On Growing Perfect Power-Law Graphs

Boleslaw Szymanski

Network Science and Technology Center and Department of Computer Science
Rensselaer Polytechnic Institute, Troy, NY 12110, USA
szymab@rpi.edu

Abstract

We outline here two new algorithms for growing perfect power-law graphs of increasing size for the given set of parameters of the desired graph, such as the average and maximum node degrees as well as desired exponent of the power law distribution of the node degrees. We present also relations of these parameters to each other and constraint of their feasible values. Both algorithms are distributed, therefore they requires just one broadcast for joining each newly added node and limited response from the existing nodes to interconnect new edges to the existing nodes. In both cases only nodes with the currently desired node degree respond to the broadcast. In making a choice of the node to which the new edge is added, additional criteria for node selection, such as distance to the newly added node, can be used. The first presented algorithm, SRA (Semi-Random Algorithm), select the currently desired node degree according to the probabilities defined by the power law distribution of node degrees. The second algorithm, SDA (Semi-Deterministic Algorithm), selects the currently desired node degree to minimize the sum of absolute differences between current and desired node degrees.

1. The Desired Properties of the Algorithm

The goal of this note is to derive an algorithm for constructing a graph with a power-law distribution of node degrees. Unlike previous solution, see for example [1, 2], we aim at having power law distribution of node degrees at every stage of construction and for the given set of power-law and graph parameters, such as exponent of the power-law, the minimum, average and maximum node degree in the graph. The construction starts with a fully connected graph with just $n_0=2k+1$ nodes, where $2k$ denotes the desired average node degree and the process continues until the graph with the desired number of nodes n_{max} is constructed. In each step of the construction, we add a new node with the number of edges equal to the half of desired average node degree $2k$. These edges are then connected to the nodes of the graph created in the previous step. We will characterize such a set of graph by the following parameters:

n – the current number of nodes in the graph, in our analysis n is not a constant, like other parameters, but rather an instance of the value, and can assume any positive value between k and n_{max}, , but we are interested in cases when n_{max}>>k, and of course $n>2k$;

$m>2k$ – the maximum degree of any node, so nodes with degree m stop accepting any new connections; below we analyze the simplest case of m being integer, so all edges have at most degree of m (we excluded the trivial of k=m/2 which will result in the graph will all nodes of degree 2k, so trivially satisfying definition of power law distribution of node degrees, there is also clear that the maximum degree of any node must be less than the number of nodes, so naturally $m<n$, and in general $n<n_{max}$).

k – as mentioned above, it is the half of average node degree, so each time a new node is added to the graph, on average *k* edges are also added; below we analyze the simplest case of *k* being integer and the growth being made with a constant number of edges added, so with each new node, exactly *k* new edges are added; it is a matter of simple extension to have instead a vector $[k_i]$ of expected frequencies (defined of course by the desired power-law distribution of node degrees) with which *i* edges are added with the newly added node, and then $k = \sum_{i=1}^{m} ik_i$;

$\gamma>0$ - the exponent of the desired power-law distribution of node degree, that defined the fraction of nodes of degree *d* as being proportional to $d^{-\gamma}$(we consider below the values of this exponent larger than 1, following the traditional assumptions, however, for certain values of *m* and *k*, this is not necessary).

It is interesting to note that the three constant parameters: *k, m,* γ , are independent of each other except that for certain values of *m*, and *k,* when there is lower bound for values of γ that that could be larger than 1.

Let n_i denotes the number of nodes with degree *i* in the graph with *n* nodes. By simple enumeration of all nodes we have:

(1) $$n = \sum_{i=k}^{m} n_i .$$

By simple enumeration of all edges (by construction, there are exactly *kn* edges in the graph with *n* edges) and taking into account that each edge belongs to two nodes, we have:

(2) $$2kn = \sum_{i=k}^{m} in_i .$$

From Eq. (1) we have $2kn = \sum_{i=k}^{m} 2kn_i$ hence, removing last element from the sum of Eq. (2) we get

(3) $$n_m = \frac{1}{m-2k} \sum_{i=k}^{m-1} (2k-i)n_i .$$

Finally, the required power-law distribution of node degrees yields the equation

(4) $$n_i = \frac{cn}{i^{\gamma}}$$ for $i<m$.

