A Grid-based Model for Generating Scale-Free Networks

Amit Kumar Verma
Department of Computer Science and Engineering
Indian Institute of Technology Ropar
Rupnagar, Punjab
2016csz0003@iitrpr.ac.in

Nikhil Prakash
Department of Telecommunication Engineering
R V College of Engineering
Bengaluru, Karnataka
nikhil07prakash@gmail.com

***Abstract*—It has been observed that the evolution of complex networks such as social networks is not a random process, there exist some key features which are responsible for their evolution. One such feature is the degree distribution of these networks which follow the power law i.e. $P(k) \propto k^{-\gamma}$ where γ is a parameter whose value is typically in the range $2 < \gamma < 3$ and such networks are called scale-free networks [4]. In this paper, we formulate a model for generating scale-free networks based on Barabási-Albert model [6], using insights from elementary Euclidean Geometry that takes into account the geometrical location of the nodes instead of their degrees for new connections. We show that our model generates scale-free networks experimentally and provide a mathematical proof for the correctness of the fact that the degree distribution in generated networks indeed follows the power law. We also validate our model on Erdös collaboration network of mathematicians.**

***Index Terms*—Scale-free networks, Power law, Degree distribution, Erdös collaboration network**

1. Introduction

We witness different types of social networks around us which influence our lives. The dynamics of these social networks change with time as nodes and connections get modified. Understanding how these social networks evolve has interested researchers in recent years, mainly due to the massive growth of online social networks [2] [3]. It has been observed that the process of evolution of these social networks is not a random process, there are some key features which are responsible for the evolution process.

One of the key features of the social networks is the degree distribution which follows the power law i.e. $P(k) \propto k^{-\gamma}$ where γ is a parameter whose value is typically in the range $2 < \gamma < 3$ and hence are called *scale-free networks* [4]. Figure 1.1 shows the degree distribution of a network following power law on the log-log plot. The power law degree distribution explains the formation of hubs and small-world phenomenon in social networks.

Various models have been proposed by researchers to explain the formation of scale-free networks [1]. But the most widely known model for generating scale-free networks is the Barabási-Albert model using *preferential attachment* mechanism [6]. In this paper, we formulate a model for generating scale-free networks based on the Barabási-Albert model that takes into account the geometrical location of the nodes instead of their degrees for new connections. We show that our model generates scale-free networks experimentally and provide a mathematical proof for the correctness of the fact that the degree distribution in generated networks indeed follows the power law. We also validate our model on Erdös collaboration network of mathematicians.

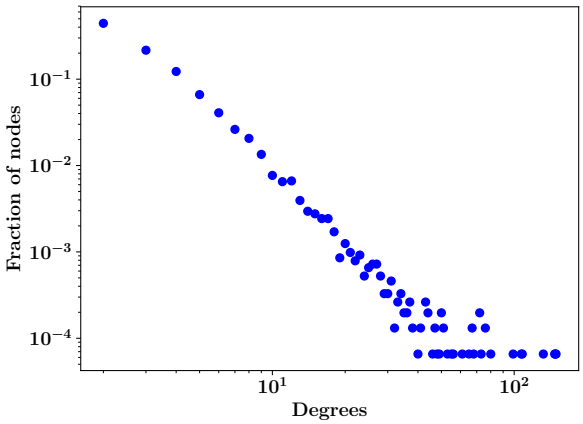


Fig. 1.1: Degree distribution of a network following power law on the log-log plot.

2. Related Work

Derek de Solla Price [5] showed that the number of citations of papers follows a heavy-tailed degree distribution. He named the reason behind the heavy-tailed degree distribution as *cumulative advantage*, which is today more commonly known as preferential attachment. Later this property of heavy-tailed distribution was replaced with power law distribution, and the networks which followed it were called scale-free networks. This term was introduced in the Barabási-Albert model [4]. They proposed a model to show the formation of scale-free networks. With an initial network of m_0 nodes, at each iteration, a new node comes and gets connected to exactly m previously present nodes. The probability by which node i is chosen for connection is given by $p_i = \frac{k_i}{\sum_j k_j}$ where k_i is the degree of node i and the denominator defines the sum of degrees of all preexisting nodes [6]. Random graph

