Graphicality transitions in scale-free networks

Charo I. Del Genio,[1] Thilo Gross,[1] and Kevin E. Bassler[2, 3]

[1]*Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Straße 38, 01187 Dresden, Germany*
[2]*Department of Physics, 617 Science and Research 1, University of Houston, Houston, Texas 77204-5005, USA*
[3]*Texas Center for Superconductivity, 202 Houston Science Center, University of Houston, Houston, Texas 77204-5002, USA*

(Dated: June 28, 2011)

We study the graphicality of power-law distributed degree sequences, showing that the fraction of graphical sequences undergoes two sharp transitions at the values 0 and 2 of the power-law exponent. We characterize these transitions as first-order, and provide an analytic explanation of their nature. Further numerical calculations, based on extreme value arguments, verify this treatment, and introduce a method to determine transition points for any given degree distribution. Our results reveal a fundamental reason why scale-free networks with no constraints on minimum and maximum degree must be sparse for positive power-law exponents, and dense otherwise.

PACS numbers: 89.75.Hc 89.75.-k 02.10.Ox 89.65.-s

Many complex systems can be modeled as networks, i.e., as a set of connections (edges) between the elements of the system (nodes) [1–3]. A characteristic of a network that affects many physical properties is its degree distribution $P(k)$, which is the probability of finding a node with k edges. In particular, considerable attention has been paid to scale-free networks, in which the distribution of the degrees follows a power-law [4–10]. For studying the properties of scale free networks, several generative models have been proposed (see [1–4] and references therein). However, no models creating scale-free networks with negative power-law exponent less than 2 have been found [11, 12].

It is well known that the mean degree of scale-free distributions with exponents less than 2 diverges in the thermodynamic limit, i.e., when the number of nodes $N \to \infty$ [2]. The divergence of the mean degree implies that such networks are not sparse, but does not preclude their existence as such. Indeed, many examples of non-sparse networks are known [13–15]. In this Letter, we present fundamental mathematical reasons for the non existence of scale-free networks without cutoff when the exponent γ in the power-law distribution of the degrees $P(k) \propto k^{-\gamma}$ lies between 0 and 2. We show that the two points $\gamma = 0$ and $\gamma = 2$ effectively correspond to transition points. We analyze the transitions by studying the behavior of Binder's cumulants for different network sizes, characterizing them as first-order. Then, we provide an analytical argument for the existence of the transitions, and offer a further numerical verification proposing a method for identifying the presence of graphicality transition points for any given degree distribution.

The generation of scale-free networks with a given degree distribution can be considered a two-step procedure. First, one creates a number of nodes and assigns to each node a number of connection "stubs" drawn from the degree distribution. The realization of the degree distribution that is thus created is called *degree sequence.* Second, one connects the stubs such that every stub on a given node links to a stub on a different node, without forming self-loops or double links.

However, not every degree sequence can be realized in a network. Sequences that admit realizations as simple graphs are called *graphical*, and we refer to their realizability property as *graphicality.* Graphicality fails trivially if the number of stubs is odd, as one needs two stubs to form every link, or if the degree of any node is equal to the number of nodes or greater, as it would be impossible to connect all its stubs to different nodes. Moreover, further conditions on the number of stubs of any given subset of nodes must be met for a sequence to be graphical [16–18].

The main result used for testing the graphicality of a sequence is the Erdős-Gallai theorem, which we state here as reformulated in [17] using recurrence relations:

Theorem 1. *Let $\mathcal{D} = \{d_0, d_1, \dots, d_{N-1}\}$ be a non-increasing sequence of integers on N nodes. Also, define $x_k = \min\{i : d_i < k+1\}$ and $k^\star = \min\{i : x_i < i+1\}$. Then, $\mathcal{D}$ is a graphical sequence if and only if $\sum_{i=0}^{N-1} d_i$ is even, and*

$$L_k \leqslant R_k \quad \forall\, 0 \leqslant k < N-1\,, \tag{1}$$

where L_k and R_k are given by the recurrence relations

$$L_0 = d_0 \tag{2}$$

$$L_k = L_{k-1} + d_k \tag{3}$$

and

$$R_0 = N - 1 \tag{4}$$

$$R_k = \begin{cases} R_{k-1} + x_k - 1 & \forall k < k^\star \\ R_{k-1} + 2k - d_k & \forall k \geqslant k^\star \end{cases} . \tag{5}$$

This formulation of the theorem has the advantage over the traditional one [19] of allowing a very fast implementation of a graphicality test [17].

