Observing the Effect of Network Growth Model on the Friendship Paradox Phenomenon

Mohammad Amin Nazerzadeh

University of Bologna
mohammad.nazerzadeh@studio.unibo.it

Abstract. In this study, the widely known Friendship Paradox is studied in the context of two main network growth models, the Uniform Random model and the Preferential Attachment model, and the dynamics of the interaction of the nodes within the network. This helps in understanding and modeling the effect of different variables on this phenomenon and as a result, engineering ways to predict wanted outcomes from the dynamics of the models.

Keywords: Friendship Paradox · Uniform Random model · Preferential Attachment model

1 Introduction

The friendship paradox is the phenomenon first observed by the sociologist Scott L. Feld in 1991 that most people have fewer friends than their friends have, on average [1]. The logic underlying the phenomenon is mathematically explored, showing that the mean number of friends of friends is always greater than the mean number of friends of individuals (Under the condition of existing variability among the number of friends of the people). If you take i as an individual and x_i as the number of friends of that individual and N as the total number of people, the average number of friends of each individual is:

$$AVG_f = \mu(x) = \frac{\sum_{i\in I} x_i}{N} \tag{1}$$

while the average number of friends of friends of an individual is:

$$AVG_{ff} = \frac{\sum_{i\in I} x_i^2}{\sum_{i\in I} x_i} = \mu(x) + \frac{\sigma(x)^2}{\mu(x)} \tag{2}$$

As individuals in the network tend to measure their social adequateness by comparing the number of their friends to the number of friends of their friends, because of the aforementioned gap between the two distributions, on average people will feel inadequacy by this comparison. It is worth noting that the degree of this gap is proportional to the variability in the distribution of the number of friends in the network.

To study the two previously mentioned distributions, two Network Growth models are proposed. To be able to make a comparison between the models, both these models will have commonalities which will be explained in the following. The first model is the **Uniform Random Model** [3] which is an extension of the Erdos-Renyi model [2] for network growth. There exists a fully connected network at first with m number of nodes (people). A new node is born at each time step. The new node has m edges to allocate to the existing nodes and will do so in a uniform manner. While in the second model, the **Preferential Attachment Model** (also known as Barabasi-Albert model) [4], the node will distribute its m edges among the existing nodes based on the probabilities proportional to the number of friends of those nodes. This is also known as a "Rich get Richer" schema. Since the only difference between the two Dynamism is how a node distributes its edges, the comparison will yield meaningful results.

Another aspect of the study is that I not only consider friendship creation through the network growth dynamism but also friendship removal through individual interaction at the end of each step of the mentioned models. This adds more realism to the models, as individuals in real life tend to both make and remove friendships. The friendship removal will follow a random uniform process. To make friendship removal compatible with the network growth models, the number of friendships to be removed is chosen as a parameter, r, that can vary between zero and m which is the number of edges that will be added at each step. In this way, we can model societies with different degrees of tendencies to make friendships.

Moreover, a noise parameter, ϵ, is added to the model to cover the other not included parameters and factors which have an effect on the behavior of the model. This will also assess the robustness of the modeling.

Finally, to further test the robustness of the results, I included churning to study its effect on the measurements. The churn will happen at a specified step and is implemented by removing randomly a specified variable proportion of the nodes in the network. In this way, one could also study the friendship paradox phenomenon under extreme circumstances.

To observe the effect of different parameters on the outcomes of the models, I will first set a baseline for the models and observe the measurements. Then at each step, I change one variable of the baseline parameters and keep the others fixed, to see the effect of that specific variable on the outcomes.

2 Results

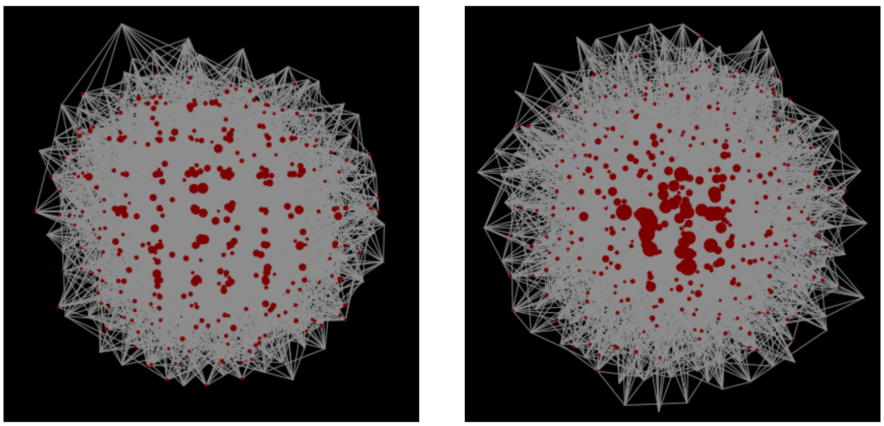
2.1 baseline models

The parameters of the baseline models are:

