boguñá, R. Pastor-Satorras, *Phys. Rev. E* **68**, 036112 (2003)
22. M.E.J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002)
23. M.E.J. Newman, *Phys. Rev. E* **67**, 026126 (2003)
24. F. Chung, L. Lu, *Annals of Combinatorics* **6**, 125 (2002)
25. J. Park, M. E. J. Newman, *Phys. Rev. E* **66**, 026112 (2003)
26. Z. Burda, Z. Krzywicki, *Phys. Rev. E* **67**, 046118 (2003)
27. A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999)
28. R. Pastor-Satorras, A. Vázquez, A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001)
29. S. Maslov, K. Sneppen, A. Zaliznyak, *Physica A* **333**, 529 (2004)
30. R. Albert, A.-L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000)
31. M. Molloy, B. Reed, *Random Struct. Algorithms* **6**, 161 (1995)
32. M. Molloy, B. Reed, *Combinatorics, Probab. Comput.* **7**, 295 (1998)