Matemática Contemporânea, Vol. 48, 167–180
http://doi.org/10.21711/231766362021/rmc4817

Network Science: a primer

Joao Meidanis

Abstract

Euler's work on the Bridges of Königsberg problem is usually regarded as the birth of Graph Theory. But combinatorialists were not alone studying graph-like structures. In the beginning of the 20th century, sociologists started studying social networks. The idea of "six degrees of separation" (any human on Earth can reach any other by at most six acquaintance links) is from this period. More recently, other scientists took to looking at real networks focusing on degree distributions, hubs, communities, preferential growth and other evolution hypotheses, and the like. In this lecture we will review the main points of modern Network Science, exploring scale-free networks, degree correlations, robustness and its relationship to spreading phenomena, e.g., pandemic modeling.

1 Introduction

This presentation aims to be a brief introduction to the ideas explored in modern network science in the last 20 years or so. Although the subject has a significantly longer history and is effectively an attempt to gather ideas and techniques from multiple knowledge fields, these last two decades

2000 AMS Subject Classification: 90B18, 90C35, 91D30, 91G45, 92C42.
Keywords and Phrases: network science, small-world property.
This research was partially supported by FAPESP grant 2018/00031-7

have witnessed the influence of a certain view over it, brought mainly by physicists.

Physics usually concerns itself with planets, heat, electrons, things of nature, that have existed forever. In contrast, most relevant networks are very recent and are all the creation of human beings: the internet, the World Wide Web, social networks. Can they be considered "things of nature", too? They can be compared to spider webs or bird nests: artifacts produced by living beings. In this respect, they can be considered things of nature.

In any case, they seem to have another property that makes them amenable to physics studies: they are often composed of a large number of small structures, so that laws of big numbers apply, much like matter is the composition of a large number of small molecules. Calculus, in the form of differential equations, can be used to study their behavior.

In the rest of this note we exemplify this approach exploring some of the properties of complex networks, such as the small-world property (another way of saying, "low diameter"), specific forms of their degree distributions, correlations between degrees of neighbors, robustness against random failures and against orchestrated attack, with applications on the spreading of news, information or diseases.

2 The Small World Property

You may have heard of "six degrees of separation", the idea that every person on Earth is separated from any other person by at most 6 acquaintance links. The exact number may not be 6, but the gist of the concept is that the diameter of the acquaintance network is small.

This surprising fact is the basis for studies trying to uncover bounds for the diameter of networks, as well as explanations for their low value. The topic has a long history, going back to the beginning of the twentieth century (see Table 1).

Low diameters may seem surprising because many networks familiar

1929	Hungarian writer F. Karinthy used the idea in some of his books [8].
1958	Mathematician M. Kochen and political scientist I. de Sola Pool wrote first mathematical analysis on the subject, published in 1978 [5], but widely circulated since 1958.
1967	Social psychologist S. Milgram [10, 13] performed the first experiment testing the idea. Found 5.2 mean.
1991	Broadway play "Six degrees of separation" by John Guare popularized the idea.
1998	Physicists D. J. Watts and S. Strogatz *Nature* paper: how a few changes in regular graph make it small-world [14].
1999	Physicists H. Jeong, R. Albert and A. L. Barabási *Nature* paper: estimated 19 degrees for WWW [1].
2011	A Facebook Data Team [2] found 4.7 mean separation for their social network at the time.

Table 1: Brief historical account of the "six degrees of separation" idea.

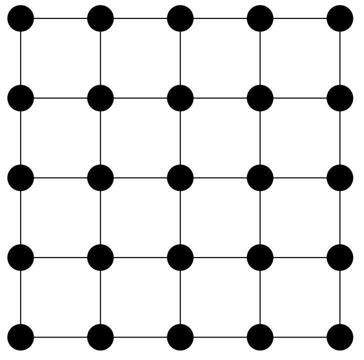


Figure 1: A 5×5 grid. Its diameter is $2(5-1) = 8$.

to us have larger diameters, when compared with the number of nodes. For instance, in an $n \times n$ grid (see Figure 1), the maximum distance is $2(n-1)$ and the number of nodes is $N = n^2$. This generalizes to any value of n, so that the diameter is proportional to $N^{1/2}$ in square grids, where N is the number of nodes. Likewise, in cubic grids, the diameter is proportional to $N^{1/3}$, and in general, the diameter is proportional to $N^{1/d}$ in d-dimensional grids. Therefore, in this type of network, the diameter varies with a positive power of N.

