Nelly Litvak
University of Twente, and
Eindhoven University of Technology
n.litvak@utwente.nl

Research

Randomness and structure in complex networks

How to mathematically describe real-life networks, such as social networks or the World Wide Web, as so-called random graphs? How do particular rules of placing random edges result in graphs that share some of the fundamental empirical properties of real-life networks? This article of Nelly Litvak is based on her lectures of PWN Vakantiecursus 2022.

Complex networks, modeled as random graphs

Many real-life systems are *networks*. A network is a set of objects connected by some relationship. For example, a railroad is a collection of stations connected by rails. In a social network, people are connected by friendships. The Internet is a network of routers connected by wires. In our brain, neurons are connected if they fire together.

A *graph* is a natural mathematical model for a network of any kind. In a graph, each object is represented as a *vertex*, and if there is a relationship between two vertices, then there is an *edge* between them. *Undirected edges* represent symmetric relations, and we draw them as lines. For example, communications between two Internet routers usually go in both directions. If the relation is not symmetric, we model this using *directed edges*, and draw them as arrows. For example, if somebody follows you on Twitter, you might not follow back. In Figure 1, we give some examples of networks, directed and undirected.

This article is about how to mathematically describe real-life networks, such as social networks or the World Wide Web, using so-called *random graphs*. Usually in research, and always in this article, the vertices of a random graph are fixed, while the edges

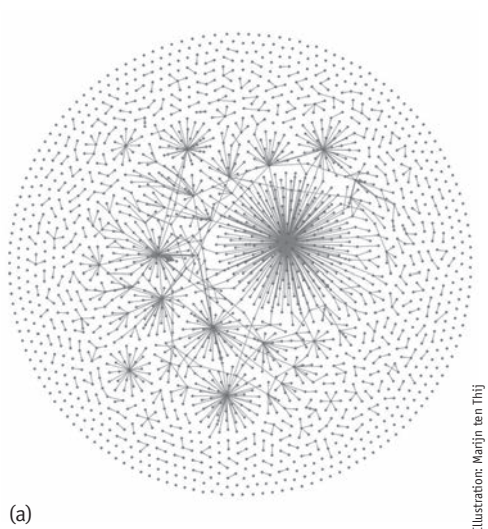

(a)

Illustration: Marijn ten Thij

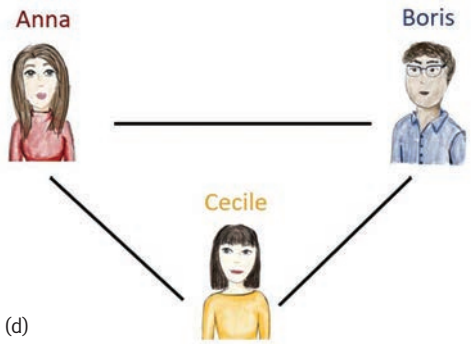

(b)

Illustration: Wikimedia Commons

# other webpages, from which a webpage can be reached by following a hyperlink	EU Web crawl 2015
average	85.743
largest	20 252 239

law.di.unimi.it/datasets.php

(c)



(d)

Illustration: Natalia Litvak

Figure 1 (a) Network of retweets about Project X in Haren, 21-09-2012, 7:00. The average number of retweets per user is small compared to the network size [11]. This network is *sparse*. (b) Any railway station can be reached by train from any other. The railway network is *connected*. (c) The number of other webpages, from which a webpage can be reached, varies greatly per webpage. The Webgraph is *scale-free* [3]. (d) Anna is a friend of Boris and Cecile. Then, Boris and Cecile are likely to be friends, too. Social networks have many *triangles*.

are placed at random. This makes sense because relationships between objects often emerge at random, like friendships in a social network. Also, even if a network is not random, such as the Internet, its structure is so complicated that it is often useful to describe it through statistical summaries and model it as a random object. A *random graph model* is in fact a set of rules, according to which the random edges are chosen. Different rules result in different models with different mathematical properties. For example, in one model, edges between all vertex pairs can be equally likely, while other models assign higher probabilities to some vertex pairs.

In this article, I will tell about several random graph models that are by now well understood, even classical. My goal is to explain how particular rules of placing random edges result in graphs that share some of the fundamental empirical properties of real-life networks. There are many such properties, but I will address four of them, listed in the next section.

Four fundamental properties of real-life networks

Surprisingly, many networks of a completely different nature, such as social networks and the Internet, share common properties. This is why we talk about a *structure* of a network. In this article I will discuss four such common structural properties.

- *Sparse*. Consider a social network. Even if the network is very large, a person can maintain only so many friendships. We say that social networks are *sparse*, meaning that the number of connections per person is limited, and does not increase very much even if the network grows. Figure 1(a) shows an interesting example: the network of retweets about Project X in Haren in 2012. A birthday invitation of a 16-year-old girl went viral in social media and resulted in a destructive riot. Dots are Twitter users and each tiny arrow (a directed edge) represents a retweet from one user to another. The figure shows this network in the morning before the riot. We see that the network is sparse, on average there are only 1.5 retweets per user. Over the night of the riot the network increased in size more than 10 times, but the average number of retweets remained small, it went only a little bit above two.
- *Connected*. Consider the network of railway stations connected by railroads, as the NS network in Figure 1(b). The vertices are the stations, and the (undirected) edges are the railway connections between them. A passenger can travel by train from any station to any other. The railroad network is *connected*. The Internet is another powerful example of a connected network. Internet is extremely complex and completely decentralized, yet, the data can be transferred across the planet from any Internet router to any other!
- *Scale-free*. In the Web graph, vertices are the webpages, and (directed) edges are the hyperlinks. By clicking on a hyperlink, we can go from one webpage to another. From how many other pages a typical webpage can be reached? In Figure 1(c) we show the average and the maximum of this number in the `.eu` domain of the Web graph in 2015. We see that on average, a webpage can be reached from 85.7 other webpages. However, the maximum is over $200\,000$ times larger than average! We say that such network is *scale-free*. This unusual term means that there is no such thing as a 'typical webpage'. The number of hyperlinks pointing to a webpage can have very different scales — from a few, to hundreds, thousands, and millions.
- *Triangles*. How do people in social networks usually meet? Often I know friends of my friends, they can be my friends, too. Groups of friends create clusters with many *triangles*, such as in Figure 1(d): if Anna is a friend of Boris and Cecile, then it is not surprising that Boris and Cecile know each other as well. Having many triangles is another typical structural property of complex networks.

Now that we have discussed the common properties of complex networks, let's see how to model them mathematically using random graphs.