- $m = 12$: number of friendship formations at each step
- $r = 2$: number of friendship removal at each step
- $\epsilon = 0.05$: noise
- having a churn or not $\rightarrow$ no churn

The simulation runs for 500 steps and will have or have not a churn based on the churn flag on step 100.

In fig.1 you can see a visualization of the simulation of the two models. The red nodes correspond to individuals in the graph and the links correspond to friendships. The size of each node is an indicator of the number of friends of that node. As expected we can observe that in the Preferential Attachment models, some nodes emerge with higher values of degree centrality compared to the Uniform Random Model.



(a) Uniform Random Baseline (b) Preferential Attachment Baseline

Fig. 1: Visualization of the simulation of the baseline models. The size of each node corresponds to the number of friends of that node

In Fig.2 you can observe the distribution of degrees in the two baseline models. Here we can observe that the number of friends of individuals in the Preferential Attachment Model follows a long-tailed distribution while it is not the case in the Uniform Random Model.

In Fig.3 you can observe the distribution of the average number of friends of friends of individuals in the two baseline models. The preferential attachment model has on average higher values (as we will also see in the following results). The amount of gap between these two distributions will also be studied later.

In Fig.4 the two measures, the average number of friends and the average of the average number of friends of individuals are plotted as a function of the time step. The left half of the graph corresponds to the uniform random model and the right half of the graph corresponds to the preferential attachment model. You can observe that as time passes the average number of friends plateaus to a fixed number in both models - in a long run it will converge to $2 \times (m - r)$. While the gap between the average of the average number of friends of individuals increases due to the *rich-get-richer* phenomenon which amplifies inequality among node degrees. Note that we know analytically the degree distribution of these

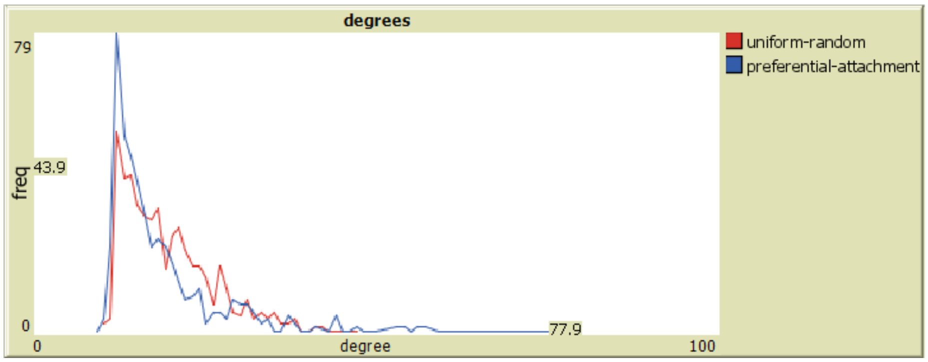


Fig. 2: Distribution of the number of friends of the two baseline models. The preferential attachment model follows a long-tailed distribution

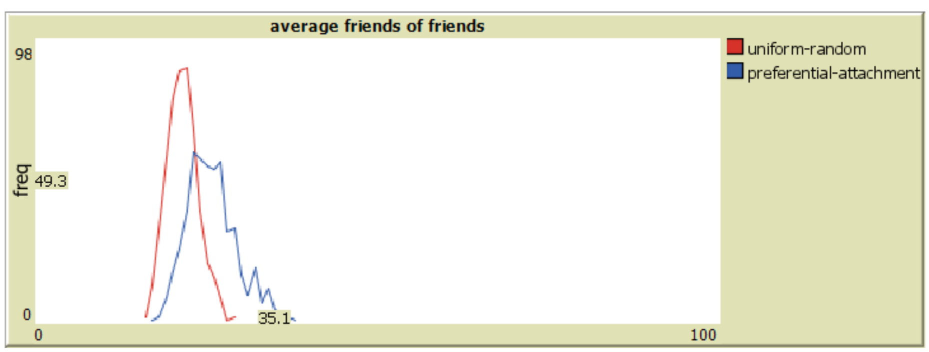


Fig. 3: Distribution of the average number of friends of friends of individuals in the two baseline models. The preferential attachment distribution has higher values due to the long-tailed distribution effect