Equivalently, graphicality can be tested by applying the Havel-Hakimi theorem, which provides a recursive constructive test [20, 21]:

Theorem 2. *A non-increasing sequence of integers $\mathcal{D} = \{d_0, d_1, \ldots, d_{N-1}\}$ is graphical if and only if the sequence obtained by connecting the first node with the first d_0 nodes is graphical.*

To investigate the dependence of the graphicality of scale-free networks on the power-law exponent γ, we performed extensive numerics, generating ensembles of sequences of random power-law distributed integers with range between 1 and $N-1$, sequence length varying from $N = 10^2$ nodes to $N = 10^5$ nodes, and γ between -2 and 4. Then, we tested each sequence for graphicality by applying Th. 1, and computed for each γ the graphical fraction g, defined as

$$g = \frac{G}{E} \,,$$

where G is the total number of graphical sequences in the ensemble and E is the number of sequences in the ensemble with an even degree sum. The results, plotted in Fig. 1, clearly show two graphicality transitions: For very large and very small exponents almost all sequences are graphical. However, at intermediate exponents there is a pronounced gap where almost no sequence is graphical. The transitions between the two phases become steeper and the transition points approach $\gamma = 0$, and $\gamma = 2$ as the system size is increased.

The dependence of g on sequence length strongly suggests that both transitions are first-order. To verify their character, we studied Binder's cumulants $U_4 \equiv 1 - \langle g^4 \rangle / \langle g^2 \rangle^2$ [22, 23]. For continuous transitions, the cumulants for different system sizes lie within a finite interval and cross at the critical point, whereas for first-order transitions the curves are flat, except for a diverging negative minimum whose position converges to the transition point with increasing system size [24–26]. In the present system the cumulants confirm that the graphicality transitions at $\gamma = 0$ and $\gamma = 2$ are first order.

To understand the origin of the transitions, we focus on the scaling of the largest degrees and of the number of lowest degree nodes in the sequences. Below, we show that the first two largest degrees are of order $O(N)$ for $\gamma \leqslant 2$, while they grow sublinearly with N for $\gamma > 2$. Also, the number of nodes with degree of order $O(1)$ increases linearly with N for $\gamma \geqslant 0$ and decreases like N^γ for $\gamma < 0$. Then, the transitions can be understood as follows. If we tried to construct a scale-free network with γ between 0 and 2, following the Havel-Hakimi algorithm, $O(N)$ nodes with unitary degree would be used to place the connections involving the first node, and then there would be no way to place all the needed edges involving the node with the second largest degree. Conversely, when $\gamma < 0$, all but a vanishingly small fraction of nodes have a degree of order $O(N)$, and in the thermodynamic limit all the nodes are able to form as many connections as needed.