Another class of familiar networks are the cycle powers. A cycle C_N is just your familiar structure with every node having degree two and all in one connected component. That's the first power of a cycle. A finite power C_N^k for an integer $k \geq 1$ is a graph with the same nodes as C_N, but with edges xy for any two nodes at distance k or less in C_N. These graphs also have large diameters, typically proportional to $N/(4k)$ for k not exceeding $N/2$. However, Watts and Strogatz [14] show that by randomly exchanging a fraction of the links in such graphs one quickly smashes the diameter to small-world ranges (see Figure 2).

Most real networks, however, display the small-world property. Table 2 shows ten networks present in our everyday lives, such as the internet, some communication networks (email messages, mobile phone calls), collaboration and citation networks, and some biological networks (trans-

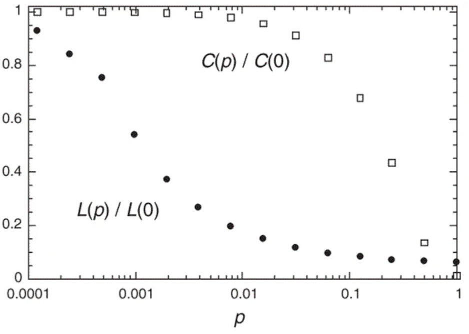


Figure 2: The diameter of a cycle power drops rapidly when we randomly exchange a fraction of its links. Notice how the diameter gets halved in this example with just 0.1% random exchanges. Adapted from ref. [14].

formation of compounds in the metabolism of bacteria *Escherichiacoli*, and yeast protein interactions). In all of them the average degree $\langle d \rangle$ is remarkably close to the ratio

$$\frac{\ln N}{\ln\langle k \rangle},$$

suggesting a logarithmic growth of $\langle d \rangle$ with respect to the number of nodes N.

3 Scale-free Networks

Is the World Wide Web (WWW) a random network? This question was raised by researchers around the turning of the century. We clarify that the meaning of "random network" can be understood in two ways: as "a network built randomly" (which is not the intended meaning but is usually taken in opposition to, e.g., "deterministic network") or as "a network uniformly chosen among all possible" (which is known in modern Network Science as the Erdős – Rényi model). The answer is no, and one of the main reasons for that is its degree distribution. In a random network, most nodes have degrees around the average, without much deviation. In contrast, in the WWW we have dramatic differences in degree. Most nodes have just a few neighbors, but a small fraction of nodes are hubs,

Network	N	$\langle k\rangle$	$\langle d\rangle$	$\ln N/\ln\langle k\rangle$
Internet Routers	192,244	6.34	6.98	6.58
WWW Documents	325,729	4.60	11.27	8.31
Power Grid Stations	4,941	2.67	18.99	8.66
Mobile Phone Calls	36,595	2.51	11.72	11.42
Email Messages	57,194	1.81	5.88	18.40
Science Collaboration	23,133	8.08	5.35	4.81
Paper Citations	449,673	10.43	11.21	5.55
Actor co-staring	702,388	83.71	3.91	3.04
E. coli Metabolism	1,039	5.58	2.98	4.04
Protein Interaction	2,018	2.90	5.61	7.14

Table 2: Real-life networks, with their numbers of nodes (N), average degrees ($\langle k\rangle$), average distance $\langle d\rangle$, and the ratio $\ln N/\ln\langle k\rangle$. Notice how close this last ratio is to the average degree, with the possible exception of email messages, showing a logarithmic dependence on the number of nodes. It is worth mentioning that the ratio $\ln N/\ln\langle k\rangle$ is also somewhat different from $\langle d\rangle$ in the power grid and citation networks, though in the other direction ($\langle d\rangle$ is larger). Adapted from *Network Science*, by Barabási and Pósfai [3].

that is, have a huge number of neighbors. The degree distribution of the WWW follows a power law, with the probability p_k of a node having degree k being roughly proportional to

$$p_k \propto k^{-\gamma},$$

where γ is a constant depending on the network. This is markedly distinct from the binomial distribution one finds in random networks [6, 7].

Besides the WWW, several other real networks have been observed to follow a power law for their degree distributions, or something derived from a power law with adjustments in the very small and very large degree ranges [4].