two models. For the uniform random model, the degree distribution converges to an exponential distribution. Which has a well-defined mean and variance. As a result, we can predict that the average of the average number of friends of individuals will also plateau for this model. On the other hand, for the preferential attachment model, as it follows a power law with an exponent equal to $\alpha = 3$, though the mean is defined, the variance is undefined and unbounded[5]. As a result, I predict that the average of the average number of friends of individuals will grow unlimitedly over time.

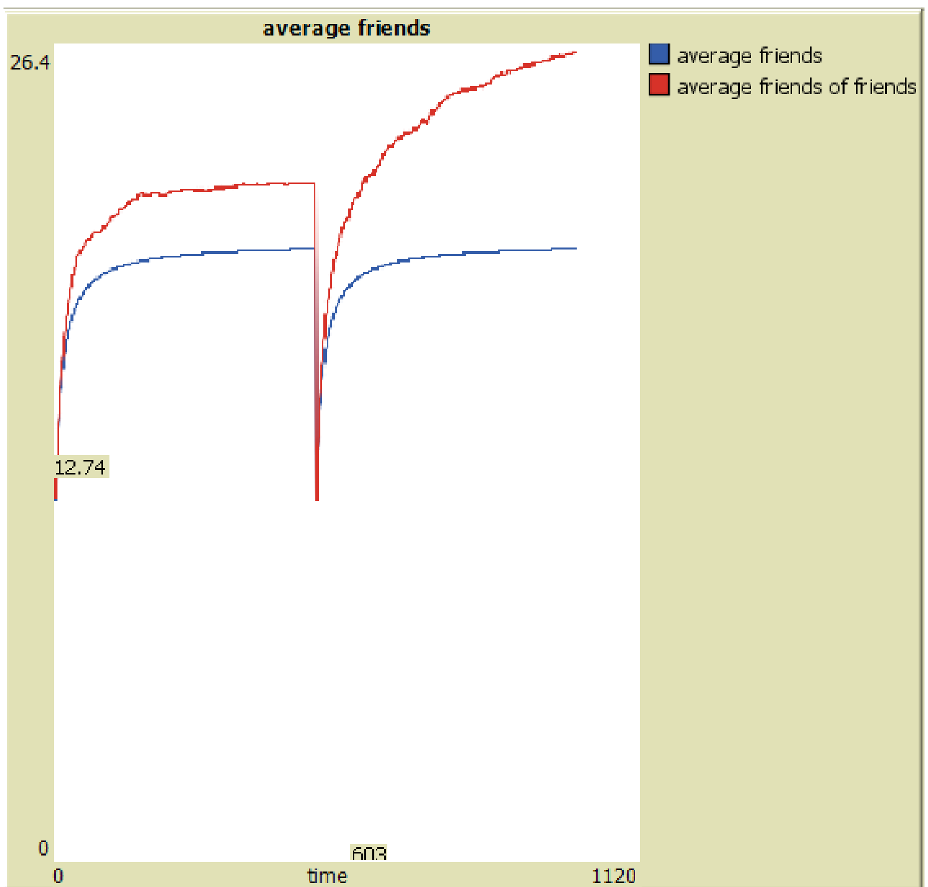
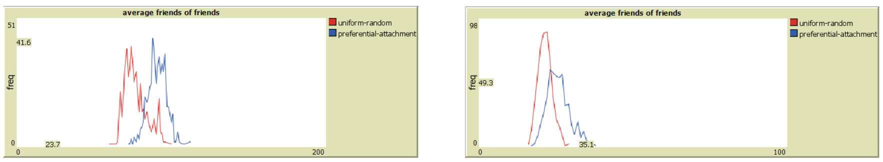


Fig. 4: plot of the average number of friends and the average of the average number of friends through time. Left half corresponds to the uniform random model and the right half corresponds the preferential attachment model

2.2 changing m (number of friendship formation)

Here the only parameter which is changed with respect to the baseline models is m which is set to 25.