Modeling sparse networks with Erdős–Rényi random graphs

Erdős–Rényi random graph

Recall that in a random graph, vertices are fixed, and edges are placed at random. Suppose we have n vertices. Possibly the simplest rule for placing random edges is to connect any pair of vertices i and j by an edge with the same probability p, independently of all other edges. This model is called an *Erdős–Rényi (E-R) random graph* after the founders of the theory of random graphs Paul Erdős and Alfréd Rényi [7]. We denote this random graph by $\mathrm{ER}_n(p)$. Figure 2 is a small example of a social network as an Erdős–Rényi random graph. The solid lines are existing connections and the dashed lines are possible but not realized connections.

You may see how E-R graphs look like for different n and p on the *Network pages* website [16]. The simulator also includes parameter $\lambda = np$, the significance of which I will explain in the next section. Look at the figures produced by the simulator: do the graphs look as you expected?

Of course, in reality, social connections are formed not at all as in the E-R model. This simple model ignores complex dependencies, heterogeneity between people, polarization, et cetera. Moreover, it is not even supposed to include all these things because Erdős and Rényi invented this model for completely different purposes: to solve difficult problems in combinatorics! This ingenious idea yielded a by now well established approach in Graph Theory, the so-called 'probabilistic method' [1]. However, as it often happens in mathematics, a model invented for solving one problem, becomes very useful for solving completely different problems in

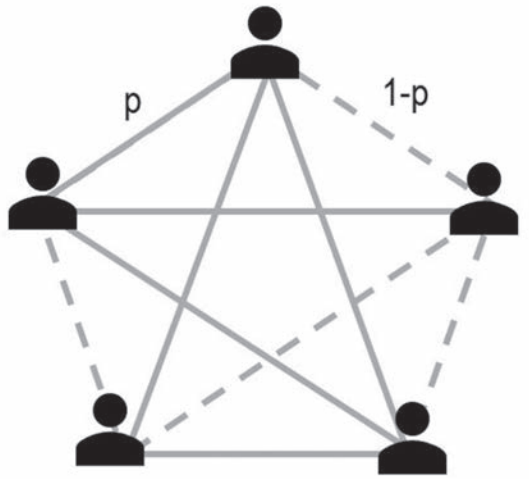


Figure 2 Social network is modeled as an Erdős–Rényi (E-R) random graph: each friendship exists with probability p.

the unknown future. With arrival of abundant network data, and explosive growth of Network Science, the E-R model is a fundamental cornerstone in studying real-life networks, because deep understanding of the simplest possible mathematical models is crucial for explaining real-world phenomena. In this section the phenomenon in question will be sparsity of real-life networks.

Sparse networks

The notion of a *sparse* network has to do with the number of edges and the number of vertices. Loosely speaking, in a sparse network, there are not too many edges per vertex on average. But how many is 'not too many'? Suppose we have a networks of 1000 vertices. If there are on average, say, 2 edges per vertex, then we will probably all agree that this network is sparse. If there are on average 500 edges per vertex, then, we will probably say that this network is not sparse, because 500 is a lot compared to 1000. But what about 20 edges per vertex? Or 50? Or 70? When does a network 'start' being sparse?

We could give a mathematical answer to this by choosing a threshold of some sort. However, the Theory of Random Graphs takes a different approach. We say that a network is *sparse* when the average number of edges per vertex remains bounded when the number of vertices n grows to infinity. In other words, sparsity is a so-called *asymptotic notion*, it is defined only in the limit, when $n \to \infty$.

Sparse Erdős–Rényi random graphs

Let us now see how sparsity works out in the E-R random graph, $\mathrm{ER}_n(p)$.

In any undirected graph of n vertices, there are in total $\frac{n(n-1)}{2}$ possible edges. Since in the $\mathrm{ER}_n(p)$ model each edge exists with probability p, there are on average

$$\frac{n(n-1)}{2} \cdot p$$

edges. To get the average number of edges per vertex, we divide the last expression by n. The result is:

$$\frac{(n-1)}{2} \cdot p \quad \text{edges per vertex on average.} \tag{1}$$

Intuitively, the network is sparse when p is small because then there are less edges per vertex. But how small should p be? Suppose we choose a very small *fixed* p, say, $p = 0.0001$. Then the average number of edges per vertex is

$$\frac{(n-1)}{2} \cdot 0.0001.$$

The most important observation about this expression is that it is *not* bounded, it grows to infinity when n grows to infinity, so our asymptotic definition of sparsity is violated.

Since even a very small fixed p doesn't give us a sparse network, we need a different approach. Here we use a powerful technique called *parametrization*. The idea is to not view p as a given constant in $(0,1)$ but make it a function of n, so $p = p(n)$. Why is it useful? Because, then we can choose the function $p(n)$ in such a way that it will 'compensate' for the growth of n in expression (1) for the average number of edges per vertex.

In order to create a sparse network, we choose

$$p := p(n) = \frac{\lambda}{n}, \quad \lambda > 0.$$

This is, by the way, the same λ that you saw in the simulator before. If we substitute this in (1), the average number of edges per vertex becomes

$$\frac{n-1}{2} \cdot \frac{\lambda}{n} < \frac{\lambda}{2}.$$

Since $\lambda > 0$ is fixed, the average number of edges per vertex is bounded by a fixed number $\frac{\lambda}{2}$, hence the network is indeed sparse.

We denote a sparse E-R random graph by $\mathrm{ER}_n(\lambda/n)$. It is important to notice that the parametrization does not change the number of parameters of the model. The original model $\mathrm{ER}_n(p)$, has two parameters: n and p. The sparse model $\mathrm{ER}_n(\lambda/n)$, too, has two parameters: n and λ. The difference is that in the sparse model, $p = \lambda/n$ decreases with n, and this gives us a sparse network for any fixed λ. This is how a very simple model allows us to describe mathematically a very essential property of real networks — their sparsity, and remains flexible enough to create sparse graphs with less or more edges, by varying λ.

I want to close this section with a small remark from my experience in teaching random graphs. When students learn about the Erdős–Rényi model, they easily remember that all edges have equal probability p, while the independence of edges is often overlooked. The independence of edges indeed doesn't play any role in modeling sparsity (because sparsity concerns only the average number of edges per vertex), but it will take a central stage later in this article, explicitly so when we will discuss the number of triangles.

Almost sure guarantee that a random graph is connected

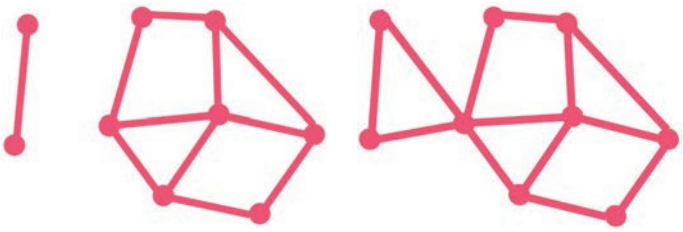
A graph is *connected* if there is a route along the edges from any vertex of the graph to any other. Graphs in Figure 1(b) and Figure 3(b) are examples of such connected graphs.

