The evolution of network science regarding topology and properties

Niklas Mayr
2110837326@stud.fh-kufstein.ac.at
FH Kufstein
Tirol, Austria

ABSTRACT

This paper highlights the most important benchmarks in the recent research history of network science.

In particular, I address the network topology of the so-called Erdős and Rényi network in terms of the network model they developed to describe random networks. Subsequently, I discuss the emerging limitations of this model in regard to real networks. Accordingly, I address on the extension of the Erdős-Rényi network by Watts and Strogatz. Due to the emergence and development of the world-wide-web and social networks as a consequence, a new topology and properties of networks have emerged in the research area of Network Science, which will be discussed using the Barabási-Albert model.

KEYWORDS

random networks, degree distribution, scale-free property, Barabási-Albert Model

1 INTRODUCTION

In the history of Network Science, it has emerged that a number of assumptions about the topology and properties of random networks do not correspond to the nature of real networks.

The scientific work of Erdős, Rényi and Gilbert reached the first milestone regarding the properties of random networks in the history. Through a literature review, In this short paper, I will examine the main cornerstones in the evolution of network science in the last decades based on high-quality literature. In this research, I will particularly focus on the historical milestones of the two researchers Watts Strogatz, as well as the research papers of Barabási and Albert.

The two first-mentioned mathematicians, Watts and Strogatz, discovered serious limitations of the Erdős, Rényi and Gilbert model concerning the properties of random networks. Within their publications, they build a model, which can successfully predict the coexistence of a high clustering and a short average path length in a random network.

However, the results of Watts Strogatz fail to explain the degree distribution in the network. At this point, I will explain the so-called scale-free network, whose properties then provide the basis for the Barabási-Albert model. The detection of scale-free networks, which follow a power-law distribution, proved the existence of hubs in real networks. This existence of nodes in a network with a significantly high degree of linkages was proved by the two observations of Preferential Attachment and Growth in scale-free networks. Preferential attachment refers to the condition that nodes, which are newly added to the network are likely to link to nodes that already have a higher degree than other nodes in the network. The study of Barabási and Albert has furthermore brought out a model for the construction of scale-free networks under the coexistence of Preferential Attachment and Growth.

Each of the evolutionary steps mentioned here with regard to the network topology and its properties, has its justification in a certain way and will probably keep it in the future.

It should be noted, however, that in explaining the origin of real network properties, such models that can capture the emergence of the network are required [2].

2 LITERATURE REVIEW

In my literary research I focus among other studies on the publications of P. Erdős and A. Rényi: "On random graphs" from 1960 and the work of A.-L. Barabási: "Linked: The new science of networks" from 2003.

2.1 Random Networks

In general, it can be said that the main objective of network science is to model the network properties and to apply the findings to use-cases in the real world. [2].

If you look at a network visualized with nodes and edges from a bird's eye view, it is often very difficult to recognize a clear architecture and the nodes seem to be connected completely randomly (Figure 1). Frequently, the composition of nodes and edges seems to be random. This makes a prediction for a possible model with regard to the network architecture quite difficult. The random network theory takes advantage of this apparent randomness to explain the creation and characterization of networks that are truly random.

One of the fundamental properties of random networks is that they consist of N nodes, where each pair of nodes is connected by a probability p [4]. Furthermore, there are two other definitions of random networks. On the one hand, there is the so-called $G(N/L)$ model. Here the N labeled nodes are connected with randomly placed L on the left. The two mathematicians Erdős and Rényi created this definition in 1959 in some published papers on random networks. On the other hand, there is another definition which states that each pair of N labeled nodes is connected by a probability

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, July 2017, Washington, DC, USA
© 2023 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

p [3]. The so-called $G(N/p)$ model was developed by Edgar Gilbert simultaneously with the work of Erdős and Rényi.

The $G(N/p)$ model thus enriches the properties of random networks by the probability with which two nodes are connected, while the $G(N/L)$ model assumes a limited number of connections in a network. Due to the common research conducted in regard to random networks, the model that finally gained acceptance in the literature is also called the Erdős-Rényi-Gilbert model. Among these is the $G(N/p)$ model, which is used more often not only because of its simpler way to compute, but also because in real networks it is very rare that there is a fixed number of nodes and links [2].

For the creation of a random network the following steps have to be considered:

- initially, assume a number N of nodes which are isolated from each other
- further, you select a random pair of nodes, and generate a random number between 0 and 1. If this number exceeds the predefined threshold for p, the selected pair of nodes will be linked together. Otherwise, they remain isolated
- the second step is repeated continuously for each of the $n(n-1)/2$ node pairs

As I mentioned before, this procedure generates a random graph network, which is called Erdős-Rényi-Gilbert network, since their mathematical work has made an important contribution to the understanding of the properties of these networks. Of course, each, randomly created network will have smaller and larger differences by the parameters N and p. For example, if you increase the value of p, the network topology will become denser [3].