978-1-5386-7902-9/19/$31.00 ©2019 IEEE

model, proposed by Erdős and Rényi, shows the formation of random graphs [7]. The graph is generated by initially having m nodes and the edges are added between two nodes with a probability p. In random graphs, the degree of each node lies around the average degree of the graph. But in the case of scale-free networks, if a node is chosen randomly from the network, then its degree can be very small or very large i.e. it does not have a meaningful internal scale hence they are scale-free [8]. Qinghua Chen and Dinghua Shi proposed an extension to Barabási-Albert model. They proposed two models named A and B. In model A, at each time step a new node gets connected with m existing nodes and new links are also added within the existing network. In model B, the same evolution rules as in Barabási-Albert model was used with the addition that some old links were deleted with anti-preferential probability. They showed that both the models generate scale-free networks [9]. S. S. Manna and P. Sen analyzed how the scale-free networks defined on the Euclidean space behaves when the probability in Barabási-Albert model is added with a length parameter l^{α} [10]. Artikov et al. suggested an approach for generating scale-free networks based on factorization methods and geographical threshold models in which two nodes are connected if they are spatially close and/or have large weights [11].

3. The Proposed Model

A. Preliminaries

We represent the network as an undirected graph G(V, E) where nodes(V) represent the network members and edges(E) represent the connections among them. In our model, we add one node at every iteration. Initially, we say that m_0 nodes are present in the network and at each time step a new node is coming and getting connected to m previously present nodes.

We have assumed that the nodes are distributed on a 2-D plane. The 2-D plane in our case is an nxn grid. The nodes are distributed as points on this grid. Initially, we start with nxn grid on which we distribute the nodes. The nodes are distributed in such a way that less number of nodes are present around the center and more number of nodes are present at the corners of the grid. The probability of placement of these nodes is given as.

$$p(\textit{node } i) = \frac{1}{\sqrt{1+|rad - r_i|}} \tag{3.1}$$

where rad is the radius of the circle inscribed in the grid and r_{i} is the Euclidean distance of $node_{\mathrm{i}}$ from the center. The intuition behind this is to make the distribution in such a way that the probability of placing a node increases as we move away from the center. The function $\sqrt{1+|rad-r|}$ decreases as r increases up-to a point and then increases as rad is a constant. Hence as we move away from the center the probability of placing a node increases. Figure 3.1 shows the placement of nodes on the grid.

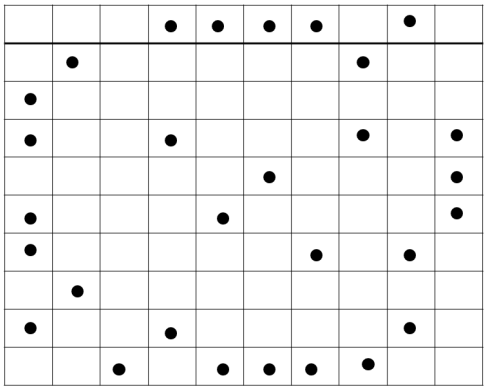


Fig. 3.1: Placement of nodes on the grid

B. Process of evolution

In our model, initially m_0 nodes are present on the grid. We start from the center of the grid and traverse in spiral order towards the corners. At each time step t, a new node is added and connected with m previously present nodes. Lemma 3.1 shows the relation between the j^{th} node in the spiral traversal and its Euclidean distance from the center.

Lemma 3.1. *Let $s(j)$ be the j^{th} node in the spiral traversal with coordinates as $x(j)$ and $y(j)$ and r be the Euclidean distance from the center of the grid then $r = \sqrt{x(j)^2 + y(j)^2}$.*

Proof. In the grid, starting from the center we move in the spiral order and whenever a new node arrives, it is connected to m previously present nodes. Figure 3.2 represents the spiral traversal. Let j be a positive integer representing the spiral