As we already know, many important real-life networks are connected. For example, in the Internet, information can be delivered from any router to any other. But when we model a network as a random graph, we place edges at random. Will such graph be connected? In this section we will solve this beautiful mathematical puzzle for the Erdős–Rényi random graph.

Isolated vertices in E-R random graph

It is easy to realize that if there are isolated vertices (vertices with no edges attached to them), then the graph is definitely disconnected. Of course, there are other ways to disconnect the graph: there can be isolated islands of $2, 3, \ldots$ vertices. Yet, as we will see, isolated vertices are a good start. Moreover, surprisingly, in the E-R model, isolated vertices define whether the graph is connected or not.

To begin with, let us compute the average number of isolated vertices. We denote this number by N_0; the sub-index zero reflects



(a) Disconnected graph. (b) Connected graph.

Figure 3 In a connected graph, all vertices are connected by a path of edges

that isolated vertices have zero edges attached to them. There are n vertices, each of which can be isolated with some probability, and we need to compute this probability first. There are $n-1$ possible edges, each edge is absent with probability $1-p$, and the edges are independent. Then, the probability that all $n-1$ possible edges of a vertex are absent, is $(1-p)^{n-1}$. The result is:

$$\mathbb{P}(\text{a vertex is isolated}) = (1-p)^{n-1},$$

where $\mathbb{P}(\cdot)$ denotes the probability of an event in the brackets. There are n vertices in total, and the average number of isolated vertices is

$$N_0 = n \cdot (1-p)^{n-1} \tag{2}$$

due to the very useful and somewhat counter-intuitive result in probability, called *linearity of expectations*. The linearity of expectations tells us that when we add random variables, the average of the sum is *always* the sum of averages, even if random variables are dependent. In this case, we can think of counting isolated vertices as adding random zeros and ones. We add 0 if the vertex is *not* isolated, and 1 if the vertex is isolated. These zeros and ones are *dependent*: for example, if we know that $n-1$ vertices are isolated, then for sure the n-th vertex is isolated as well (I leave it to the reader to explain why). Nevertheless, on average, each vertex adds $(1-p)^{n-1}$ to the sum, and the total average is $n \cdot (1-p)^{n-1}$ thanks to the linearity of expectations.

Sparse E-R random graphs are disconnected

Expression (2) is already sufficient to conclude that large sparse E-R graphs are very likely to be disconnected! Indeed, take a sparse E-R random graph $\mathrm{ER}_n(\lambda/n)$, where $p = \lambda/n$. Then (2) becomes

$$n\left(1-\frac{\lambda}{n}\right)^{n-1} \approx ne^{-\lambda}, \tag{3}$$

where the approximation is quite good for large n, and follows from the famous limit:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}. \tag{4}$$

This approximation (3) tells us that the number of isolated vertices in a sparse E-R random graph grows roughly linearly with n. When there are so many isolated vertices on average, it is very unlikely that there will be none of them in the graph. The exact mathematical result is as follows:

$$\lim_{n\to\infty}\mathbb{P}(ER_n(\lambda/n) \text{ is connected}) = 0.$$

Connectivity threshold

Let us now think of connectivity of the $\mathrm{ER}_n(p)$ model when p changes from 0 to 1. When $p=0$, there are no edges, so the graph is disconnected. When p grows, there will be more edges, and when $p=1$, all edges are present, so the graph is connected and complete. Then, there should be a range of values of p, for which the graph is likely to be connected. We already saw that $p=\lambda/n$, as in sparse E-R graphs, is not large enough to connect all vertices. For connectivity, we need larger p, but it turns out that 'a little bit' larger p is already sufficient. The transition from disconnected graphs to connected ones occurs when we choose a slightly different parametrization:

$$p = \frac{a\ln(n)}{n}, \quad \text{for some } a>0. \tag{5}$$

Indeed, let us substitute this p in (2). Then we get

$$\begin{aligned} N_0 &= n\cdot(1-p)^{n-1} = n\cdot\left(1-\frac{a\ln(n)}{n}\right)^{n-1} \\ &\approx n\cdot e^{-a\ln(n)} = n\cdot n^{-a} = n^{1-a}, \end{aligned} \tag{6}$$

where the approximation holds analogously as in (4). We see that the average number of isolated vertices is approximately n^{1-a}. Whether this number is big or small when n goes to infinity, depends on a. When $a<1$, we have that $1-a$ is positive, so n^{1-a} grows to infinity. In this case, a large network has many isolated vertices, the graph is disconnected. However, if $a>1$, then $1-a$ is negative, and n^{1-a} is vanishing. In this case, on average, there are no isolated vertices, and the probability that a graph is connected, converges to 1 as n goes to infinity. In probability theory we say that in this case the graph is connected *almost surely*. For large finite graphs of size n, this means that the corresponding probability is smaller than 1 but very close to 1, and becomes even closer when n grows. The exact mathematical result is:

$$\text{If } a<1, \text{ then } \lim_{n\to\infty}\mathbb{P}\left(\mathrm{ER}_n\left(\frac{a\ln(n)}{n}\right) \text{ is connected}\right) = 0.$$

$$\text{If } a>1, \text{ then } \lim_{n\to\infty}\mathbb{P}\left(\mathrm{ER}_n\left(\frac{a\ln(n)}{n}\right) \text{ is connected}\right) = 1.$$

The abrupt transition from disconnected to connected graphs is a beautiful example of a phenomenon known as *phase transition*. Like water turns into ice at zero degrees Celsius, isolated vertices vanish when connection probability p exceeds the so-called *connectivity threshold*

$$p_{\text{connectivity}} = \frac{\ln(n)}{n}.$$

We see this in Figure 4(a) and 4(b), a graph here has 100 vertices, and $p_{\text{connectivity}} \approx 0.046$.

Disappearance of disconnected islands

One may wonder, what about other ways of making a graph disconnected? Can there be islands of $2, 3, \ldots$ vertices? To answer this question, we consider *isolated pairs*. An example of a graph with an isolated pair is shown above in Figure 3(a).