In Fig.5 you can compare the distribution of the average number of friends of friends of individuals in this case and the baseline models. Increasing m had the effect of increasing the discrepancy between the two distributions compared to the baseline.



(a) m = 25 (b) m = 12

Fig. 5: Increasing the total capacity of new nodes to distribute edges in the graph in each step, increases the discrepancy between the distributions of the average number of friends of friends of individuals of the two models. Note that the scale of the x-axis in the figure. a is 2 times bigger

In Fig. 6 you can also observe that the gap between the average number of friends and the average number of friends of friends of the models increased in comparison to the baseline model.

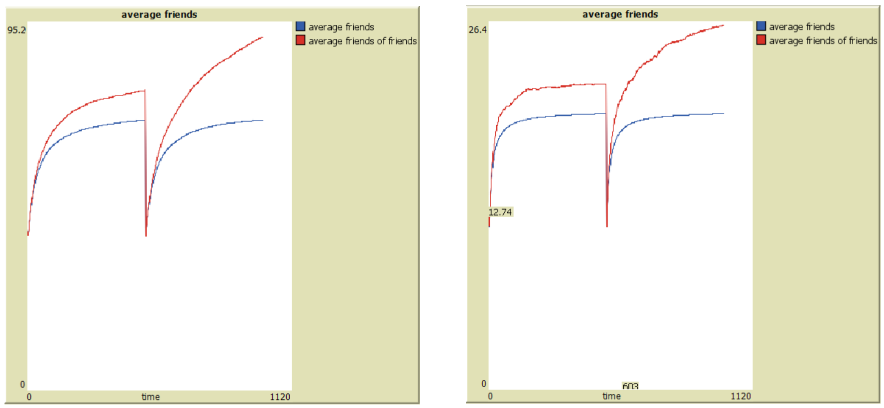
2.3 changing r (number of friendship removal)

Here the only parameter which is changed with respect to the baseline models is r which is set to 8.

In this case, a change in the behavior of the preferential attachment model happens. Due to its nature, after some time passes some nodes will lose all their friends and will never be able to receive friendship from the newly introduced individuals. As a result, the long-tailed distribution will be corrupted as is the case in the figure. 7

2.4 changing ϵ (noise)

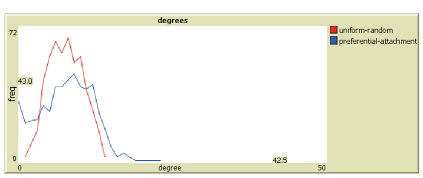
Here the only parameter which is changed with respect to the baseline models is ϵ which is set to 0.5. By changing this parameter, an interesting phenomenon happens. As the noise is applied both when making friendships and losing friendships, on average half of the friendships creation will not happen and half of the friendships removal will not happen. So in this case at each step on average, $\frac{12}{2} - \frac{2}{2} = 5$ new friendships will be formed and as a result, the average number of friends will decrease, and then plateau over time (Fig.8). Also in the uniform



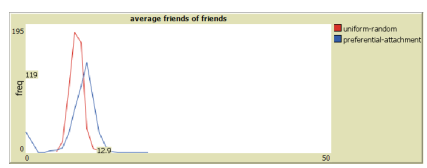
(a) m = 25

(b) m = 12

Fig. 6: Increasing the total capacity of new nodes to distribute edges in the graph in each step, increases the gap between the average number of friends of individuals and of the average number of friends of friends of individuals of the two models. Note that the scale of the y-axis in the figure. a is 3.6 times bigger



(a) distribution of the number of friends

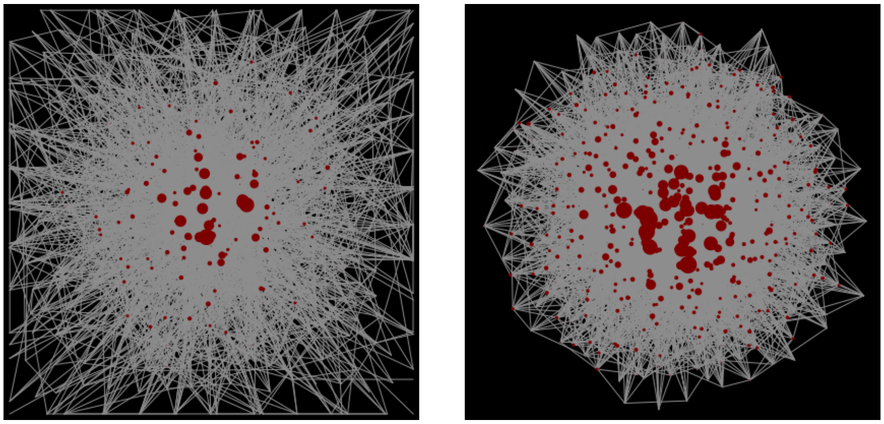


(b) distribution of the number of average friends of friends

Fig. 7: You can see that the preferential attachment model has a peak around zero in both figures. This is due to the fact that some individuals will lose all their friends and will never be able to make new friendships. This phenomenon decreases the distinction between the two models.

