Discrete Scale Invariance in Scale Free Graphs

Matías Graña *

Departamento de Matemática, FCEyN, Universidad de Buenos Aires, Ciudad Universitaria, Pab. 1 (1428) Ciudad de Buenos Aires, Argentina.

Juan Pablo Pinasco

Instituto de Ciencias, Universidad de General Sarmiento, J.M.Gutierrez 1150, (1613) Los Polvorines, Buenos Aires, Argentina.

Abstract

In this work we introduce an energy function in order to study finite scale free graphs generated with different models. The energy distribution has a fractal pattern and presents log periodic oscillations for high energies. This oscillations are related to a discrete scale invariance of certain graphs, that is, there are preferred scaling ratios suggesting a hierarchical distribution of node degrees. On the other hand, small energies correspond to graphs with evenly distributed degrees.

Key words: Scale free graphs, discrete scale invariance, log periodic oscillations
PACS: 89.75; 05.10.-a; 05.10.Ln

1 Introduction

In the last few years a huge amount of research on graphs has been achieved in physics and mathematics, mainly due to the increasing importance of complex networks like the Internet and the World Wide Web, and the role of social networks in the propagation of diseases, infections, rumors and news, in both real and virtual environments such as populations or web-based communities.

Despite the different origins, methods and perspectives in those works, certain concepts appeared and pervaded the network literature, like small worlds, clustering, centrality, scale free degree distribution, and preferential attachment,

* Corresponding author: `matiasg@dm.uba.ar`
Email addresses: `matiasg@dm.uba.ar` (Matías Graña), `jpinasco@ungs.edu.ar` (Juan Pablo Pinasco).

among others, [1], [2], [3]. They suggest that simple underlying mechanisms could describe and explain the network growth and their emerging properties. To this end, several networks based on real data were analyzed (see [4], [5] and the references therein), and many theoretical models of network growth were proposed (see Table 3 in [2]).

Also, there are still many questions and problems which remain unanswered. Perhaps the main one is the origin of the power law degree distribution of nodes in real networks. This distribution seems to be observed in several real networks (although some criticism appeared in the last year, [6], [7], [8]). Roughly speaking, the power law distribution says that the number of nodes in a graph with k links for $k > k_0$ has a decay proportional to $k^{-\alpha}$, for certain $\alpha > 1$. Several theoretical models were proposed to explain this phenomenon. Among them, let us mention *edge redirection* of Krapivsky and Redner [10], *preferential attachment* of Barabasi and Albert [9] (which can be thought of as a particular case of the previous one for certain initial conditions), and *attachment to edges* of Dorogovtsev, Mendes and Samukhin [11].

Although much attention was paid in the theoretical literature in the thermodynamical limit (when time and number of nodes tend to ∞) in these models, not too much is known for graphs with a bound in their size ([11], [12] being the main exceptions). Indeed, many of the computer-based simulations tend to concentrate on large graphs, while small ones usually get ignored. Yet, real networks are finite, and in many cases they are very small (like networks of metabolic reactions, proteins interactions, digital electronic circuits, sexual interactions or food webs). Our intention in this paper is to concentrate on these small graphs. Specifically, we concentrated on graphs with a number of nodes ranging from 50 to 100,000. On the other hand, we were able to generate large amounts of graphs for each model, which reduced the noise and statistical fluctuations, and helped see some otherwise hidden properties.

Another question, which motivates our research, was the discrimination among the different graphs generated from these models and the analysis of their properties. We may consider this question as a basic one concerning the internal topology of the network, and the quality of their connections. Let us note that different networks with the same node distribution respond in distinct ways to edges or nodes removal depending on the connectivity between hubs, that is, if the nodes with high number of links are connected or not between them. This question was considered by Newman in [13], who coined the term *assortativity* in order to describe this phenomenon.

In order to study this question, we introduce the linking energy of a graph G, defined as

$$E_2(G) = \sum_{x \sim y} |d(x) - d(y)|^2,$$

where $d(x)$ is the degree of the node x, i.e., its number of links, and the sum runs over the pairs of connected nodes.