Denote the average number of isolated pairs by N_2. Let us now compute this number. There are $\frac{n(n-1)}{2}$ pairs in total. A pair is isolated, if the two vertices in the pair are connected to each other, this happens with probability p. Furthermore, each of the two vertices in the pair must be isolated from the other $n-2$ vertices in the graph. The probability of this for each vertex is $(1-p)^{n-2}$, and for two vertices simultaneously, is $[(1-p)^{n-2}]^2$. Altogether, we get

$$N_2 = \frac{n(n-1)}{2}\cdot p\cdot\left[(1-p)^{n-2}\right]^2.$$

If we now substitute $p = \frac{a\ln(n)}{n}$ as in (5), we get

$$\begin{aligned} N_2 &= \frac{n(n-1)}{2}\cdot\frac{a\ln(n)}{n}\cdot\left[\left(1-\frac{a\ln(n)}{n}\right)^{n-2}\right]^2 \\ &\approx \frac{n^2\left(1-\frac{1}{n}\right)}{2}\cdot\frac{a\ln(n)}{n}\cdot n^{-2a} \\ &\approx \frac{1}{2}\cdot\frac{a\ln(n)}{n}\cdot n^{2(1-a)}. \end{aligned}$$

Compare this to the expression (2) for N_0. If $a>1$ then $n^{2(1-a)}$ goes to zero even faster than n^{1-a}. In addition, $\frac{a\ln(n)}{n}$ goes to zero,

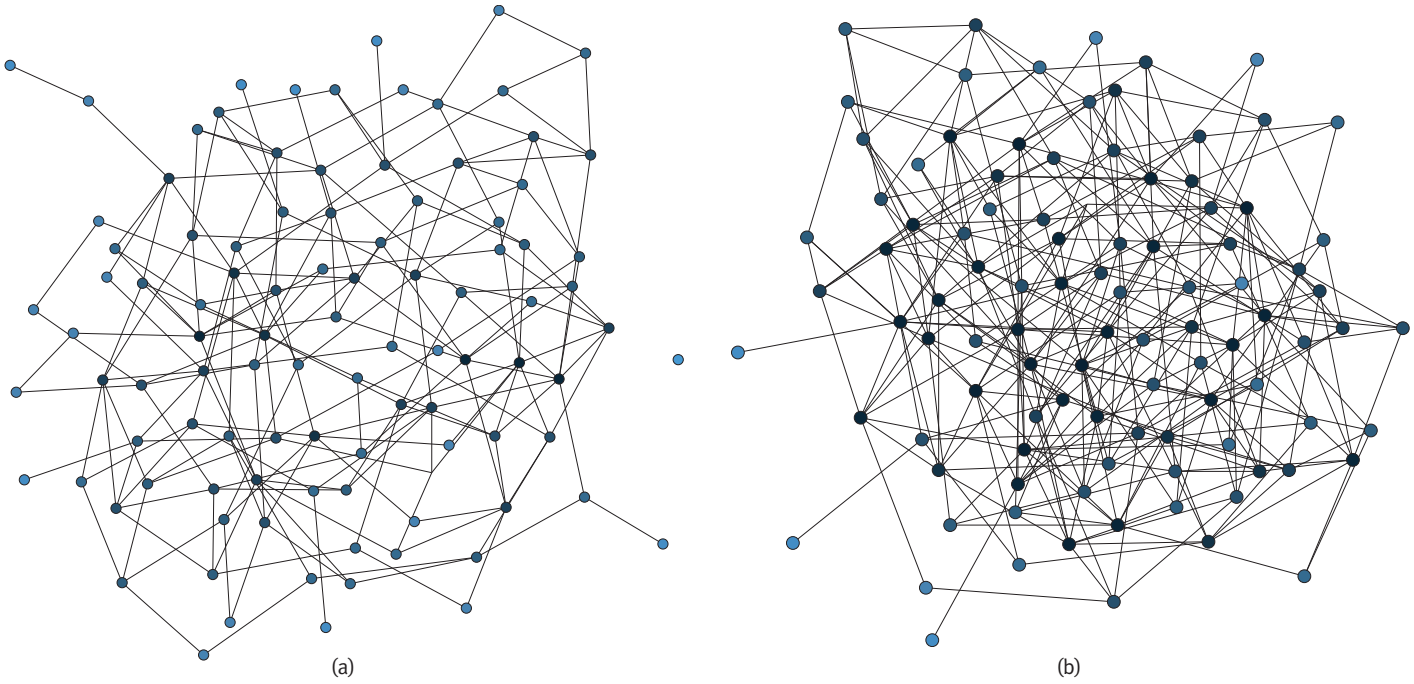


Figure 4 (a) E-R random graph with $n = 100$, $p = 0.04$. The graph is disconnected. (b) E-R random graph with $n = 100$, $p = 0.05$. The graph is connected.

as well. Hence, we conclude that above the connectivity threshold, isolated pairs are even less likely than isolated vertices.

In fact, we can mentally visualize how the graph 'gets frozen' when p increases. You can observe how this happens in the simulator on Network Pages [16]. The simulator adds edges one by one, and this is equivalent to gradually increasing p. The process starts with completely disconnected vertices. Then, p and the number of edges grow, and connected islands appear. As the graph becomes denser, large disconnected islands start merging together. Gradually we do not see anymore islands of 5, 4, 3, 2 vertices. Finally, when p exceeds $p_{\text{connectivity}}$, the last isolated vertices disappear, and the graph becomes connected. And now you also know why the process evolves this way, and you can explain it with just a couple of formulas!

Mathematics of scale-free networks

Real-life networks are often *scale-free*} In this section we want to express this scale-free property in mathematical terms, and get some insight in how this phenomenon emerges.

It will be convenient to introduce a bit of terminology. The number of edges attached to a vertex is called the *degree* of this vertex. For instance, in Figure 5, the degree of vertex 2 is equal to 3.

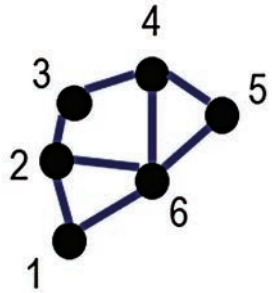


Figure 5 A degree of a vertex is the number of edges attached to it. For example, the degree of vertex 2 is 3.

The scale-free property is nothing else but a large variability of degrees. In real-life networks, there is often a relatively small but notable group of vertices with extremely high degrees, much higher than average. This is the case, for instance, in the World Wide Web, as we saw in Figure 1(c).

Sparse E-R graphs are not scale-free

The title of this subsection already gives away the result, but we are mainly interested in why sparse E-R graphs are not scale-free. An intuitive explanation is that the scale-free property also means large *heterogeneity* of vertices: some vertices attract many more connections than others. In E-R random graph this is clearly not the case: all edges have the same probability, and in that sense, all vertices are symmetric. Yet, since the edges are placed at random, it is also true that degrees of the vertices are random, so there will be some variability among them. The main message of this section is that this variability, solely due to randomness of the edges, is insufficient to create a scale-free network.

Let us see how this looks in formulas. Every vertex has $n-1$ possible edges, each edge exists with probability λ/n, independently of anything else. Then the degree of a vertex follows the well-known *Binomial distribution*, and we can write down the probability that the degree of a vertex is k:

$$P(\text{degree of a vertex is } k) = \frac{(n-1)!}{(n-1-k)!k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-1-k}, \quad k = 0,1,\ldots,n-1. \tag{7}$$

Now, if we take the limit $n \to \infty$, this converges to

$$\lim_{n\to\infty} P(\text{degree of a vertex is } k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0,1,\ldots. \tag{8}$$

I leave it to the reader to derive this limit. A hint: the limit (4) will be useful again. In particular, the limiting probability that a degree is zero (obtained by substituting $k = 0$), is $e^{-\lambda}$. This is of course

the same expression as the probability that a vertex is isolated, that we used in (3). The probability distribution in (8) is called the *Poisson* distribution. The parameter λ is the mean, or, the average of the Poisson distribution. This is consistent with the sparse E-R model because

$$\text{average degree in } \mathrm{ER}_n(\lambda/n) = (n-1)\cdot\frac{\lambda}{n} \approx \lambda.$$

Again, the approximation is accurate when n is large.