Networks that display a power law degree distribution are said to be *scale-free*, because, for a certain range of the exponent γ, the first moment of the distribution exists, so you have an average degree, but the second moment diverges, so the standard deviation is unbounded as the number of nodes grows. Therefore, there is not a "scale" that would have been provided by the standard deviation. Even in finite networks the moment divergence is visible. For instance, in the WWW we have that the in-degree satisfies

$$k_{in} = 4.60 \pm 1546.$$

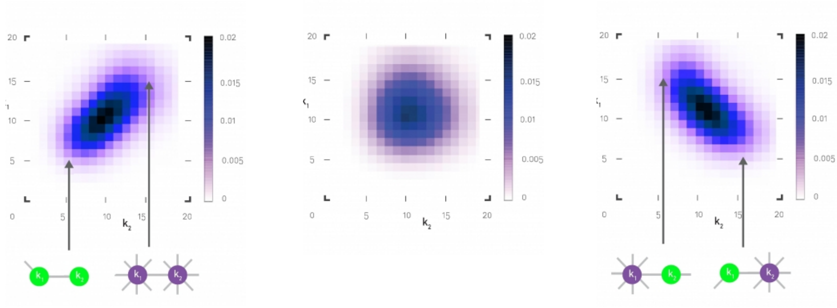
4 Degree Correlations

This section is concerned with the following question: do links connect nodes with similar degrees or with widely differing degrees? A network can be classified in three categories according to the answer to this question:

Assortative : favors links between nodes with similar degrees

Neutral : no particular bias in links with respect to endpoint degrees

Disassortative : favors links between nodes with widely different degrees



Assortative Neutral Disassortative

Figure 3: Heatmaps of the degree correlation matrix of three networks. High probabilities along the main diagonal suggest assortativity, while high probabilities along the secondary diagonal suggest dissasortativity. If neither occurs, we can say the network is neutral. Adapted from *Network Science*, by Barabási and Pósfai [3].

To determine the category of a given network, three approaches have been proposed. The first one is to build the degree correlation matrix, which is just a matrix with entries $e_{i,j}$ equal to the probability of finding a node with degrees i and j at the two ends of a randomly selected link. Plotting this matrix as a heatmap and inspecting the result can help us decide the network category. Figure 3 shows examples of matrices of each kind. Notice that in this figure the main diagonal runs from bottom left to top right.

Then we have the degree correlation function, denoted by $k_{nn}(k)$, which returns, given a degree value k, the average degree of the neighbors of all degree-k nodes. If this function is an increasing function, chances are the network is assortative. On the contrary, when this function decreases with k, this indicates a dissasortative network. Absence of these tendencies points to a neutral network. Figure 4 shows three examples taken from the networks in Table 2.

Finally, a third way of measuring the degree correlation of a network is to compare its degree correlation matrix $e_{i,j}$ with what would be expected for a randomly linked network with the same degree distribution, where

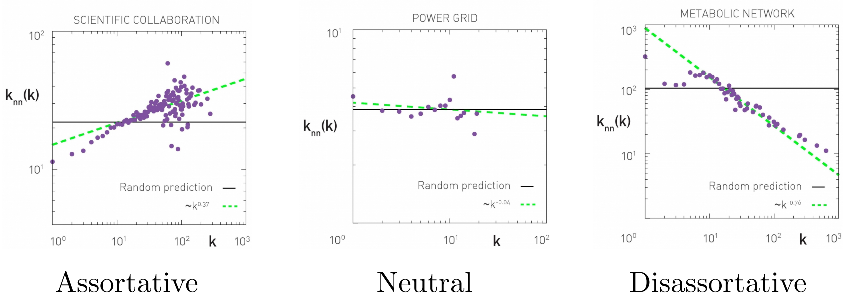


Figure 4: Degree correlation functions for three of the real networks listed in Table 2. Assortativity is associated with an increasing $k_{nn}(k)$. Adapted from *Network Science*, by Barabási and Pósfai [3].

the degrees at each extremity of a link are independent random variables.

5 Network Robustness

Network robustness aims at answering questions of the form: How resilient is a real network to random failures? How easy is it to break it with planned attacks? For us here, random failures will mean removal of nodes at random from the network, while planned attacks will refer to the removal of nodes in decreasing order of their degrees. Thus, hubs are destroyed first on a planned attack. Also, to "break" is to disconnect the giant component, that is, the largest connected component of the network.

From classical work on random graphs we know a useful condition for the existence of a giant component in a random network:

$$\langle k \rangle > 1.$$

However, most real networks are not random! Luckily, Molloy and Reed [11] generalized this result to networks still having their links chosen at random, but required to follow a given degree distribution. The condition then becomes:

$$\langle k^2 \rangle > 2\langle k \rangle.$$

For unrestricted random networks, where $\langle k^2 \rangle = \langle k \rangle(\langle k \rangle + 1)$, this formula is equivalent to the previous one. However, for other types of networks, e.g., scale-free networks, we have to rely on the latter.