Our original intention was to compute the expected linking energy for certain scale-free network generating models, which enabled us to classify the graphs generated for each models as either assortative or not. One of the first observations, was that the underlying mechanism responsible of networks generation in each model could generate both assortative and non-assortative networks. Another interesting fact is the correlation of the assortativity and the linking energy E_2, and the fractal structure of this correlation. These results will be published in a separated work.

However, we also obtained an unexpected result when we computed E_2 for the models mentioned above, since the distribution of energies presents log-periodic oscillations in the tail. This oscillatory behavior is more apparent and becomes amplified within models which tend to produce a low number of highly connected hubs. The oscillatory behavior is not present for random or regular graphs, even when they are small.

Surprisingly, this shed some light on the first problem, the node distribution. The presence of log-periodic oscillations is associated to fractal complex dimensions, which are due to discrete scale invariance instead of a scale free similarity ([14], [15]). Indeed, this suggests a stronger order in the networks other than that predicted by the scale free models, since fractal complex dimensions are related to (mathematical) self similar fractals such as the triadic Cantor set. Hence, we may classify the graphs generated with any of those models as continuous or lacunary scale free. A similar dichotomy is present for fractals generated with iterated functions systems [16].

We conjecture that the nodes distribution itself must also exhibit the oscillations. To this end we provide some evidence from simulations in the Appendix. At a first sight, this seems to contradict the results in [12], however, this is not the case. Although each node appears with a definite probability following a power law for different models, a particular realization could present a lacunary structure corresponding to a discrete set of invariant scales. Further work is required to settle this question. However, several real networks that appeared in the literature seem to show such oscillations. Among them, we mention the World Trade Web [17], or web-based communities (see Fig. 1, 5 and 8 in [18]). Also, two theoretical models of networks without growth were recently proposed in [19] and [20], where oscillations are apparent. The existence of oscillations on the degree distribution was related to the existence of different hierarchies among the nodes of the graph, see [21], [22] where theoretical models were proposed to study this kind of networks. Hence, our results show that classical models based on preferential attachment or edge redirection can be used to generate non homogeneous networks. The oscillation of the nodes distribution of hierarchical networks was proved recently in [23], and also cellular networks were analyzed, showing the existence of log periodic oscillations which suggest the presence of a discrete hierarchy.

2 Main Results

2.1 The linking energy

Let G be any graph, and let us denote $x \sim y$ whenever there exists a link between the nodes x and y. Let $d(x)$ be the degree of x, that is, the number of links attached to x.

We define the *linking energy* of G as

$$E_2(G) = \sum_{x \sim y} |d(x) - d(y)|^2.$$

This is a discrete version of the energy which arises for the tension of a system of strings attached at the nodes, with heights proportional to the number of connections. For regular networks, it coincides with the variational form of the Dirichlet energy of a discrete Laplacian.

Let us note that the energy spectra for graphs with N nodes is finite. For a tree on N nodes, the maximum value of the energy is attained for the central tree where all the nodes are connected to one of them, with an energy of

$$(N-1)(N-2)^2 \sim N^3.$$

Let us note as well that the energy is an even number:

$$E_2(G) \equiv \sum_{x \sim y} d_x - d_y \equiv \sum_{x \sim y} d_x + d_y = \sum_x d_x^2 \equiv \sum_x d_x = \sum_{x \sim y} 2 \equiv 0 \pmod 2$$

For regular lattices, such as the triangular, square or hexagonal lattice, this energy is proportional to the number of boundary nodes, since the degree is the same for all the internal nodes. Also, a random (Erdős-Renyi) graph has lower energy than a scale free one due to the small number of hubs in it.

For networks with power law degree distribution, this degree sequence is not enough to characterize several topological and dynamical properties. Graphs with the same degree sequence behave differently, depending on whether or not the hubs are connected among them, a problem considered by Newman who introduced the *assortativity* to measure the connectivity between hubs. The linking energy could be used in order to study this problem.