2.1.1 Degree Distribution of the Erdős-Rényi-Gilbert model . In this subsection I would like to mention the degree distribution of random networks. In the creation and observation of random networks, it can be seen that some nodes have a significant number of links, while others have a very small number of links, or are completely isolated (Figure 1). These differences in the distribution are explained by the degree distribution pk, which indicates the probability of a random node having the degree k.

The shape of the distribution of a random network, follows the form of a binomial distribution (Figure 2). However, for $N >>< k >$ this binomial distribution can also be approximated by a Poisson distribution. Even though both distributions have the same properties, they are expressed by different parameters. While the binomial distribution depends on the two parameters p and N, the Poisson distribution only relies on the parameter $< k >$. Due to its simplicity, the Poisson distribution is often preferred as an approximation in calculations. Due to the fact that most real networks are sparse, this limitation leads to the fact that the degree distribution can be well approximated by the Poisson distribution [1].

In summary, the Poisson distribution is only an approximation of a random network. Its key feature is that the properties are independent of the network size and depend only on a single parameter $< k >$ (average degree). In the case of smaller networks (10^2) the degree distribution deviates strongly from the Poisson distribution and can be explained much better by a binomial distribution. The larger a network is ($10^{3,n}$), the more the degree distribution necessarily deviates from the Poisson distribution [1].

Thus, in a random network, all individuals represented by nodes have a comparable number of connections. The problem with this is that it disregards those individuals that have more connections than others. The relation to reality is missing, because "famous" individuals are likely to have more connections than the average individual. This is the logical consequence of an important property of random networks: "in a large random network, the degree of most nodes is in the narrow vicihity of <k>" [3].

2.1.2 Watts-Strogatz model. In 1998, the two scientists, Duncan Watts and Steven Strogatz, published their extension of the random network model in Nature magazine. Their extension is based on two observations.

- small-world property
- high clustering

Watts and Strogatz found that the average distance in real networks between two nodes is logarithmically dependent on N, instead of

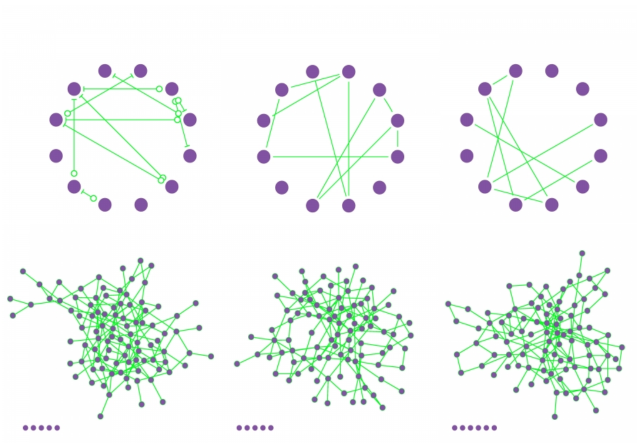

Figure 1: Different degree distribution of nodes in a random network

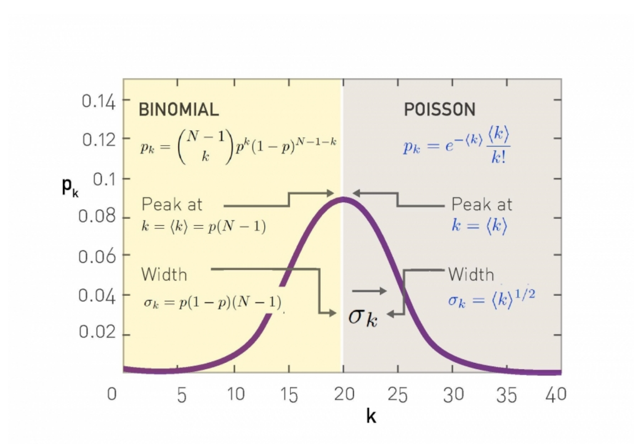

Figure 2: Binominial vs Poisson Distribution

following a polynominal expected regular lettices. Furthermore, it turns out that the average clustering coefficient of real networks is significantly higher than expected in random networks [1]. The Watts-Strogatz model interpolates between a regular grid, which on the one hand has high clustering, but no small-world phenomena, and a random network, which has low clustering but corresponds to small-world phenomena. As an extension of the Erdős-Rényi-Gilbert model, the Watts Strogatz model predicts a Poisson Distribution as well [5].

The research of Watts and Strogatz proved an important assumption in the evolution of network science: "Real networks are not random". Evidence for this can be seen, for example, in an assumed random society. According to this assumption, an Austrian student would be just as likely to have links in a small African town as to establish links in his own town. Another example is the network of proteins, where the interactions between proteins must follow strict biochemical rules [1].