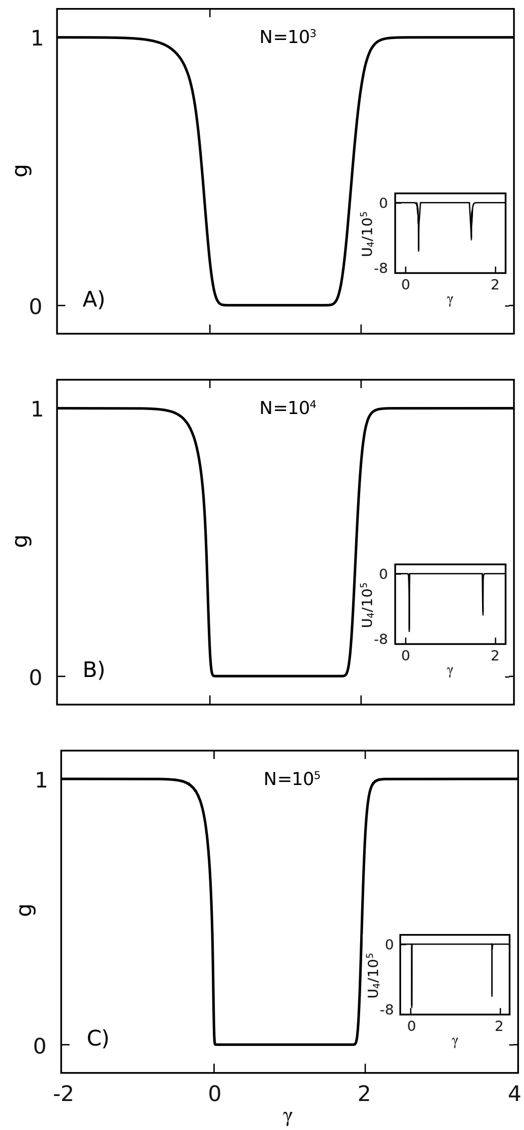

Figure 1: Graphicality transitions in scale-free networks. The plots of graphical fraction g vs. power-law exponent γ show two transitions at $\gamma = 0$ and $\gamma = 2$. Binder's cumulant, shown in the insets, clearly identifies the character of the transitions and confirms the transition points.

To see this, calculate the expected maximum degree of a scale-free sequence $\widehat{d}$ as:

$$\widehat{d} = \max\left\{ x : \; N \sum_{k=x}^{N-1} \frac{k^{-\gamma}}{H_{N-1,\gamma}} \geqslant 1 \right\} \,, \qquad (6)$$

where $H_{a,b}$ is the a^{th} generalized harmonic number of exponent b

$$H_{a,b} = \sum_{t=1}^{a} t^{-b} \,.$$

When $N \gg 1$, Eq. 6 becomes

$$N \int_x^{N-1} \frac{k^{-\gamma}}{H_{N-1,\gamma}} \mathrm{d}k = 1 \, . \tag{7}$$

Because of the dependence of the behavior of the generalized harmonic number on the exponent, we solve this equation for different values of γ.

Solving the integral for $\gamma > 1$ gives

$$\frac{N}{(1-\gamma) H_{N-1,\gamma}} \left[(N-1)^{1-\gamma} - x^{1-\gamma} \right] = 1 \, . \tag{8}$$

Equation 8, for $N \gg 1$ implies

$$x = \left[\frac{N}{(\gamma - 1) H_{N-1,\gamma}} \right]^{\frac{1}{\gamma-1}} \sim N^{\frac{1}{\gamma-1}} \, .$$

Therefore, because of the upper bound of the degrees of a sequence at $N-1$, if $1 < \gamma \leqslant 2$, the value of the largest degree grows linearly with the number of nodes N.

For $\gamma = 1$, the integral in Eq. 7 gives

$$\log(N-1) - \log(x) = \frac{H_{N-1}}{N} \, ,$$

where $H_{N-1} = H_{N-1,1}$ is the $(N-1)^{\mathrm{th}}$ harmonic number. To solve the above equation note that the right-hand side vanishes in the limit of large N. This can be seen by an application of l'Hôpital's rule, noticing that for $\gamma \geqslant 0$

$$\frac{\partial H_{N-1,\gamma}}{\partial N} = \gamma \left[\zeta(\gamma + 1) - H_{N-1,\gamma+1} \right]$$

and

$$\lim_{N \to \infty} H_{N,\gamma} = \zeta(\gamma) \, ,$$

where ζ is Riemann's zeta function. Then, the solution of the equation in the thermodynamic limit is $x \sim N$.

Next, for $0 \leqslant \gamma < 1$, from Eq. 7 yields once more Eq. 8, hence we get

$$(N-1)^{1-\gamma} - x^{1-\gamma} = \frac{(1-\gamma) H_{N-1,\gamma}}{N} \, . \tag{9}$$

As in the previous case, the right hand side vanishes for large N, and the solution is that $x \sim N$.