We can now use the Poisson formula to check the scale-free property. For instance, assume that the average degree is $\lambda = 100$ (similar to the data in Figure 1(c)), and let us compute the probability that a vertex has degree $k = 1000$, a modest 10-fold of the average, much smaller than typical large degrees in the real data. Substituting the numbers in (8) we get

$$\lim_{n\to\infty} P(\text{degree of a vertex is } 1000 \mid \lambda = 100) = \frac{100^{1000}}{1000!}e^{-100}. \quad (9)$$

Although 100^{1000} is impressive, the $1000!$ in the denominator is a humongous number. Expression (9) has a stunning 610 zeros after the decimal point! Moreover, the probability that degree is *at least* 1000 is not much larger:

$$\begin{aligned}
&\sum_{k=1000}^{\infty}\frac{100^k}{k!}e^{-100}\\
&= \frac{100^{1000}}{1000!}e^{-100}\left(1+\sum_{k=1}^{\infty}\frac{100^k}{(1000+k)\cdot(1000+k-1)\cdots 1001}\right)\\
&< \frac{100^{1000}}{1000!}e^{-100}\left(1+\sum_{k=1}^{\infty}\left(\frac{1}{10}\right)^k\right)\\
&= \frac{100^{1000}}{1000!}e^{-100}\cdot\left(1+\frac{1}{9}\right).
\end{aligned}$$

which is the same expression as in (9), only multiplied by ≈ 1.11, so we basically get the same answer. This is because the Poisson distribution is known to stay very close to its average. It is fundamentally not scale-free!

Power laws

A model that is able to capture the scale-free phenomenon, is the so-called *power law*. In this model the probability that a vertex has degree k is approximately proportional to a negative power of k, thus the name 'power law'. For the time being, we will write this as

$$\begin{aligned}
&P(\text{degree of a vertex is } k) \approx C\cdot k^{-\tau},\\
&\text{where } \tau > 2, C > 0, k > k_{\min} > 0.
\end{aligned} \quad (10)$$

Clearly smaller τ means higher probability of k. In particular, even very large k are quite likely to occur when τ is small. However, we need τ to be greater than 2 because otherwise large values are so likely that the average degree is infinite, and therefore the network is not sparse anymore. I encourage you to check this using the formula for the average, or, mean degree

$$\text{average degree} = \sum_{k=k_{\min}}^{\infty} k\cdot P(\text{degree of a vertex is } k).$$

(*Hint:* Think of harmonic series.)

Let us now verify that the power laws indeed model the scale-free phenomenon. Take $\tau = 2.5$, $k_{\min} = 33$, $C = 288.67$. Then the average degree is approximately again 100, but now we get

$$\begin{aligned}
P(\text{degree of a vertex is } 1000) &= 288.67\cdot 1000^{-2.5}\\
&= 0.0000091287 \approx 10^{-5}.
\end{aligned}$$

This probability, about one in a hundred thousand, is not very small in a network of millions, even billions vertices such as the World Wide Web. And the probability that the degree is *at least* 1000, is even much bigger:

$$\begin{aligned}
\sum_{k=1000}^{\infty} 88.67\cdot k^{-2.5} &\approx 288.67\cdot\int_{1000}^{\infty} x^{-2.5}dx\\
&= 192.45\cdot 1000^{-1.5} \approx 0.006.
\end{aligned}$$

This is a sizable proportion of vertices especially in a very large network. These results are even not very surprising because, definitely, the $k^{-\tau}$ decreases much slower than $\lambda^k/k!$.

When a model correctly captures a phenomenon however, it does not mean that the model is accurate, and there can be other suitable models. After power laws were found for the first time in the paper by three brothers Faloutsos [8] on the connectivity of the Internet, and then, explosively, in most other types of networks [2], there has been a lot of discussion about whether this model really holds. An interesting milestone in this discussion was the 2019 paper 'Scale-free networks are rare' by Broido and Causet [5]. The authors' statistical analysis resulted in the conclusion that (10) holds in but a modest fraction of real-life networks. Almost immediate answer to these results was the paper by Voitalov, van der Hoorn, van der Hofstad and Krioukov [12], called 'Scale-free networks well done', with a playful hint to the steak analogy. This paper has reminded of the fact that the power law is a much broader notion than the rather stringent formula (10). Mathematically, power laws are defined more precisely as

$$P(\text{degree of a vertex is } k) = C(k)\cdot k^{-\tau}, \quad \text{where } \tau > 2,\ k > 0, \quad (11)$$

with function $C(k)$ being a so-called *slowly varying* function. The term 'slowly varying' means that $C(k)$ may grow or decrease not faster than any arbitrarily small power of k. Formally, this can be written as

$$\lim_{k\to\infty}\frac{\log(C(k))}{\log(k)} = 0. \quad (12)$$

Of course, $C = \text{const}$ is a special case of a slowly varying function. The point in [12] is that (12) is an *asymptotic* expression, so for finite k we can never be sure whether (11) holds or not. The authors in [12] suggested their own method how to check whether the power law model is acceptable, and many real-life networks indeed do pass this test.

The debate on power laws in 2019 attracted a lot of attention and even hit the media [15]. I enjoyed this scientific discussion, but personally I do not find it extremely important whether power laws exist in reality or not. Mathematics describes the world with abstract objects, and in this case I find power laws a very suitable object because it does capture the essence of the phenomenon in hand: the high variability of degrees in real-life networks.

Preferential attachment

While we can observe, measure and model the scale-free phenomenon, the question remains, why are the networks scale-free? The *preferential attachment* model is an attempt to answer this question with a dynamic mathematical model of network growth.

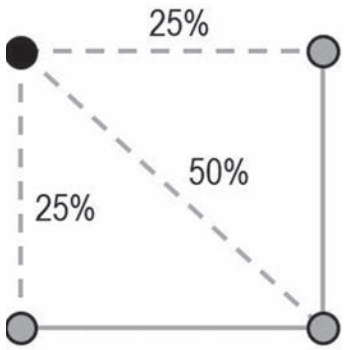


Figure 6 A new (black) vertex arrives in the network, and connects to existing vertices with probabilities proportional to their current degrees. Dash lines denote possible edges, and the number next to a dashed line is the probability of this edge.