2.2 The scale-free property

The World Wide Web is a good example of a real world network. Especially when you talk about the existence of hubs, nodes with a significantly high degree. The presence of hubs is not unique in the www. Hubs are also an expression of a deeper organizational principle, which is also called the scale-free property [1]. The physicist Hawoong Jong mapped the nd.edu domain in 1998, with a cumulative total of about 300,000 documents and 1.5 million links. The sequence of images shows an increasingly enlarged local region of the network (Figure 3).

A major difference between a scale-free network and a random network is in the tail of the degree distribution. While a random network can be approximated by the Poisson distribution, the scale-free network follows the so-called power-law distribution. The key difference between these two distributions is in their different shapes. While the nodes in a random network have a comparable degree and hubs are "forbidden", the power-law distribution even expects hubs. Moreover, the more nodes there are in a network, the larger these hubs become. In a power-law network, most nodes have only a few links, which are held together by a few highly linked nodes. Therefore, it can be seen that random networks have a scale, while scale-free networks have a lack of scale [3].

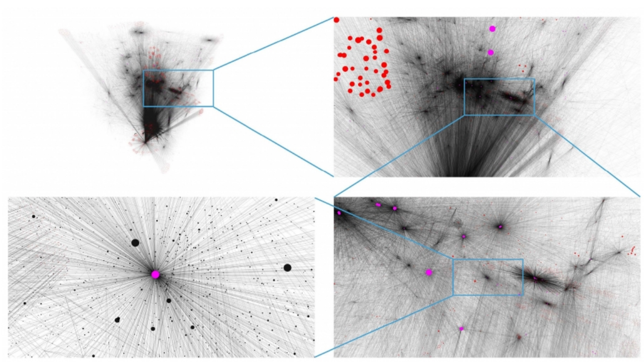

Figure 3: Mapping of the nd.edu domain by Hawoong Jeong

In conclusion, scale-free networks have played a very important role in the development and evolution of network science. The reasons for this are that firstly a lot of networks are scale-free and therefore the role of the scale-free property is unavoidably important. Nevertheless, it must be said that not all networks are scale-free. An example is the power-grid network, which consists of generators and switches and is held together by transmission lines. Another important point for the important role of scale-free networks, is the presence of hubs. These fundamentally change the network characteristics, similar to the small-world phenomena.

2.3 The Barabási-Albert Model

The World Wide Web has few websites that have an exceptionally large number of links. Examples would be google.com or facebook.com. The existence of hubs in scale-free networks, as already mentioned, is the big difference compared to random networks. Therefore, the question arises: why are hubs missing in random networks? This question can be answered relatively easily on the basis of two points. On the one hand, a random network assumes that there is a fixed number of nodes. In a scale-free network, it expands by adding nodes. On the other hand, a random network assumes that the link between two nodes is random. In the world of scale-free networks, newly added nodes prefer to link to nodes that already have a higher degree than other nodes in the network [3].

These two important concepts of scale-free networks are named Growth and Preferential Attachment. These main properties are the cornerstones of the Barabási-Albert model, which can create scale-free networks. At the end of the 1990's, the two mathematicians developed this model.

Preferential attachment follows a probabilistic mechanism. When a new node is added to the network, preferential attachment implies that between linking to a node with degree = 2 and one with degree = 4, the chance that this node will choose the node with degree = 4 is twice as likely. The concept of growth just means, that there exists a continuously process to add a new node at each time step [1].

The two properties growth and preferential attachment must coexist. If, for example, preferential attachment is missing, this leads to a growing network which has a stationary but exponential degree distribution. In contrast, the lack of growth would mean that, on the one hand, stationary would be missing and the network would converge to a complete graph [1].

3 CONCLUSION

The foundations placed by Erdős, Rényi and Gilbert regarding the properties of random networks will probably still be relevant in the future. Only through this work it was possible to uncover the small-world phenomenon by Watts-Strogatz. Thus if we want to explain the origin of particular network properties, other models have to be chosen that can explain the network origin and its growth. The Barabási-Albert model has shown that some real networks, such as the www, can be explained by preferential attachment and growth in a power-law distribution [2].

REFERENCES

[1] Albert-László Barabási. 2012. *Luck or reason.* 10.1038/nature11486.
[2] Albert-László Barabási. 2016. *Network Science.* http://networksciencebook.com.
[3] Jeong Barabási. 2020. *Mean-field theory for scale-free random networks.* 10.1016/s0378-4371(99)00291-5.
[4] Renyi Erdös. 2020. *On random graphs I. Dedicated to 0, Vargo, at the occasion of his 50th birthday networks.* https://snap.stanford.edu/class/cs224w-readings/erdos59random.pdf.
[5] Strogatz Watts. 1998. *Collective dynamics of 'small-world' networks.* 10.1038/30918.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009