Note that we cannot enforce the power-law distribution for the nodes with maximum degree *m* because their frequency is defined by Eq. (3). Only for certain values of *k, m, n* and γ, this equation will yield $n_m = \frac{cn}{m^\gamma}$.

Using Eq. (4) to substitute in the right hand side of Eq. (3), we get

(5) $$n_m = \frac{cn}{m-2k}\sum_{i=k}^{m-1}\frac{2k-i}{i^\gamma}.$$

Clearly, the nodes with maximum node degree will also have frequency defined by the power-law if γ is selected such that

Using Eq. (4) to substitute in the right hand side of Eq. (3), we get

(6) $$\frac{m-2k}{m^\gamma} = \sum_{i=k}^{m-1}\frac{2k-i}{i^\gamma}.$$

Since *m>2k,* than the left hand side of Eq. (6) is always positive, its derivative for γ is $-ln(m)(m-2k)/m^\gamma$ while its value approaches $(1-2k/m)m^{-\gamma+1}$ when γ tends to infinity. The right hand side of this inequality can be initially negative, but for large γ it must be positive, its value approaches $k^{-\gamma+1}$ when γ tends to infinity and it has the derivative $-\sum_{i=k}^{m-1}\ln(i)\frac{2k-i}{i^\gamma}$.

It is easy to show then, that the right hand side decreases slower than the left hand side and therefore at most one unique value of γ can satisfy Eq. (6). The condition for the unique solution to exits is that for $\gamma=0$, the right hand side is smaller than the left hand side, hence $m-2k \geq 2k(m-k)+(m-1)m/2+k(k+1)/2$ which reduces to $(m-2k)^2 \geq k^2+3k$ and since $m>2k$ and $(k+1)2<k^2+3k<(k+1.5)^2$.

Using Eq. (4) to substitute in the right hand side of Eq. (3), we get

(7) $$m \geq 3k+2.$$

In short, for *m* greater or equal to 3*k+2,* there exists a unique value of γ for which the constructed graph will have power-law distribution for all node degrees.

Now, we can use Eq. (1) to compute the constant c from Eq. (4)

$$n = \sum_{i=k}^{m-1}\frac{cn}{i^\gamma} + \frac{cn}{m-2k}\sum_{i=k}^{m-1}\frac{2k-i}{i^\gamma} = \frac{cn}{m-2k}\sum_{i=k}^{m-1}\frac{m-i}{i^\gamma}$$ so

(8) $$c = \frac{m-2k}{\sum_{i=k}^{m-1}\frac{m-i}{i^\gamma}}.$$

To be independent of the graph size *n*, we will use frequency $f_i = \frac{n_i}{n}$ of nodes with degree *i*, and substituting c using Eq. (8), we get from Eq. (4)

(9) $$f_i = \frac{m-2k}{i^{\gamma} \sum_{j=k}^{m-1} \frac{m-j}{j^{\gamma}}}$$ for *i<m.*

Using Eq. (5) and c substitution from Eq. (8) we get also

(10) $$f_m = \frac{1}{\sum_{j=k}^{m-1} \frac{m-j}{j^{\gamma}}} \sum_{i=k}^{m-1} \frac{2k-i}{i^{\gamma}}$$

It is easy to check, using Eq. (9) and Eq. (10) that, as needed, we have

$$\sum_{i=k}^{m} f_i = \sum_{i=k}^{m-1} \frac{m-2k}{i^{\gamma} \sum_{j=1}^{m-1} \frac{m-j}{j^{\gamma}}} + \frac{1}{\sum_{j=k}^{m-1} \frac{m-j}{j^{\gamma}}} \sum_{i=k}^{m-1} \frac{2k-i}{i^{\gamma}} = 1$$

so we can use this equality to replace Eq. (10) with a simpler one:

(11) $$f_m = 1 - \sum_{i=k}^{m-1} f_i$$

Eq. (9) and Eq. (11) express frequencies *f's* as simple functions of *m*, *k* and γ.

Let's consider now a growth of the graph from its size *n* nodes to the size of *n+1* nodes. The added node has on average *k* edges originating at it which are then connected to the existing nodes, so on average it increases by 1 the number of nodes with degree *k*, i.e. $n'_k = n_k + 1$.