This model formalizes the mechanism known as 'rich get richer', or 'popular get ever more popular'. It works roughly as follows. Assume a network starts with three vertices, like the three gray vertices in Figure 6. When the next vertex arrives (the black vertex in the figure), it can make one of three possible connections (dashed lines). For that, the new vertex uses the rich get richer mechanism: the probability to connect to any of the gray vertices is proportional to the number of connections they already have. In Figure 6, one of the gray vertices has two connections, and therefore it has a higher chance to get one more. Clearly, the more connections a vertex gets, the easier it is to attract even more new connections. This is the rich get richer mechanism at work! It is quite natural that with such dynamics, some vertices become extremely well connected.

Barabási and Albert in 1999 [2] put forward the preferential attachment model as a plausible mathematical explanation for the emergence of scale-free networks. This became a very influential work, paper [2] has more than 43K citations on Google Scholar, and counting! That said, it rarely happens that such an influential model and its brilliant application in a rising new area of science, has no history in earlier work. Indeed, the model is not new. In 1965, Derek de Solla Price suggested a very similar model for networks of scientific citations [10]: papers that are already frequently cited, tend to receive many new citations. Even earlier, in 1927, Yule proposed a similar model for the evolution of biological species [13]: large populations of species have large offspring. And even before that, in 1925, great mathematician George Pólya and his PhD student Florian Eggenberger published a paper about a mathematical model that is now called *Pólya's urn scheme*. The scheme works as follows. We have a number of green and red balls in an urn. We choose one ball at random, and then return it to the urn together with another ball of the same color. Figure 7 illustrates how it works.

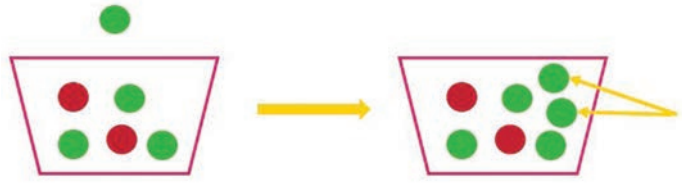


Figure 7 Pólya's urn scheme. Initially we have 4 green balls and 2 red balls. We randomly chose a green ball, and therefore we add another green ball to the urn. As a result, we have 2 red balls and 5 green balls. It is now even more likely to choose a green ball.

Here we have more green balls in an urn, so it is more likely to choose a green ball, as in the figure. After that, we add another green ball, so in the next round, it's even more likely to choose a green ball again. This is exactly the rich-get-richer mechanism. Pólya's urn is another example of a mathematical model that appeared before its applications. Mathematical techniques developed for Pólya's urn scheme, are often used to analyze preferential attachment dynamics in networks.

Emergence of power laws

Intuitively, it is clear that preferential attachment helps some vertices to get lots of connections, making power laws plausible. In this section we will go one step beyond intuition and see why it happens, mathematically. There are many ways to do this. In this article I use the argument in lines of [9, Section 8.3]. The way I derive it here is not strictly rigorous, we may call it a *heuristic* argument. However, it does give mathematical reasons why degrees in the preferential attachment model follow the power law distribution (10).

Suppose the network starts at time $t=0$ with vertex zero without edges. At time $t=0$, vertex 1 arrives with one edge and connects to vertex 0. After that, vertices appear in the network one by one, and we will number them $2,3,\ldots,t$, so t is our 'current time', and, including vertex 0, there are $t+1$ vertices at time t. Denote by $D_i(t)$ the average degree of vertex i at time $t \geq i$. With arrival of vertex t, the average degree of i in the previous step, $D_i(t-1)$, changes according to the preferential attachment rule. There are many versions of this rule, but we assume the simplest one: a new vertex makes only one connection, and attaches to one of the existing vertices with probability exactly proportionally to their degree. So, the probability that vertex t attaches to vertex i is,

$$\frac{D_i(t-1)}{D_0(t-1)+D_1(t-1)+\cdots+D_{t-1}(t-1)} = \frac{D_i(t-1)}{2(t-1)}, \tag{13}$$

where the denominator equals to $2(t-1)$ because in any graph all degrees sum up to twice the number of edges, and at time $t-1$ we have $t-1$ edges because each vertex $1,2,\ldots,t-1$ arrives with one edge.

Let us now look how the average degree of vertex i changes at time t:

$$\begin{aligned} D_i(t) &= D_i(t-1) + 1\cdot\frac{D_i(t-1)}{2(t-1)} \\ &= D_i(t-1)\cdot\left(1+\frac{1}{2(t-1)}\right) \\ &= D_i(t-1)\cdot\frac{2(t-1)+1}{2(t-1)} \qquad (14) \\ &= D_i(t-1)\cdot\frac{(t-1)+1/2}{t-1}. \qquad (15) \end{aligned}$$

We can now iterate (15) back in time till $D_i(i)=1$:

$$\begin{aligned} D_i(t) &= D_i(t-1)\cdot\frac{(t-1)+1/2}{t-1} \\ &= D_i(t-2)\cdot\frac{(t-1)+1/2}{t-1}\cdot\frac{(t-2)+1/2}{t-2} \\ &= \cdots = D_i(i)\cdot\frac{(t-1)+1/2}{t-1}\cdot\frac{(t-2)+1/2}{t-2}\cdots\frac{i+1/2}{i} \\ &= 1\cdot\frac{\Gamma(t+1/2)}{\Gamma(i+1/2)}\cdot\frac{\Gamma(i)}{\Gamma(t)}. \end{aligned} \tag{16}$$

We can now use the property of the gamma function $\Gamma(\cdot)$:

$$\frac{\Gamma(x+\alpha)}{\Gamma(x)} \approx x^{\alpha} \quad \text{for large enough } x \text{ and fixed } \alpha.$$

If t and i in (16) are large, then we have an approximate expression for the average degree of vertex i at time t:

$$D_i(t) \approx \frac{t^{1/2}}{i^{1/2}} \quad \text{when } t \text{ and } i \text{ are large enough.}$$

We already see the power $1/2$ appearing, it remains only to translate this to the power law (10). This is where our derivation becomes 'heuristic'. Remember that $D_i(t)$ is the average degree of vertex i, but let us pretend that this is the exact degree (even if it is not necessarily an integer number). Then the fraction of vertices with degree k, among the total of $t+1$ vertices, is

$$\frac{1}{t+1} \cdot \big[\text{the range of } i \text{ such that } D_i(t) \in (k-0.5, k+0.5)\big].$$

Now we need to find, which i are in the required range:

$$k-0.5 < \frac{t^{1/2}}{i^{1/2}} < k+0.5.$$

Solving for i, we get:

$$\frac{t}{(k+0.5)^2} < i < \frac{t}{(k-0.5)^2}.$$

The length of the interval that contains i above, is:

$$\begin{aligned}\frac{t}{(k-0.5)^2} - \frac{t}{(k+0.5)^2} &= t \cdot \frac{1}{(k-0.5)^2} \cdot \left(\frac{(k+0.5)^2 - (k-0.5)^2}{(k+0.5)^2} \right) \\ &= t \cdot \frac{2k}{(k-0.5)^2 (k+0.5)^2} \\ &\approx 2t \cdot k^{-3}, \quad \text{when } k \text{ is large enough.}\end{aligned}$$

It remains to divide the last expression by $t+1$:

$$2 \cdot \frac{t}{t+1} \cdot k^{-3} \approx 2k^{-3},$$

and we have obtained a power law, as in (10), with $C=2$ and $\tau=3$.