2.2 E_2 in a random (Erdős-Renyi) graph

Let G be a graph in the Erdős-Renyi model, with d nodes. Here, two nodes connect each other with probability p. Let us fix an order for the nodes, say $x_1, \cdots, x_d$, and for $1 \leq i < j \leq d$ let $X_{ij} = \#\{k \sim i,\ k \neq j\}$ and let $Y_{ij} = \#\{k \sim j,\ k \neq i\}$. In other words, X_{ij} is the degree of the node i disregarding a possible connection to node j, and similarly for Y_{ij}. Then, X_{ij} and Y_{ij} are binomial, with probability

$$p(X_{ij} = k) = p(Y_{ij} = k) = \binom{d-2}{k} p^k (1-p)^{d-2-k}.$$

Thus, the mean and the variance are given by

$$< X_{ij} >= p(d-2),$$

$$\sigma^2_{X_{ij}} =< X^2_{ij} > - < X_{ij} >^2 = (d-2)p(1-p).$$

Furthermore, the variables X_{ij} and Y_{ij} are independent. Hence,

$$\begin{aligned} < (X_{ij} - Y_{ij})^2 > &= 2 < X^2_{ij} > -2 < X_{ij} >^2 \\ &= 2(d-2)p(1-p). \end{aligned}$$

Now, the mean value of E_2 is given by

$$\begin{aligned} < E_2 > &= \sum_{i<j} p < (X_{ij} - Y_{ij})^2 > \\ &= \binom{d}{2} p \cdot 2(d-2)p(1-p) \\ &= d(d-1)(d-2)p^2(1-p). \end{aligned}$$

For other models of graphs, the distribution of energies gives better information than the mean value of E_2.

2.3 Models of graph generation

We explain now the studied models. All of them consist of the following steps:

- One begins with a fixed seed of nodes and links.

- At each step, a new node is added to the graph, and
- a fixed number of links emanate from the new node to the existing ones.

These models differ in the way the targets for the new nodes are chosen. We considered the following models:

KR **Edge redirection**: one target node is selected at random with uniform probability. However, with probability $1 - r$, this target is changed by the node it points to.

BA **Preferential attachment** (or *linear* P.A.): one target node is selected at random, with a probability proportional to the degree of the nodes.

DMS **Attach to edges**: one link is selected at random with uniform probability (among the links). The new node connects to both ends of the chosen link.

The labels are put after the authors of these models, see [9], [10], [11]. We consider separately the BA model since its energy distribution is different to the one corresponding to the KR model depending on the initial configuration.

2.4 Numerical Experiments

We computed the energy E_2 for several graphs constructed according to these models. We present in this subsection some of these computations. For the KR model, we used $r = 0.5$, which gives a power law with exponent 3 for the degree distribution, the same exponent of the other two models (see [2]).

We present the results in Figures 1 to 1: in Figure 1 we show E_2 for 10^7 graphs on 100 nodes produced with BA model. In Figure 1 we show E_2 for 10^7 graphs on 100 nodes, produced with KR model, with $r = 0.5$. In Figure 1 we show E_2 for 10^8 graphs on 100 nodes, produced with DMS model. Oscillations in this case become more apparent using logarithmic scale, shown as an inset.

Here, the energy spectra has a maximum close to 100^3, although the observed energy levels are far from this value in the examples. Moreover, there are no forbidden levels at least for $E_2 \leq 40,000$. This fact also shows that the oscillations are not caused by the discreteness nor the sparsity at high values of the energy.

The oscillations persist even varying the width of the bins in the histogram. This could be caused by the fractal pattern in the energy distribution, since each oscillation exhibits a strong self-similarity with a sequence of sub-oscillations, see Figure 2. This fractal pattern is clearly shown in the integrated density of energies,

$$N_2(E) = \int_0^e d\delta(E_2 - E).$$

Fig. 1. Nodes degree distribution of 100-node graphs. Up: Attach to Edges, 10^8 realizations (inset: log scale allows oscillations to become visible). Middle: Linear Preferential Attachment, 10^7 realizations. Down: Edge Redirection, 10^7 realizations.

As was pointed out in [5], the analysis of the cumulative distribution is better than binning in order to avoid statistical fluctuations. In Figure 3 we show different amplifications of $N_2(E)$ for KR with $r = 0.25$. The structure resembles a Cantor devil staircase.