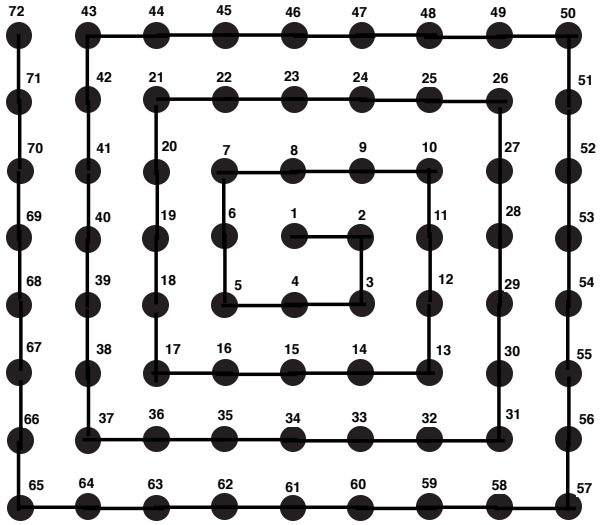


Fig. 3.2: Spiral Traversal

traversal order starting at j = 1 from the center of the grid, which is our origin, i.e. $(0, 0)$. The traversal pattern is.

$$\underbrace{RDLLUU}_{1^{\text{st}}\text{ block}} + \underbrace{RRRDDDLLLLUUUU}_{2^{\text{nd}}\text{ block}} + \ldots.$$

where R, D, L, U represents right, down, left and up respectively. Let us call each $RDLU$ group a block, as shown in above pattern. The first block is the intial $RDLLUU$, second block is $RRRDDDLLLLUUUU$ and so on. We can see that the k^{th} block consist of the following steps.

$$R^{2k\text{-}1}D^{2k\text{-}1}L^{2k}U^{2k} \tag{3.2}$$

The net effect of the above steps, i.e. a single block, is to move one step up and left on the grid, which can also be verified from the figure 3.2. Since our starting position is $(0, 0)$, after the traversal of k blocks, we would have moved to $(-k, k)$ on the grid. We can also deduce from equation 3.2 that the total number of steps traversed after the completion of first k blocks is given by.

$$T = 2k(2k + 1) \tag{3.3}$$

Given the spiral traversal order j, we can find the block in which the corresponding node lies using the following algorithm.

Algorithm 1 Algorithm to find the block

$m \leftarrow \lfloor \sqrt{j} \rfloor$
if m is odd **then**
  $block = \frac{1}{2}(m-1)+1$
else
  if $j > m(m+1)$ **then**
    $block = \frac{m}{2}+1$
  else
    $block = \frac{m}{2}$
  end if
end if

Let us assume that the j^{th} node in the spiral traversal is in the $(k+1)^{\text{th}}$ block. That means, it has already traversed first k blocks. Hence we can say that, the x and y coordinates of j^{th} node is $j - (2k(2k+1))$ steps ahead of $(-k, k)$. Using this fact, we can compute $x(j)$ and $y(j)$ coordinates of the j^{th} node in spiral traversal as.

$$x(j) = \begin{cases} j - 4k^2 - 3k - 1 & \text{if } 2k(k+1) < j \leq (2k+1)^2 \\ k+1 & \text{if } (2k+1)^2 < j \leq 2(k+1)(2k+1) \\ 4k^2 + 7k + 4 - j & \text{if } 2(k+1)(2k+1) < j \leq 4(k+1)^2 \\ -k-1 & \text{if } 4(k+1)^2 < j \leq 2(k+1)(2k+3) \end{cases}$$

$$y(j) = \begin{cases} k & \text{if } 2k(k+1) < j \leq (2k+1)^2 \\ 4k^2 + 5k + 2 - j & \text{if } (2k+1)^2 < j \leq 2(k+1)(2k+1) \\ -k-1 & \text{if } 2(k+1)(2k+1) < j \leq 4(k+1)^2 \\ j - 4k^2 - 9k - 6 & \text{if } 4(k+1)^2 < j \leq 2(k+1)(2k+3) \end{cases}$$

Hence after finding the coordinates $x(j)$ and $y(j)$, we can represent the Euclidean distance r in terms of j as.

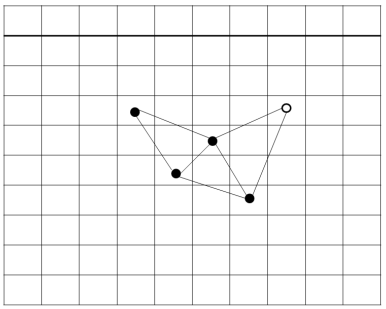
$$r = \sqrt{x(j)^2 + y(j)^2} \tag{3.4}$$