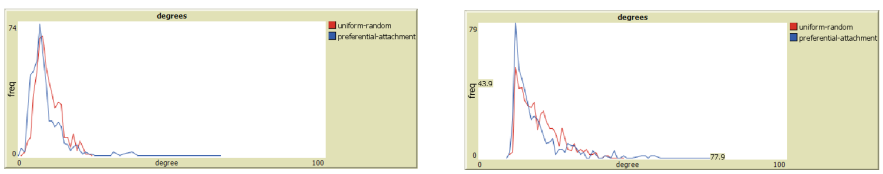
random model, the average friends of friends on average will also decrease but on the hand, in the preferential attachment model, the average friends of friends on average will increase (Fig.8). That is due to the fact that the number of nodes with a high degree of centrality will decrease (Fig.9), and as a consequence, we will have few nodes with a very high degree and other nodes with a very low degree (as you can see in Fig.10, the degree distribution of the preferential attachment model is more strongly heavy-tailed compared to the baseline models.) It also pushes the distribution of the number of average friends of friends of the preferential attachment model to flatten which is shown in Fig.11.

Fig. 8: Increasing the noise to 0.5 results in a decrease in the average number of friends in both models and the average number of friends of friends of individuals in the uniform random model. But it does not mitigate the average number of friends of friends of individuals in the preferential attachment model



(a) $\epsilon = 0.5$ (b) $\epsilon = 0.05$

Fig. 9: Increasing the noise to 0.5, will result in a higher concentration of friendships around just few central nodes and intensifying the friendship paradox phenomenon



(a) $\epsilon = 0.5$ (b) $\epsilon = 0.05$

Fig. 10: You can see that the tail of the preferential attachment models in both case end at nearly the same point in space for both cases. This is happening although the number of edges in the network of the figure on the left is roughly half of the edges of the network of the figure on the right.

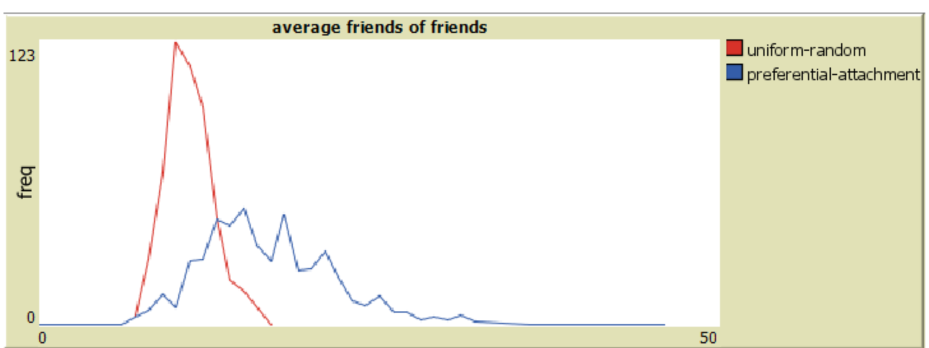
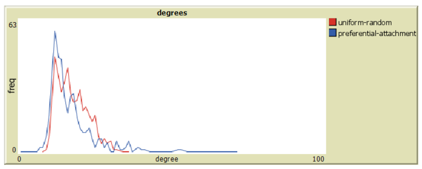


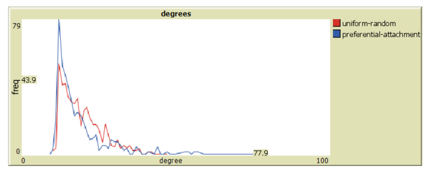
Fig. 11: The distribution of the average number of friends of friends is flattened in the preferential attachment model. This is due to the fact that almost all people are connected to few highly centered degree nodes.