It is interesting to note that the three models shown above share the same node-degree distribution. In all of the cases, degrees decay as a power-law of exponent 3. However, the energy distribution presents some differences. For instance, it is clear that their modes differ: they are close to 4400 (BA), 5700 (KR), and 20000 (DMS). Their oscillations also differ both in the place where they begin and in their amplitudes.

Let us emphasize the difference between KR and BA models. Although KR was proposed as a model to easily simulate BA, both depend on the seed in different ways (for instance, KR needs a seed in which each node has exactly one outgoing edge, while BA can be defined even for undirected graphs). Furthermore, the energy is highly dependent on the seed, and we show in Figure

Fig. 2. Edge redirection, 100 nodes, $r = 0.25$, histogram over 10^8 graphs

Fig. 3. Edge redirection, 200 nodes, $r = 0.1$, 10^7 graphs

4 the energy distribution for 100-node-KR graphs with three different seeds. Thus, the energy can also be used to discriminate between seeds within the same model.

Moreover, for r close to zero, the KR model gives almost only high energy

Fig. 4. Edge redirection, 100 nodes, $r = 0.5$, histogram over 10^7 graphs. Three different seeds were used, showing that KR depends highly on the seed, and in turn that energy depends highly on it.

graphs, together with an interesting behavior of the oscillations, which become highly amplified. In the limit, the resulting distribution is a Dirac delta. We show in Figure 5 the energy distribution for $r = 0.25$.

2.5 Discrete Scale Invariance and log-periodic oscillations

Let us consider a model in which the first node has no outgoing edge, while the others have exactly one, as it happens in KR or BA models. Then, a graph with d nodes has $d-1$ edges, and for each edge $x \sim y$, $|d(x) - d(y)| \leq d-2$. Thus, $E_2(G) \leq E_2^{max} := (d-1)(d-2)^2$. This level of energy is reached with probability 1 in KR when $r = 0$.

Now, consider the graph G_k depicted in Figure 6. It has 1 node at the center, k nodes with degree 2, k nodes "in a second row", attached to the previous k nodes, and $d - 2k - 1$ nodes with degree 1 and at distance 1 from the origin.

The energy of G_k is given by

$$e_k := E_2(G_k) = (d - 2k - 1)(d - k - 2)^2 + k(d - k - 3)^2 + k.$$

If $k \ll d$,

$$e_k \sim (d-k)^3.$$

Fig. 5. Edge redirection, 200 nodes, $r = 0.25$, histogram over 2×10^7 graphs. In the inset, distribution of high-energy graphs.

Fig. 6. Graph with k edges in a second row

Within the KR model, if r is close to 0, this sort of graphs have non-negligible probability, and they explain the peaks at energies close to E_2^{max}, which can be seen at the inset in Figure 5. If one pays a closer attention to one of these peaks, one will see that each of them has sub-peaks with smaller probability; this fact is more apparent when the number of nodes is bigger (see Figure 7, which is made with 2000-node graphs).

The sub-peaks are due to graphs similar to G_k, but for which the k nodes at second row do not attach to those in first row so regularly. The number of energy levels around the peak associated to G_k is related to the number of graphs on k nodes (more precisely, to the number of *forests*, i.e. disjoint union of trees, on k leaves). This means that when k grows, these peaks get wider, and eventually they become close enough to seem oscillations. Such a phenomenon can be seen in Figure 7.

Fig. 7. Sub-peaks become close. KR, 2000 nodes, $r = 0.20$, E_2.

Since, for $k \ll d$, $\ln(e_k) - \ln(e_{k+1}) \sim \frac{3}{d-k} \sim \frac{3}{d}$, the peaks can be considered a log periodic structure. The log periodicity of oscillations was related to the existence of a *discrete scale invariance* (as opposed to a continuous one), which is due to the existence of preferred scaling ratios. For self similar fractals like the triadic Cantor set or the Sierpinski gasket, the existence of *complex* fractal dimensions was proposed, since the imaginary part leads to log periodic corrections of the scaling. In our setting, the existence of preferred scaling ratios corresponds to the existence of different hierarchies in the distribution of node degrees in a given graph. That is, the degrees are not evenly distributed, but lacunary.