□

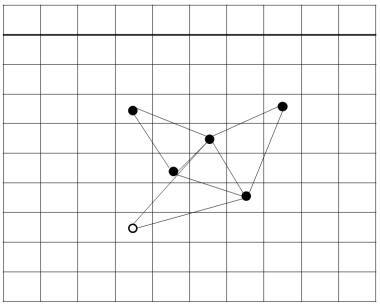
Each new node is getting connected to m previously present nodes with probability s.t its likelihood of getting connected to nodes which are closer to center of the grid is more as compared to other nodes. In this way we try to ensure the preferential attachment based on the geometrical location in our model. Hence at time t the probability that the new node chooses some previously present node i is given by.

$$\Pi(\textit{node } i) = \frac{\frac{1}{r_i}}{\sum_{j=1}^{t-1} \frac{1}{r_j}} \tag{3.5}$$

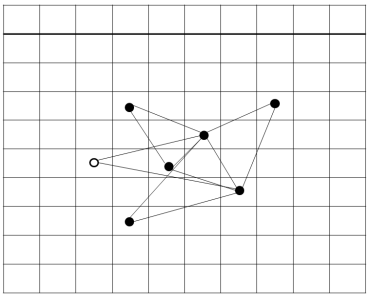
where r_i represents the Euclidean distance of node i from the center of the grid and the denominator defines the sum of Euclidean distance of all preexisting nodes. So, each time a new node comes, m nodes are chosen with the above probability for the connections. Figure 3.3 shows the evolution of network where $m = 2$ connections are made with the new incoming node at each time step.



(a)



(b)



(c)

Fig. 3.3: Evolution of network. White circles are the newly added nodes which are getting connected with $m = 2$ previously present nodes.

4. Experiments and results

For the experimental purpose, we made a tool replicating our model in Python language. In each experiment, we generate a grid of size nxn and distribute the nodes according to the equation 3.1. The spiral traversal starts from the center of the grid and moves towards the corners. At each time step, a new node is connected to m previously present nodes according to equation 3.5. Our result shows that the generated networks are scale-free as their degree distribution on the log-log plot is linear. Figure 4.1 shows the degree distribution of a network generated by our model on the log-log plot.

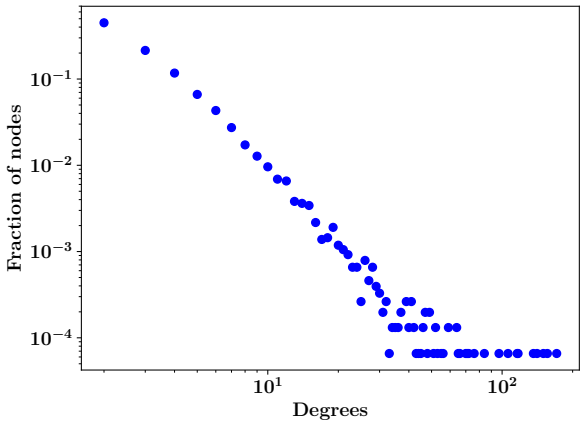


Fig. 4.1: Degree distribution of the generated network on the log-log plot with 15202 nodes and 30398 edges.

5. Mathematical Analysis

In this section, we will prove that the degree distribution of generated networks follows the power law. The evolution process starts with initial m_0 nodes and at each iteration, a new node is added to the network. We have assumed that the number of connections made by the new node is always more than or equal to 2 ($m \geq 2$).

Lemma 5.1. *The degree distribution of node i is given by $k_i(t) = B(\frac{t}{t_i})^n$*

Proof. Let k_i be the degree distribution of node i and a real time dependent variable. Then the rate of acquiring links by node i is given by the following equation.

$$\frac{\partial k_i}{\partial t} = m\Pi(\textit{node } i) = m\frac{\frac{1}{r_i}}{\sum_{j=1}^{t-1}\frac{1}{r_j}} \tag{5.1}$$

where m represents the number of links a new node connects with. We know that the size of grid is nxn hence we can say that $r_j < n$. Thus we rewrite equation 5.1 as.

$$\frac{\partial k_i}{\partial t} = m\frac{\frac{1}{r_i}}{\sum_{j=1}^{t-1}\frac{1}{r_j}} < m\frac{\frac{1}{r_i}}{\sum_{j=1}^{t-1}\frac{1}{n}} \tag{5.2}$$

Summing the denominator of equation 5.2 over $t-1$, we get.