Let a_i denote the average number of nodes that increase their degree from *i* to *i+1* in one step of growth (hence, decreasing the number of nodes of degree *i* by a_i and increasing the number of nodes with degree *i+1* by a_i) by connecting to a newly added node. Of course, each existing node can add at most one connection to a newly added node. Hence, we have $f_k(n+1) = f_k n + 1 - a_k$, so $f_k = 1 - a_k$ and finally

(12) $a_k = 1 - f_k$.

Similarly, $f_i = a_{i-1} - a_i$ for $k<i<m$ and therefore by induction

(13) $$a_i = 1 - \sum_{j=k}^{i} f_j$$ for $k \le i < m$

Finally, $f_m = a_{m-1}$ so

(14) $a_{m-1} = f_m$.

Since we are adding k edges in each step of growth, then $\sum_{i=k}^{m-1} a_i = k$. However

$$\sum_{i=k}^{m-1} a_i = m - k - \sum_{i=k}^{m-1} (m-i) f_i$$

so the required constraint can be rewritten as $m - 2k = (m-2k) \sum_{i=k}^{m-1} \frac{m-i}{i^{\gamma} \sum_{j=k}^{m-1} \frac{m-j}{j^{\gamma}}}$

Clearly, the sum on the right hand side sums up to one, so this restriction is always satisfied.

Finally, another restriction is that all frequencies must be positive. From Eq. (9) and Eq. (10), it is clear that because m>2k, it is necessary and sufficient that $\sum_{i=k}^{m-1} \frac{2k-i}{i^{\gamma}} \geq 0$, which can be rewritten as

(15) $2k \sum_{i=k}^{m-1} i^{-\gamma} \geq \sum_{i=k}^{m-1} i^{-\gamma+1}$.

If this condition is not satisfied, it is always sufficient either to appropriately increase γ or to sufficiently decrease *m*. Other changes to these parameters or parameter *k* may or may not, depending on the particular values of the parameters, also cause the inequality of Eq. (15) to be satisfied. For sufficiently large γ any value of *m* will be feasible. Indeed, we have

$$\sum_{i=a}^{\infty} \frac{1}{i^{\gamma}} = \sum_{j=0}^{\infty} \sum_{i=a2^j}^{a2^{j+1}-1} \frac{1}{i^{\gamma}} < \sum_{j=0}^{\infty} \frac{1}{(a2^j)^{\gamma-1}} = \frac{2^{\gamma}}{(2^{\gamma}-1)a^{\gamma-1}}.$$

Rewriting condition in Eq. (15), we get $\sum_{i=k}^{2k} \frac{2k-i}{i^{\gamma}} \geq \sum_{i=2k+1}^{\infty} \frac{i-2k}{i^{\gamma}} = \sum_{j=1}^{\infty} \frac{j}{(2k+j)^{\gamma}} = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \frac{1}{(2k+i)^{\gamma}}$.

So finally

(16) $\sum_{i=k}^{2k} \frac{2k-i}{i^{\gamma}} \geq \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \frac{1}{(2k+i)^{\gamma}}$.

Thus, for γ*=3,* we have for the right hand side of inequality in Eq. (16):

$$\sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \frac{1}{(2k+i)^3} < \sum_{j=1}^{\infty} \frac{4}{3(2k+j)^2} < \frac{8}{3(2k+1)}.$$

Hence, in this case for at least *k=1*, any value of *m* would be feasible (and there is a numerical evidence that this holds for any *k*). Similar reasoning for γ*=4* shows that for at least *k=1,* and *2* again any value of *m* is feasible.

2. Algorithms

2.1. Random Interconnection

The algorithm starts with the initial configuration of fully interconnected graph of 2*k+1* nodes, and then, in each subsequent step, the newly joining node generates *k* random numbers, r_1 … r_k. Let $v_i = \sum_{l=k}^{i} \frac{a_l}{2}$ and $v_{k-1}=0$. The newly added *i*-th edge will be connected to node of such degree *j* that $v_{j-1} \le r_i < v_j$. All *k* desired node degrees are broadcast in one message. The nodes of the desired degrees respond in time proportional to their fitness to make a connection. If no node responds (because either there are no nodes of the given degree, or all nodes of that degree already are connected to a new node), nodes with node degree by one lower respond, and so on.

The advantage of this algorithm is that it creates different graphs in different runs. However, random selection of the node degree to which the new edge is connected may occasionally create fairly large (few percent) differences between the desired and actual number of nodes of the give degree.