2.5 adding churn to the baseline models

Injecting churn into the baseline model shows the robustness of the two models. In the following figures, you can see the comparison between adding churn to the baseline or not. The behavior is the same and it shows that the models are robust with respect to churn.

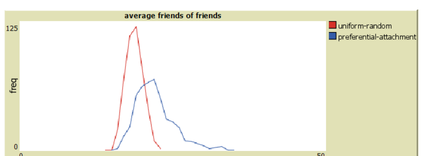


(a) with churn

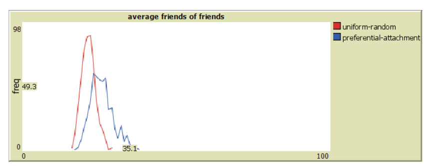


(b) without churn

Fig. 12: The behavior of the two models does not have a substantial change after injecting a churn. There is a decrease in the peak of the preferential attachment distribution that is due to the fact that churn happens randomly and may remove some nodes with a high degree of centrality

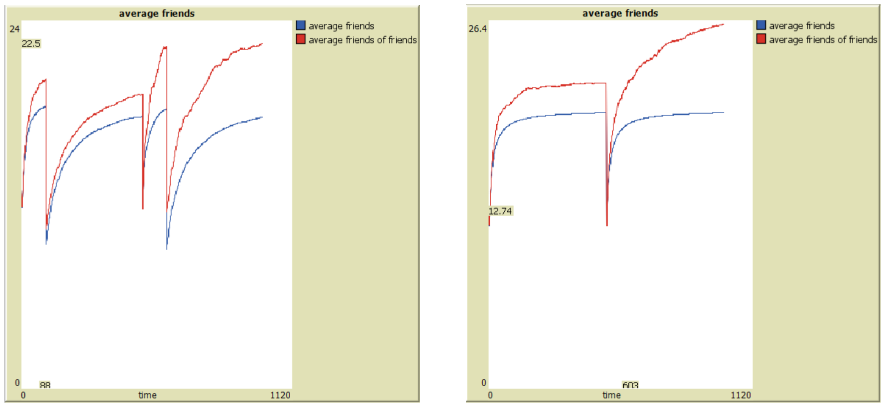


(a) with churn



(b) without churn

Fig. 13: The behavior of the two models does not have a substantial change after injecting the churn. The distribution of the preferential attachment model is flattened due to the fact that churn may have removed some nodes with a high degree of centrality. After that other nodes may be substituted in a competition to replace the highly centered remove nodes and this will result in a more flattened distribution of average number of friends of friends



(a) with churn (b) without churn

Fig. 14: On both figures the left half corresponds to the behavior of the uniform random model and the right half corresponds to the behavior of the preferential attachment model. You can see that the churn happens at the 100th step during the simulation of both but it does not affect the trend of the measurements and they will start increasing with the same trend as before

3 Conclusion

In this study, we observed how the friendship paradox phenomenon appears in two different network growth models. We observed that this phenomenon is more intensified in the case of the preferential attachment model. We observed that increasing m had an effect of increasing the discrepancy between the distribution of the average number of friends of individuals of the two models and increasing the gap between the average number of friends of individuals and the average number of friends of friends of individuals of the two models. It can model the fact that in modern societies in big cities, the feeling of inadequacy of members of the society is increased and intensified. We observed that by increasing r, the scale-free property of the preferential attachment model is corrupted. It can predict that individual interactions can mitigate the inadequacy effect that one can perceive by comparing the number of their friends to others. We observed that adding high noise to the models can lead to almost all people in the society being connected to few highly centered degree nodes. In the end, adding churn showed the robustness of our models to an extreme abrupt condition that could happen to the whole dynamic.

References

1. Feld, S. L. (1991). Why Your Friends Have More Friends Than You Do. American Journal of Sociology, 96(6), 1464–1477. http://www.jstor.org/stable/2781907

2. Erdos, P.L., Rényi, A. (1984). On the evolution of random graphs. Transactions of the American Mathematical Society, 286, 257-257.
3. Erdős, P.; Rényi, A. (1959). "On Random Graphs. I" Publicationes Mathematicae. 6: 290–297.
4. Albert, Réka; Barabási, Albert-László (2002). "Statistical mechanics of complex networks" (PDF). Reviews of Modern Physics. 74 (47): 47–97
5. Newman, M. E. J. (2005). "Power laws, Pareto distributions and Zipf's law". Contemporary Physics. 46 (5): 323–351