$$\frac{\partial k_i}{\partial t} < m\frac{\frac{1}{r_i}}{\frac{t-1}{n}} \tag{5.3}$$

for large values of t we cen neglect -1 term, hence equation 5.3 becomes.

$$\frac{\partial k_i}{\partial t} < \frac{\frac{mn}{r_i}}{t} \tag{5.4}$$

According to our rules of evolution $m \geq 2$ and $r_i \geq 1$, thus we can say that $\frac{m}{r_i} \leq k_i$. Hence equation 5.4 becomes.

$$\frac{\partial k_i}{\partial t} < \frac{nk_i}{t} \tag{5.5}$$

For some constant c, we can rewrite equation 5.5 as.

$$\frac{\partial k_i}{\partial t} + c = \frac{nk_i}{t} \tag{5.6}$$

Solving the differential equation in 5.6, with the initial condition that node i joins the network at time t_i and connects with m nodes, we get.

$$k_i(t) = A \cdot t + B \cdot (\frac{t}{t_i})^n \tag{5.7}$$

where $A = \frac{c}{n-1}$ and $B = \frac{mn-m-ct_i}{n-1}$. For large values of t, we can neglect $A \cdot t$ term, hence equation 5.7 becomes.

$$k_i(t) = B(\frac{t}{t_i})^n \tag{5.8}$$

□

Lemma 5.2. *The degree distribution of networks generated by our model follows the power law.*

Proof. First, we will calculate that given the degree distribution of node i at time t i.e. $k_i(t)$, what is the probability that $k_i(t) < k$, where k is some arbitrary positive integer. Hence we can write.

$$P(k_i(t) < k) = P(t_i > \frac{B^{\frac{1}{n}}t}{k^{\frac{1}{n}}}) \tag{5.9}$$

As described in the evolution process, we start the spiral traversal with an initial m_0 nodes on the grid, and at each time step a new node with m connections is introduced in the network. Using this fact, we can write the probability density function of node i at time t as.

$$P(t_i) = \frac{1}{m_0 + t} \tag{5.10}$$

Substituting the equation 5.10 in equation 5.9, we get.

$$P(k_i(t) < k) = 1 - P(t_i \leq \frac{B^{\frac{1}{n}}t}{k^{\frac{1}{n}}}) = 1 - \frac{B^{\frac{1}{n}}t}{k^{\frac{1}{n}}(m_0 + t)} \tag{5.11}$$

Taking the derivative of equation 5.11 to get the probability density.

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{1}{n} \cdot \frac{B^{\frac{1}{n}}t}{m_0 + t} \cdot \frac{1}{k^{\frac{1}{n}+1}} \tag{5.12}$$

From equation 5.12, we can say that.

$$P(k) \propto k^{-\gamma} \tag{5.13}$$

where γ is $\frac{1}{n}+1$. This shows that the networks generated by our model follows power law distribution. This completes our proof. □

6. Validation

In this section, we will validate our proposed model on a real-world network, i.e. *Erdös collaboration network*.

A. Erdös Collaboration Network

Paul Erdös was one of the most prolific mathematicians of 20^{th} century. He was known for wandering the world and collaborating with mathematicians. The Erdös collaboration network has as vertices all mathematicians, with an edge joining u and v if u and v have published a joint research paper. One of the nodes p of the network is Paul Erdös. The distance from a vertex u to p is known as $u's$ *Erdös number* [12]. The mathematicians who have collaborated with Paul Erdös have Erdös number 1 and the mathematicians who have collaborated with those mathematicians have Erdös number 2 and so on.

B. Dataset

In our experiment, we combined the datasets provided by the National Institute of Standards and Technology and The Erdös Number Project, Oakland University. We collected the Erdös collaboration network of 2002 from NIST in the edgelist format[1]. It contains 6927 mathematicians as vertices, including Paul Erdös and 11850 edges representing coauthorship. The names of all the mathematicians are also provided in the dataset. To get the Erdös number of mathematicians we downloaded the dataset provided by The Erdös Number Project, Oakland University containing names of all the coauthors of Paul Erdös[2]. After combining both the dataset, a total of 507 mathematicians were found to have an Erdös number 1 in our dataset, and rest all have Erdös number 2.