We wish to stress that this fact was obtained independently for the degree distribution of hierarchical scale free graphs in [22], and some models of graph generation were proposed there. Here, the energy shows that those kind of networks are obtained also from classical models of graphs generation like the BA, KR, or DMS.

3 Conclusions

In this work we introduced the linking energy E_2, and computed it for several small graphs. We generated the graphs with preferential attachment, edge redirection and attachment to edges models. These models are known as responsible of power law distribution of the nodes degree, and the graphs generated are considered scale free. We also consider random (Erdős-Renyi) and regular lattice graphs.

The frequency distribution of energies shows log periodic oscillations in the tail for the graphs generated with scale free models, which are associated to complex fractal dimensions and discrete scale invariance. They are not present for random or regular graphs.

There are two interpretations of the terms *scale free*. In the first one, the range of the nodes degree distribution crosses several scales, instead of being concentrated around the mean number of connections. The second one suggest that the nodes degree distribution presents details at every interval of the range following a power law.

Our results show the presence of discrete scale invariance in high energies graphs: they are scale free in the first sense, although the nodes degree are not evenly distributed between the minimum and maximum values, but clustered in hierarchies. This is a fact observed in small networks gathered from real data. Also, for small energies, the distribution does not show the oscillations and the graphs seems to be scale free in the second sense. Hence, the models considered are able to generate both kind of graphs, and the energy could be used to discriminate between them.

Acknowledgements

MG is partially supported by Fundación Antorchas and CONICET.
JPP is partially supported by Fundación Antorchas and ANPCyT.

A Some evidence on degree oscillation

Scale free networks may be defined by having a degree distribution $d(k) \propto k^\gamma$, and, among others, the models studied here are shown to produce such networks. However, this is true in the thermodynamical limit. This is also true for finite networks if one takes the average of $d(k)$ along the graphs produced by the model, as proved in [12]. However, we conjecture that the degree distribution of some finite network oscillates. To be precise, we introduce the quantity $o = o(f)$ below as a mean to measure oscillation. We use the variation of degree distributions: let $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be a function which eventually vanishes (i.e., $\exists n_0$ such that $f(n) = 0 \ \forall n \geq n_0$). For such an f, we take its variation

$$V(f) = \sum_{n \in \mathbb{N}} |f(n) - f(n+1)|.$$

# nodes	$r = 0.8$	$r = 0.6$	$r = 0.4$	$r = 0.2$	$r = 0.1$
100	1.873	2.238	2.609	2.909	3.006
200	1.619	2.008	2.423	2.805	2.955
400	1.433	1.808	2.245	2.699	2.900
800	1.302	1.642	2.083	2.594	2.844
1600	1.210	1.506	1.937	2.491	2.787
3200	1.145	1.397	1.808	2.394	2.730
6400	1.099	1.310	1.695	2.300	2.674

Table A.1
Normalized variation $o(d \times n)$ for finite graphs. Averages over 10^6 realizations.

Thus, $V(f)$ provides a measure as to how f oscillates. Notice that if $f(n) \geq f(n+1) \geq 0$ for all $n \in \mathbb{N}$, then $V(f) = f(1)$. Therefore, the quantity $o(f) := \frac{V(f)}{f(1)} \in [1, +\infty]$ is a normalized way to see how far f is from being monotone decreasing ($o(f) = 1$ meaning f is indeed monotone decreasing).

We averaged the quantity $o(f)$ for the functions $f(n) = n \times d(n)$, where $d(n)$ is the number of nodes with degree n in networks constructed by Edge Redirection model. The results are reproduced in table A.1. Notice that since theoretically $d(n) \propto n^{\gamma}$ with $\gamma < -1$, then $f(n)$ should still be monotone decreasing. Taking f instead of d, however, makes oscillations on high degrees more evident.

As one can see from the table, $f(n) = n \times d(n)$ approaches a monotone decreasing function when the number of nodes increases. However, the simulation gives evidence that degrees do oscillate for finite networks.

References

[1] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, *Classes of Behavior of Small-World Networks*, Proc. Natl. Acad. Sci. **97**, 11149 (2000).