2.2. Selecting Interconnections to Increase Power-Law Distribution Adherence

To decrease random divergence from the power law, the algorithm sketched here uses this divergence in the currently created graph to select the connections of the newly added edged. Namely, for each edge, the algorithm computes which node degree has the node count most divergent from the desired one if a node with this node degree will not be selected for connection to the edge being currently added. This computation is done at each existing node of the current graph in response to a broadcast of the request for connection to newly created edges by the node being currently added. In a sequence of *k* responses, those existing nodes for which the node degree count would diverge the most from the number required by the power-law send prioritized responses (each response is proportional to the fitness of the responding node).

This algorithm is deterministic and reduces the differences between the desired and actual number of nodes of the give degree by the order of magnitude compared to the previously described randomize algorithm.

3. Examples

Two examples of computations of frequencies for the described algorithms are shown below.
The first example assumes m=5 and uses three different values of γ= 2, 1, and 0.

m/γ	**a4**	**a3**	**a2**	**f5**	**f4**	**f3**	**f2**
5/2	0.5906	0.6510	0.7584	0.5906	0.0604	0.1074	0.2416
5/1	0.5517	0.6552	0.7931	0.5517	0.1034	0.1379	0.2069
5/0	0.5	0.6667	0.8333	0.5	0.1667	0.1667	0.1667

The second example assumes m=10 and uses several values of γ ranging from 1.34 to 6. The values of *a'* coefficients are shown first followed by the corresponding values of frequencies *f*.

	a9	a8	a7	a6	a5	a4	a3	a2
10/2.12	0.1055	0.1234	0.1464	0.1769	0.2193	0.2815	0.3815	0.5655
10/2	0.0925	0.1133	0.1395	0.1738	0.2205	0.2878	0.3929	0.5797
10/2.12	0.1055	0.1234	0.1464	0.1769	0.2193	0.2815	0.3815	0.5655
10/1.35	0.0005	0.0449	0.0969	0.1592	0.2359	0.3340	0.4666	0.6620
10/1.34	-0.001	0.0437	0.0962	0.1590	0.2363	0.3349	0.4678	0.6634
10/3	0.1728	0.1786	0.1868	0.1991	0.2186	0.2522	0.3180	0.4739
10/4	0.2107	0.2121	0.2145	0.2185	0.2259	0.2414	0.2790	0.3979
10/5	0.2287	0.2291	0.2297	0.2310	0.2336	0.2403	0.2608	0.3468
10/6	0.2378	0.2379	0.2381	0.2384	0.2394	0.2422	0.2529	0.3132

f10	f9	f8	f7	f6	f5	f4	f3	f2
0.1055	0.0179	0.0230	0.0305	0.0423	0.0623	0.1000	0.1840	0.4345
0.0925	0.0208	0.0263	0.0343	0.0467	0.0673	0.1051	0.1868	0.4203
0.1055	0.0179	0.0230	0.0305	0.0423	0.0623	0.1000	0.1840	0.4345
0.0005	0.0444	0.0520	0.0623	0.0767	0.0981	0.1326	0.1955	0.3380
-0.0012	0.0449	0.0525	0.0628	0.0772	0.0986	0.1330	0.1955	0.3366
0.1728	0.0058	0.0082	0.0123	0.0195	0.0337	0.0658	0.1559	0.5261
0.2107	0.0015	0.0024	0.0040	0.0074	0.0154	0.0376	0.1189	0.6021
0.2287	0.0004	0.0006	0.0012	0.0027	0.0067	0.0204	0.0860	0.6532
0.2378	0.0001	0.0002	0.0004	0.0009	0.0028	0.0107	0.0603	0.6868

Acknowledgement

The author acknowledges helpful discussions of this problem with Dr. Eyuphan Bulut of RPI. This research was sponsored by the US Army Research Laboratory and the UK Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001 and under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defense, or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

Bibliography

[1] Dangalchev, Ch. (2004). "Generation models for scale-free networks". *Physica A* **338.**

[2] Dorogovtsev, Mendes, J.F.F. , Samukhin, A.N. (2000). "Structure of Growing Networks: Exact Solution of the Barabási—Albert's Model". *Phys. Rev. Lett.* **85**(21): 4633.