How to generate a scale-free network from a random graph

João Gama Oliveira

Departamento de Física, Universidade de Aveiro,
Campus Universitário de Santiago, 3810-193, Aveiro

Abstract

A numerical analysis of the distribution of the number of vertices at distance d for Erdős-Rényi (ER) random graphs is presented. We observe that, while for $d < \bar{\ell}$ (average distance between vertices in the graph) the distribution is Poisson like, for the region where $\bar{\ell} \lesssim d \lesssim \ell_{max}$ the distribution becomes power law with exponent -2, a result already obtained analytically in Ref. [4]. This finite size effect allows, as an application, to generate a scale-free (SF) network (starting from an ER random graph) with γ exponent equal to 2. Also, we find that, although the γ exponent is independent of the parameter pN of the starting ER random graph, the behavior of the clustering coefficient as a function of the degree k, as well as correlations between nearest neighbors, in the generated SF network, clearly depend on that parameter. A possible mechanism, which potentially self-organizes random networks to SF correlated networks, is suggested.

Some basic definitions

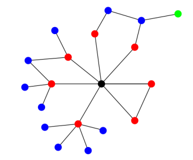
– *Intervertex distance*, ℓ_{ij} : Length of the shortest path between a pair of vertices

– *Nearest neighbors* : A pair of vertices separated by a distance d are d^{th} nearest neighbors

1^{st} neighbors are separated by a distance of one edge, 2^{nd} neighbors by two edges, and so on.

– d^{th} *neighbors' shell of vertex* i : The set of vertices which are d^{th} neighbors from a vertex i

The figure illustrates an example of a small, simple graph with 1^{st} neighbors (red), 2^{nd} neighbors (blue) and a 3^{rd} neighbor (green) from a given vertex (black).



– $P_d(s)$: The probability that, by choosing, uniformly, at random, a vertex i in a graph, then the number of its d^{th} nearest neighbors is s

In particular, if $d = 1$, then $P_1(s)$ is the degree distribution of the graph (where *degree* is the number of edges connected to a given vertex).

Some basic definitions

– *Average distance* between pairs of vertices in a connected component of size n :

$$\bar{\ell} = \frac{2}{n(n-1)} \sum_{i,j>i} \ell_{ij}$$

– *Diameter*, ℓ_{max} : The largest of all distances ℓ_{ij}

– *Remote distance* d : A distance in the interval $\bar{\ell} \lesssim d \lesssim \ell_{max}$

– *Erdős-Rényi random graph* : A graph where the presence or absence of an edge between two vertices is independent of the presence or absence of any other edge, so that each edge may be considered to be present with independent probability p. Its degree, k, follows a distribution

$$p_k = \binom{N}{k} p^k (1-p)^{N-k} \simeq \frac{z^k e^{-z}}{k!}$$

where $z = pN$ is the average degree.

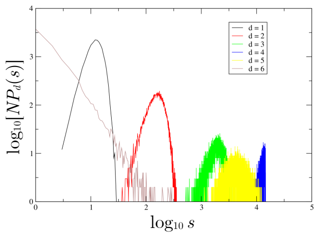
– *Scale-free network* : A net with power-law degree distribution $p_k \sim k^{-\gamma}$

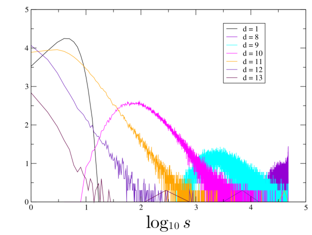
In terms of the notation previously defined, the degree distribution $p_s = P_1(s)$, so that $P_d(s)$ can be seen as a generalization of the concept of degree distribution.

Distribution of the number of vertices at distance d in an Erdős-Rényi random graph – numerical results

$NP_d(s)$ is plotted in the following figure for two ER random graph realizations:

$N = 2{\times}10^4$, $z = 13$	$N = 10^5$, $z = 5$
$\bar{\ell} \sim \log N / \log(pN) \simeq 3.86$	$\bar{\ell} \sim \log N / \log(pN) \simeq 7.15$
$\ell_{max} = 6$	$\ell_{max} = 14$





The slope -2 of the log-log plots when d is the remote distance (6 and 12 for the above realizations), provides evidence of the power law $P_d(s) \sim s^{-2}$.

Generating a scale-free net from an ER random graph

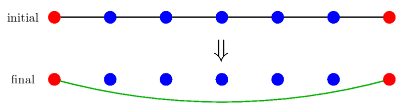
Taking advantage of the power-law distribution of the number of remote d^{th} neighbors in an ER random graph, by shortcutting all shortest paths of such length d in such a graph, it is possible to convert it into a SF network.