We can vary τ, by adjusting the model a little bit. For instance, the attachment probability to vertex i is proportional to

$$\big[\text{degree of } i\big] + a$$

for some $a > -1$. Interestingly, this changes τ, we get $\tau = 3+a$. We may also allow vertices to arrive with $m>1$ edges, then we must take $a > -m$, and the power law exponent is $\tau = 3 + a/m$.

Finally, I want to notice that the heuristic argument above does capture the essence because the degrees in the preferential attachment model tend to stay close to their averages. Making this statement precise, however, is not easy. It requires some work and involves several advanced probabilistic techniques that have been developed only recently, starting with paper [4] that presented such rigorous analysis for the first time in 2001.

Mathematical model for triangles

Real-life networks have many triangles, as illustrated in Figure 1(d): if Anna knows Boris and Cecile, then Boris and Cecile likely know each other as well. This phenomenon is called *triangle closure*. In this section we will talk about how to capture triangle closure in a mathematical model.

Sparse Erdős–Rényi random graphs don't have many triangles

Denote by Δ_n the average number of triangles in a sparse E-R random graph, $\mathrm{ER}_n(\lambda/n)$. We will now find a formula for Δ_n to see how large this number can be. First of all, the number of possible triplets of vertices in a graph is:

$$\binom{n}{3} = \frac{n(n-1)(n-2)}{6}.$$

Next, a triplet of vertices i,j,k forms a triangle if and only if all three edges ij, jk, ki exist in the graph, as in Figure 8.

The probability of this event is

$$\left(\frac{\lambda}{n}\right)^3 = \frac{\lambda^3}{n^3}. \tag{17}$$

Then, the average number of triangles Δ_n is the fraction $\frac{\lambda^3}{n^3}$ of all $\frac{n(n-1)(n-2)}{6}$ possible triangles:

$$\Delta_n = \frac{n(n-1)(n-2)}{6}\frac{\lambda^3}{n^3} = \frac{n(n-1)(n-2)}{n^3}\frac{\lambda^3}{6} \approx \frac{\lambda^3}{6}. \tag{18}$$

For instance, if $\lambda = 9$, there are on average 121.5 triangles. Maybe, 121.5 is not a very small number, but the main implication of formula (18) is that this number does not grow when n grows. Even if the number of vertices n is very large, the average number of triangles is stuck at 121.5. In a networks of several million vertices, having only 121.5 triangles on average doesn't conform at all to the triangle closure phenomenon.

Before going any further, let us do what I always ask my students to do after answering the question: analyze the answer.

First of all, clearly, we will never see 121.5 triangles, this number is the *average*. In this context, it means the following. If we generate many graphs of size n, count the number of triangles in each of them, and then compute the average, then we will get something close to 121.5. Of course, since the edges are placed at random, the actual number of triangles will be different in every realization of $\mathrm{ER}_n(\lambda/n)$. In fact, one can prove that the number of triangles in an $\mathrm{ER}_n(\lambda/n)$ random graph converges to the Poisson distribution with parameter $\frac{\lambda^3}{6}$ when n goes to infinity.

Second, why exactly don't sparse E-R random graphs have many triangles? Which model assumptions are responsible for this result? There are in fact only two assumptions: all edges have the same probability $\frac{\lambda}{n}$, and are independent. Of course, if we increase the edge probability, then the number of triangles increases as well. Nevertheless, when the edge probability is $\frac{\lambda}{n}$, the number of triangles is bounded even for very large λ. To let the number of triangles

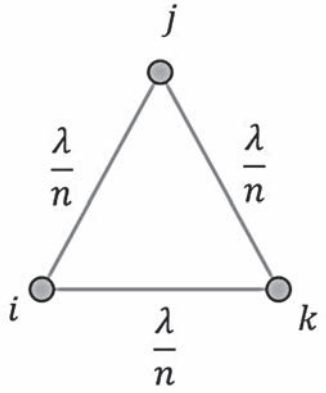


Figure 8 Triangle ijk in $\mathrm{ER}_n(\lambda/n)$. The probability of each edge is λ/n.

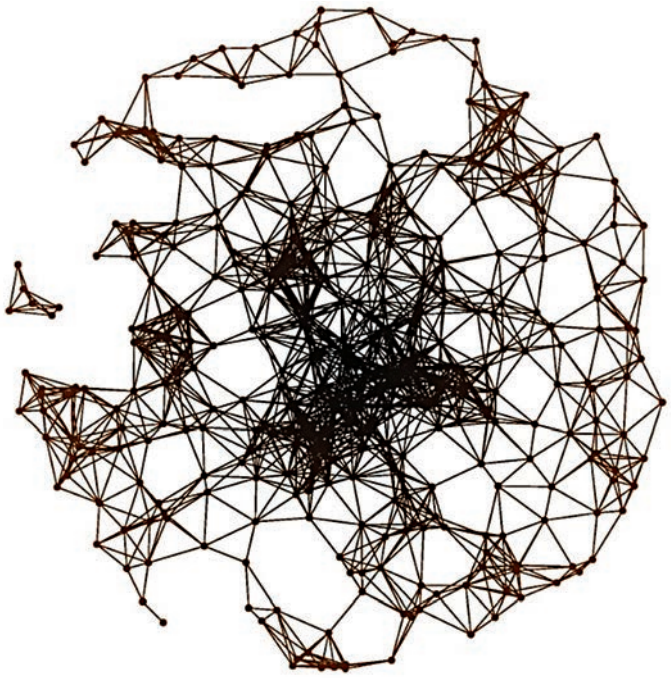


Illustration: Pim van der Hoorn

Figure 9 A geometric random graph.

grow with network size, we need a different parametrization. For instance, if $p = p(n) = \frac{\lambda}{\sqrt{n}}$, then $\Delta_n \approx \frac{\lambda^3\sqrt{n^3}}{6}$ (I leave it to the reader to verify this themselves). But then, the average degree $\approx \frac{\lambda n^{1/2}}{2}$, and the graph isn't sparse anymore. In a sparse graph, the main culprit that fails the triangle closure is the *independence* of edges. Indeed, triangle closure tell us that existence of two edges makes the third edge more likely, and this is simply not the case in the E-R model. To solve this, we could explicitly introduce complicated dependencies between edges. While such models exist, here I will explain a more elegant and now common approach: introducing geometry.