C. Experiments and Results

The degree distribution of Erdös collaboration network on the log-log plot is shown in figure 6.1. Based on the degree distribution we can say that the Erdös collaboration network is scale-free. To validate our proposed grid based model for generating scale-free networks, we modeled the Erdös collaboration network on a grid where the center of the grid is Paul Erdös and the Euclidean distance of nodes from the center is represented as the Erdös number of the mathematicians.

Our results are positive as the highest degree node is at the center, i.e. Paul Erdös and most of the other higher degree nodes have Erdös number 1. Around 12847 degrees are associated with just 507 nodes having Erdös number 1, whereas only 10346 degrees are associated with 6420 nodes having Erdös number 2. Thus we can see that, as we move away from the center of the grid the degree distribution of the nodes decreases drastically, as proposed in our model. Therefore Erdös collaboration network validates our proposed grid based model for generating scale-free networks.

[1]http://vlado.fmf.uni-lj.si/pub/networks/data/Erdos/Erdos02.net

[2]https://files.oakland.edu/users/grossman/enp/Erdos0.html

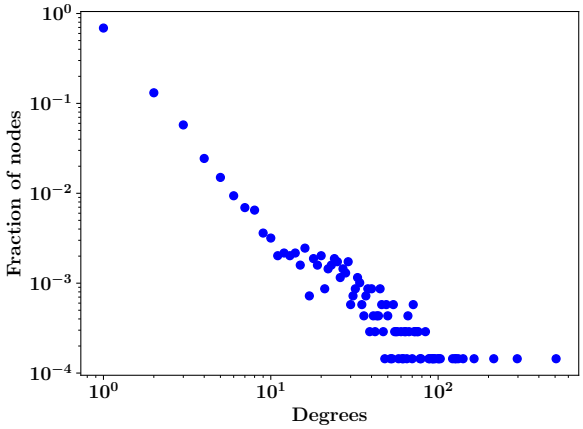


Fig. 6.1: Degree distribution of the Erdös collaboration network on the log-log plot with 6927 nodes and 11850 edges.

Conclusion

We have worked on the generation of scale-free networks. To capture the notion of the evolution of these networks we have formulated a grid based model. Unlike the Barabási-Albert model, our model is not purely combinatorial but uses insights from elementary Euclidean geometry. The experiments gave positive results in which we saw that the degree distribution of the generated networks follow the power law, thus confirming that they are scale-free networks. We have also proved its correctness. Further, the Erdös collaboration network validated our model in the real world scenario. Our model can also be used to explain the evolution of other complex networks such as location-based online social networks, airport networks, earthquake networks etc.

References

[1] C. Dangalchev, "Generation models for scale-free networks," in Physica A: Statistical Mechanics and its Applications, vol. 338.3-4, pp.659-671, 2004.

[2] T. Hellmann and S. Mathias, "Evolution of social networks," in European Journal of Operational Research, vol. 234, no. 3, pp. 583-596, 2014.

[3] A. K. Verma, and M. Pal, "Evolving Social Networks via Friend Recommendations," in 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2015 IEEE, 2015.

[4] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," in science, 286(5439), pp. 509-512, October 1999.

[5] D. Price, "Statistical studies of networks of scientific papers," in Statistical Association Methods for Mechanized Documentation: Symposium Proceedings, vol. 269, p. 187, US Government Printing Office, 1965.

[6] R. Albert and A.-L. Barabsi, " in Statistical mechanics of complex networks," Reviews of modern physics, vol. 74, no. 1, p. 47, 2002.

[7] P. ERDds and A. R&WI, "On random graphs I," in Publ. Math. Debrecen, vol. 6, pp. 290-297, 1959.

[8] A.-L. Barabasi, "Linked: How everything is connected to everything else and what it means," in Plume Editors, 2002.

[9] Q. Chen and D. Shi, "The modeling of scale-free networks," in Physica A: Statistical Mechanics and its Applications, vol. 335, no. 1, pp. 240-248, 2004.

[10] S. S. Manna, P. Sen, " Modulated scale-free network in Euclidean space." in Physical Review E, 66(6), 2002.

[11] A. Artikov, A. Dorodnykh, Y. Kashinskaya, E. Samosvat, "Factorization threshold models for scale-free networks generation." in Comput. Soc. Netw. 3(1), 4, 2016.

[12] C. Goffman, "And what is your erdos number?," in The American Mathematical Monthly, vol. 76, no. 7, pp. 791-791, 1969.