[2] R. Albert and A-L. Barabási, *Statistical Mechanics of Complex Networks*, Rev. Mod. Phys. **74**, 47 (2002).

[3] Juyong Park and M. E. J. Newman, *The Statistical Mechanics of Networks*, submitted to Phys. Rev. E; (`cond-mat/0405566`).

[4] S.N. Dorogovtsev and Mendes J F F, *Evolution of Networks*, Advances in Physics **51**, 1079 (2002); (`cond-mat/0106144`).

[5] M. E. J. Newman, *Power laws, Pareto Distributions and Zipf's Law*, Contemporary Physics **46**, 323 (2005); (`cond-mat/0412004`).

[6] D. Achlioptas, A. Clauset, D. Kempe and C. Moore, *On the Bias of Traceroute Sampling; or, Power-law Degree Distributions in Regular Graphs*, Proc. STOC (2005); (cond-mat/0503087).

[7] M.L. Goldstein, S.A. Morris and G.G. Yen, *Problems with Fitting to the Power-law Distribution*, Eur. Phys. J. B **41**, 255 (2004); (cond-mat/0402322).

[8] M. Mitzenmacher, *A Brief History of Generative Models for Power Law and Lognormal Distributions*, Internet Mathematics **1** (2) 226 (2004).

[9] A-L. Barabási and R. Albert, *Emergence of Scaling in Random Networks*, Science, **286**, 509 (1999).

[10] P. L. Krapivsky and S. Redner, *Organization of Growing Random Networks*, Phys. Rev. E **63**, 066123 (2001); (cond-mat/0011094).

[11] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, *Generic Scale of the "Scale-free" Growing Networks*, Phys. Rev. E **63**, 062101 (2001); (cond-mat/0011115).

[12] P. L. Krapivsky and S. Redner, *Finiteness and Fluctuations in Growing Networks*, J. Phys. A **35**, 9517 (2002).

[13] M. E. J. Newman, *Mixing Patterns in Networks*, Phys. Rev. E **67**, 026126 (2003); (cond-mat/0209450).

[14] D. Sornette, A. Johansen, A. Arneodo, J.-F. Muzy and H. Saleur, *Complex Fractal Dimensions Describe the Internal Hierarchical Structure of DLA*, Phys. Rev. Lett. **76**, 251 (1996).

[15] D. Sornette, *Discrete Scale Invariance and Complex Dimensions*, Physics Reports **297**, 239 (1998) (cond-mat/9707012).

[16] Levitin, M. and Vassiliev, D. *Spectral Asymptotics, Renewal Theorem and the Berry Conjecture for a Class of Fractals*, Proc. London Math. Soc. (3) **72** 188 (1996).

[17] M.A. Serrano and M. Boguñá, *Topology of the World Trade Web*, Phys. Rev E **68**, 015101(R) (2003).

[18] W. Bachnik, S. Szymczyk, P. Leszczynski, R. Podsiadlo, E. Rymszewicz, L. Kurylo, D. Makowiec and B. Bykowska, *Quantitive and Sociological Analysis of Blog Networks*, Acta Physica Polonica B **36**, No. 10 2435 (2005); (physics/0506051).

[19] Yan-Bo Xie, Tao Zhou and Bing-Hong Wang, *Scale-free Networks Without Growth*, (cond-mat/0512485).

[20] V.C. Barbosa, R. Donangelo and S.R. Souza, *Emergence of Scale-free Networks from Local Connectivity and Communication Trade-offs*, (cond-mat/0512526).

[21] Barabasi and Ravasz, *Hierarchical Organization in Complex Networks*, Phys. Rev E **67**, 026112 (2003).

[22] K. Suchecki and J.A. Holyst, *Log-Periodic Oscillations in Degree Distributions of Hierarchical Scale-Free Networks*, Acta Physica Polonica B **36** (8), 2499 (2005); (cond-mat/0503291).

[23] W. X. Zhou, Z. Q. Jiang, and D. Sornette, *Exploring self-similarity of complex cellular networks: The edge-covering method with simulated annealing and log-periodic sampling*, Arxiv preprint cond-mat/0605676, (2006).