As we saw, in scale-free networks $\langle k^2 \rangle$ is much larger than $\langle k \rangle$, so they're pretty robust against random failures. However, they're an easy prey of planned attacks: there are not that many hubs, and the mere removal of a small fraction of hubs is enough to cause the break-up of the giant component.

Network robustness has applications beyond the analysis of a single network. When we consider the underlying networks for the transmission of diseases or information, we may study the dynamics of their spread as a time-varying subset of the network links representing transmission paths. The giant component represents the fraction of the target population being affected by the phenomenon. This is the subject of the next section.

6 Spreading Phenomena

It could be a virus, or it could be news (real or fake). But it spreads along a network, until it reaches all nodes, or a fraction of the nodes. In any case, such phenomena are of interest and their study is high impacted by the type of underlying network.

Before the advent of modern network science, epidemiologists just made assumptions essentially considering the underlying network to be a random network. However, most real networks are not random, being more closely related to a scale-free network or one of its variants. The introduction of scale-free graphs leads to important consequences, implying a considerably faster spread of diseases.

A popular tool to study these phenomena is the SIR model, where the entire population is divided into three groups: the susceptible individuals (S), the infected individuals (I), and the removed individuals (R) [9]. People go from susceptible to infected according to a given rate β, but also depending on the number of contacts they have. Thus, the degrees of the underlying network play a role here. Removed individuals either die from the disease or recover, in this last case acquiring immunity against the infection. We model that with a single rate μ expressing the speed with which people go from infected to removed.

In this model, there's no going back to previous groups. Once infected, a person never becomes susceptible again. And, once removed, a person never gets infected again. Therefore, it is appropriate for diseases such as measles, in which recovered individuals become resistant to the disease for a long time afterwards.

The classical model relies on the following differential equations:

$$\begin{aligned}\frac{ds}{dt} &= -\beta\langle k\rangle i(1-i-r)\\ \frac{di}{dt} &= -\mu i+\beta\langle k\rangle i(1-i-r)\\ \frac{dr}{dt} &= \mu i\end{aligned}$$

where s, i, and r are, respectively, the fractions of people in the susceptible, infected, and removed groups, β is the infection rate, and μ is the recovery rate. There is no closed formula for the general solution of these equations, but it is possible to infer that the relationship between β and μ decides the fate of the infection, that is, whether it dies out or stays within the population forever.

However, given the scale-free nature of most spreading networks, a different set of equations was proposed in which, instead of the average value of k, a distinct equation for each value of k is used [12]. The modern equations take the form:

$$\begin{aligned}\frac{ds_k}{dt} &= -\beta k\theta_k(1-i_k-r_k)\\ \frac{di_k}{dt} &= -\mu i_k+\beta k\theta_k(1-i_k-r_k)\\ \frac{dr_k}{dt} &= \mu i_k\end{aligned}$$

Notice the introduction of a new function θ_k, indicating the number of infected nodes in the neighborhood of a susceptible node of degree k.

This function replaces i in the estimation of infection instantaneous speed. Instead of three equations, we now have three equations for each value of k, from 1 to k_{max}, the maximum degree in the network. The functions θ_k are responsible for linking equations with different values of k.

The new formulation leads to staggering conclusions, such as the fact that even pathogens that are hard to pass from one individual to another can spread successfully, causing an epidemic in scale-free networks.

References

[1] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *Nature*, 401:130–131, Sep 1999.

[2] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In Michael W. Macy and Wolfgang Nejdl, editors, *WebSci 2012: Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42, New York, NY, USA, Jun 2012. Association for Computing Machinery.

[3] Albert-László Barabási and Márton Pósfai. *Network Science*. Cambridge University Press, 2016.

[4] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nature Communications*, 10(1017):1–10, 2019.

[5] Ithiel de Sola Pool and Manfred Kochen. Contacts and influence. *Social Networks*, 1(1):5–51, 1978.

[6] Paul Erdös and Alfred Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

[7] Edgar Nelson Gilbert. Random graphs. *Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[8] Frigyes Karinthy. *Minden másképpen van (Ötvenkét vasárnap)*. Athenaeum, Budapest, 1929. English translation of chapter Chain-Links (acc. 2020-11-18): `https://edisciplinas.usp.br/pluginfile.php/4205012/mod_resource/content/1/Karinthy-Chain-Links_1929.pdf`.

[9] William Ogilvy Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A*, 115:700–721, 1927.

[10] Stanley Milgram. The small-world problem. *Psychology Today*, 1(1):61–67, May 1967.

[11] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6(2–3):161–180, 1995.

[12] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

[13] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, Dec 1969.

[14] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, Jun 1998.

Joao Meidanis
University of Campinas
Theory of Computing
meidanis@unicamp.br