Finally, for $\gamma < 0$, from Eq. 8 one gets again Eq. 9. However, in this case the right-hand side grows as $N^{-\gamma}$. Since γ is negative, one can rewrite Eq. 9 for large N as

$$(N-1)^{1+|\gamma|} - x^{1+|\gamma|} = N^{|\gamma|} \, ,$$

which implies again that $x \sim N$.

The same arguments can be applied to the scaling of the second largest degree, with identical results. Now, consider the number A of nodes with unitary degree. For large N, $A = N / H_{N-1,\gamma}$. Thus, when $\gamma \geqslant 0$, $A \sim N$, whereas, if $\gamma < 0$, then $A \sim N^{\gamma}$.

Then, to formally check the transition mechanism, explicitly write Inequality 1 for $k = 1$. The left-hand-side consists of the sum of the largest and the second largest degrees, which can be obtained using the same argument as above. To compute the right hand side, first notice that $k^\star$ must be greater than 2. In fact, by definition it cannot be 0, as this would imply that the higiest degree in the sequence would be 0. Moreover, in our case it cannot be 1, as this would imply that the second highest degree in the sequence would be 1, in contradiction with what demonstrated above. Also, by the definition of x_k, it follows that x_1 equals $N - A$. Thus, applying Eq. 5, the right hand side is simply $2N - 2 - A$. Therefore, the inequality reads

$$\left[\frac{(\gamma-1) H_{N-1,\gamma}}{N} + (N-1)^{1-\gamma} \right]^{\frac{1}{1-\gamma}} + \left[\frac{2(\gamma-1) H_{N-1,\gamma}}{N} + (N-1)^{1-\gamma} \right]^{\frac{1}{1-\gamma}} \leqslant 2N - 2 - \frac{N}{H_{N-1,\gamma}} \, .$$

A numerical solution shows that for $N \gg 1$ the above inequality is indeed satisfied only when $\gamma < 0$ or $\gamma > 2$, confirming the transition mechanism.

Note that one can also study the inequality in the presence of a cutoff in the distribution, by replacing every $(N-1)$ with the cutoff value. Cutoffs have been observed in real-world scale-free networks [8, 9], and are sometimes imposed in constructing procedures for different purposes, such as making the degree-degree correlations uniform [27, 28]. Notably, the effect of cutoffs is that the inequality is always satisfied, thus making the transitions disappear. Once again, this confirms the origin of the transitions.

The above treatment indicates that one can build a "representative" finite length sequence $S = \{s_0, s_1, \ldots, s_{N-1}\}$ for the purpose of studying the graphicality of infinite systems, by applying extreme value arguments. A finite sequence maximizing the

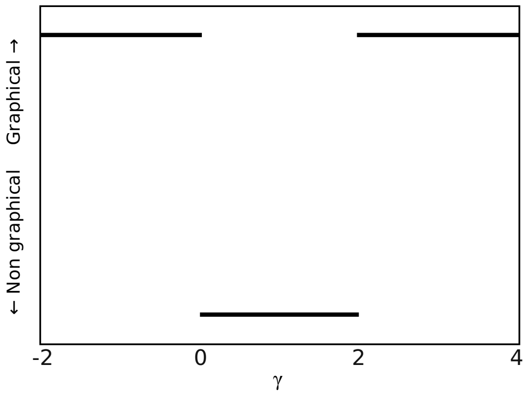

Figure 2: Graphicality of representative finite-size sequences of length $N = 10^6$ for scale-free distributions vs. power-law exponent γ. The graphicality transitions points, at $\gamma = 0$ and $\gamma = 2$ are correctly identified.

degrees of the nodes for a given degree distribution will best approximate the graphicality of an infinite sequence, especially since broken graphicality is always caused by an excess of stubs in some subset of nodes. Therefore, for a length N, and any degree distribution $P(d)$, the elements of the sequence are given by the family of functionals

$$s_i = \max\left\{ s^\star : N \sum_{d=s^\star}^{d_M} P(d) \geqslant i+1 \right\} ,$$

where d_M is the largest allowed degree. In general, $d_M = N-1$, but the full generality of its value allows cutoffs to be accounted for. Increasing the number of nodes in the representative sequence will improve the accuracy in the determination of transition points, as it will better approximate an infinitely large system.