Specifically, the method to generate a SF network from an initial ER graph is as follows :
- start from a disconnected set of vertices, labeled in the same manner as in the ER graph
- in this new set of vertices, connect by an edge every pair of vertices which are separated by the remote distance d in the original ER graph

The graph thus generated will have a power-law degree distribution equal to the distribution of the number of d^{th} neighbors in the original ER graph, i.e. $p_k \sim k^{-2}$.

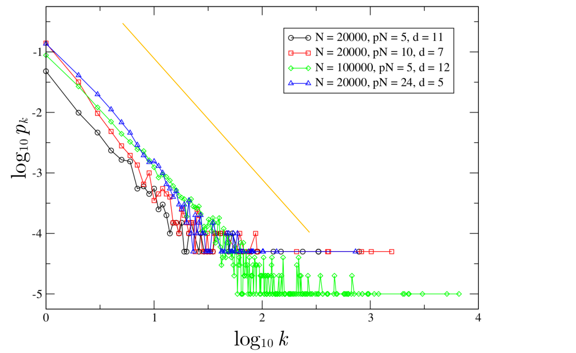
Illustration of the procedure for a simple graph, a chain of length 6 and letting d=6 :

The figure shows the application of the algorithm above for an initial simple case.



Results on generated scale-free networks – degree distribution

p_k of networks generated from 4 ER random graphs with the parameters indicated in the box :

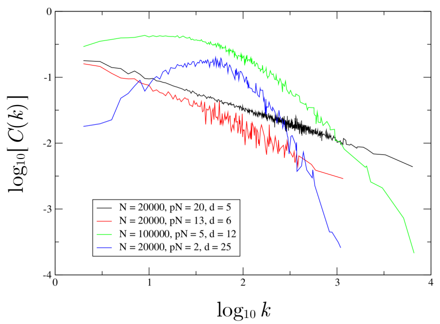


The straight line has slope -2, therefore $p_k \sim k^{-2}$.

Results on generated scale-free networks – clustering coefficient

The *clustering coefficient* of vertex i is defined as $C_i = \dfrac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}$ constituting a measure of the density of triangles in a network.

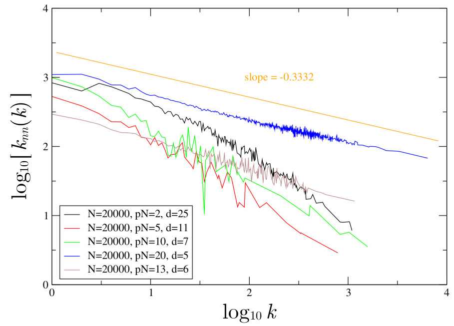
The following figure plots the dependence of C_i on k_i for the SF networks generated from the ER random graphs with the parameters indicated in the box :



Results on generated scale-free networks – degree correlations

A compact way of quantifying degree correlations is by computing the mean degree of neighbors of a vertex as a function of the degree k of that vertex, $k_{nn}(k)$.

The decreasing functions plotted below indicate that the obtained SF networks exhibit *disassortative* degree correlations (typically found in information, technological, and biological networks).



Summary and conclusion

The distribution $P_d(s)$ for ER random graphs was numerically observed to follow power law, s^{-2}.

SF nets, with degree distribution $p_k \sim k^{-2}$, were generated from ER random graphs.

The method for such generation is based on the replacement of long shortest paths by single edges.

For these SF nets, $C(k)$ and $k_{nn}(k)$ were numerically computed, allowing the observation that the presented method generates loopy and disassortative correlated nets; this confirms the presence of large loops and long-range structural correlations in ER random graphs.

A potential real-world example of such contraction of long paths to single edges could be found in the use signal repeaters in communication networks, namely, in situations when a number of repeaters could be replaced by a single (more efficient) repeater.

Other mechanisms, e.g. whenever minimization of shortest path distances is at play, could potentially serve as examples of random networks' organization to scale-free, correlated, networks.

References

[1] B. Bollobás, Random Graphs (Cambridge University Press, second edition, 2001).
[2] M. E. J. Newman, S. H. Strogatz, D. J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E **64**, 026118 (2001).
[3] A.-L. Barabási, R. Albert, Emergence of Scaling in Random Networks, Science **286**, 509 (1999).
[4] S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin, Metric structure of random networks, Nucl. Phys. B **653**, 307 (2003).
[5] R. Albert, A.-L. Barabási, Statistical Mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).
[6] S. N. Dorogovtsev, J. F. F. Mendes, Evolution of networks, Adv. Phys. **51**, 1079 (2002).
[7] M. E. J. Newman, The structure and function of complex networks, SIAM Review **45**, 167 (2003).