We computed representative sequences with $N = 10^6$ for power-law distributions, and tested them for graphicality. The results, shown in Fig. 2, are consistent with the simulations and with the analytical treatment, showing once again transitions at $\gamma = 0$ and $\gamma = 2$.

In conclusion, we discussed two sharp transitions in the graphicality of scale-free degree sequences, in correspondence of the values 0 and 2 of the power-law exponent γ. We presented numerical evidence by means of extensive simulations, and characterized these transitions as first-order. To explain the transitions, we provided an analytical treatment of the sequences, showing that in the thermodynamic limit it is indeed impossible to build a graph whose degrees are distributed following a power law if the exponent γ lies between 0 and 2. Finally, we used an extreme value argument to offer further verification of our analysis, and introduced a metod to identify graphicality transition points in any given degree distribution. Our results explain the lack of examples and generative models for scale-free networks with exponents between 0 and 2, as these appear to be precluded by fundamental mathematical constraints. This implies that large scale-free networks whose degree distribution is a power-law over the full range $[1, N-1]$ are always sparse when the exponent is greater than 2, and always dense when it is less than 0.

The authors gratefully acknowledge Zoltán Toroczkai, Hyunju Kim and Chiu-Fan Lee for fruitful discussions and helpful comments on the manuscript. KEB was supported by NSF grant No. DMR-0908286.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[3] S. Boccaletti et al. Phys. Rep. **424**, 175 (2006).
[4] G. Caldarelli, *Scale-free networks – complex webs in nature and technology* (Oxford University Press, Oxford, United Kingdom, 2007).
[5] D. J. de Solla Price, Science **149**, 510 (1965).
[6] S. Redner, Eur. J. Phys. B **4**, 131 (1998).
[7] R. Albert, H. Jeong, and A.-L. Barabási, Nature **401**, 130 (1999); H. Jeong et al. Nature **407**, 651 (2000); H. Jeong et al. Nature **411**, 41 (2001).
[8] L. A. N. Amaral et al. Proc. Natl. Acad. Sci. USA **97**, 11149 (2000); F. Liljeros et al. Nature **411**, 907 (2001).
[9] M. E. J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).
[10] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[12] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[13] V. Spirin and L. A. Mirny, Proc. Natl. Acad. Sci. USA **100**, 12123 (2003).
[14] A. H. Y. Ton et al. Science **303**, 808 (2004).
[15] P. Hagmann et al. PLoS Biology **6**, 1479 (2008).
[16] H. Kim et al. J. Phys. A – Math. Theor. **42**, 392001 (2009).
[17] C. I. Del Genio et al. PLoS One **5**, e10012 (2010).
[18] H. Kim et al. *in preparation.*
[19] P. Erdős and T. Gallai, Mat. Lapok **11**, 477 (1960).
[20] V. Havel, Časopis Pěst Mat **80**, 477 (1955).
[21] S. L. Hakimi, J. Soc. Ind. Appl. Math. **10**, 496 (1962).
[22] K. Binder, Z. Phys. B – Cond. Mat. **43**, 119 (1981).
[23] K. Binder, Phys. Rev. Lett. **47**, 693 (1981).
[24] K. Binder and D. P. Landau, Phys. Rev. B **30**, 1477 (1984).
[25] K. Vollmayr et al. Z. Phys. B **91**, 113 (1993).
[26] D. P. Landau and K. Binder, *A guide to Monte Carlo simulations in statistical physics* (Cambridge University Press, Cambridge, United Kingdom, 2000).
[27] S. Mossa et al. Phys. Rev. Lett. **88**, 138701 (2002).
[28] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).