Geometric random graphs

A fundamental way to create triangle closure is to place vertices in a multi-dimensional space, and let vertices connect when they are close to each other. Sometimes such geometry is already there, for example, in the network of airports connected by direct flights, each airport has a location, and there are many quite short (so, local) direct flights. On a slightly more abstract level, a 'location' can be defined through the properties of the vertex, for example, we may 'locate' people in a multi-dimensional space depending on their age and interests. Then, in a social network, similar people are more likely to be friends. This idea is very natural and appealing. In fact, many network scientists believe that only geometric random graphs can give us realistic models for complex networks.

Geometric networks naturally have many triangles, as we can see in the example in Figure 9. This is easy to explain intuitively: if vertices j and k are connected to vertex i, then j and k are likely to be close to i in the space, and thus they are likely to be close to each other as well, so edge jk has a high chance to appear.

In this article we consider the simplest sparse geometric random graphs and prove that they indeed have many triangles. In this simple model, we place n vertices in a unit square (with side 1) uniformly at random, and we connect two vertices if the distance between them is at most r. We call this model $\mathrm{Geom}_n(r)$. Figure 10(a) shows an example of such model with $n = 11$.

To start, let us compute the probability of an edge. Edge ij appears if vertex j happens to be in a circle of radius r around vertex i. Since j is placed randomly on the area 1, the probability of such edge is:

$$P(j \text{ is in the circle of radius } r \text{ from } i) = \frac{\pi r^2}{1} = \pi r^2. \qquad (19)$$

Of course, we assume that the area of a circle, πr^2, is smaller than 1, otherwise most vertices are connected, and the model is not very interesting.

Looking critically at formula (19), a sharp reader may notice that the formula does not work if i is closer than r to the boundary. This does not change the essence of the results but calculations become much messier when we have to take the boundary into account. This is why geometric models are often studied not on a square, but on a *torus*, where boundaries are merged together. I recommend a nice animation of how a square becomes a torus [14]. In words, on a torus, the boundaries are merged, and therefore vertices close to the boundary become close to each other. Figure 10(b) shows that in our small example, replacing a square by a torus results in one more, 'over the boundary' edge kl. We stick to the torus in this article so that all vertices are symmetric, and the probability (19) is correct for every edge ij regardless the location of i.

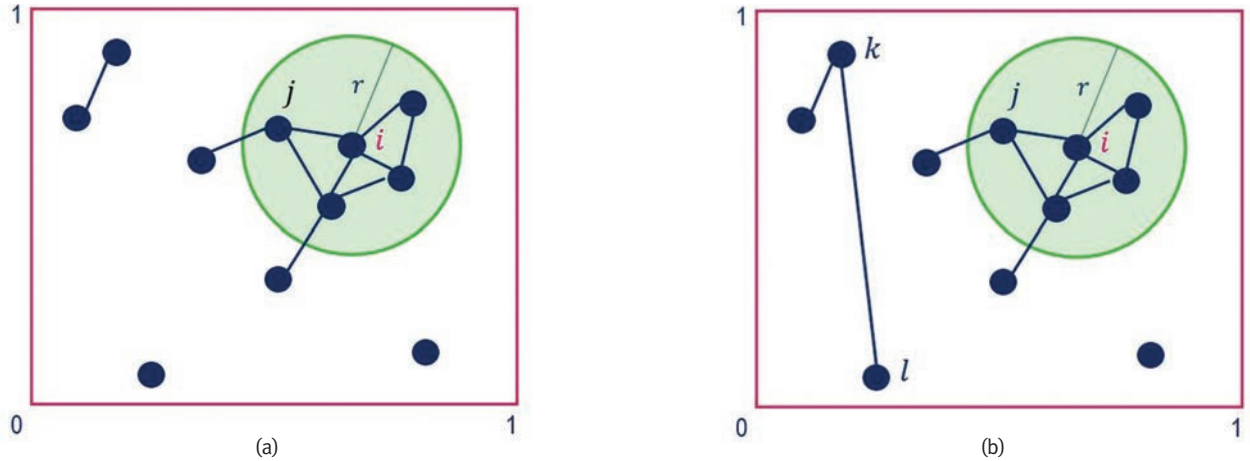


Figure 10 (a) A $\mathrm{Geom}_{11}(r)$ random graph on a unit square. Vertices within distance r from each other form an edge. (b) A $\mathrm{Geom}_{11}(r)$ random graph on a torus. The distance between k and l 'over the boundary' is less than r, therefore, on a torus, there is an edge kl.

Sparse geometric random graphs

To make the geometric random graph sparse, we again will use parametrization. As in the E-R model, every vertex has $n-1$ potential neighbors, and we have already computed the probability of connection in (19). Then the average degree of a vertex is

$$[\text{the average degree in } \mathrm{Geom}_n(r)] = (n-1)\cdot \pi r^2. \qquad (20)$$

Recall that the graph is sparse when its average degree doesn't grow with n. We can achieve this with parametrization

$$r = \frac{\mu}{\sqrt{n}}, \quad \text{where } \mu > 0.$$

Then the average degree in (20) becomes

$$(n-1)\pi\left(\frac{\mu}{\sqrt{n}}\right)^2 < \pi\mu^2.$$

We conclude that $\mathrm{Geom}_n(\mu/\sqrt{n})$ graph is sparse. And now it remains to prove that this sparse graph has many triangles.

Many triangles in sparse geometric random graphs

Calculations with circles are a little bit messy, that's why we will derive a lower bound for the average number of triangles using squares instead of circles. A lower bound is sufficient for our purposes because if the lower bound is large, then the actual number of triangles is even larger.

To get our lower bound, we assume that vertex i can be at any location, and we draw a square with diagonal r (and side $\frac{r}{\sqrt{2}}$) with i in a center, as in Figure 11. All vertices in this square are connected to i, but also to each other because they are at the distance less than r. The probability that j is in the square with center at i, is

$$\left(\frac{r}{\sqrt{2}}\right)^2,$$

and same holds for k. Since locations of j and k are independent from each other, the probability that both j and k are in the square is

$$\mathbb{P}(j \text{ and } k \text{ are in the square with side } \frac{r}{\sqrt{2}} \text{ and center at } i)$$
$$= \left(\frac{r}{\sqrt{2}}\right)^2\left(\frac{r}{\sqrt{2}}\right)^2 = \frac{r^4}{4}. \qquad (21)$$

In a sparse geometric graph, with $r = \frac{\mu}{\sqrt{n}}$, this probability is

$$\frac{\mu^4}{4n^2}.$$

The actual probability of triangle ijk is larger than that because there are many other configurations for such triangle to appear. All three vertices being in one square is only one of the possibilities.

Now recall that in any graph of size n there are $\frac{n(n-1)(n-2)}{6}$ possible triangles. Then, on average, the number of triangles in a sparse geometric graph is at least

$$\frac{n(n-1)(n-2)}{6}\cdot\frac{\mu^4}{4n^2} \approx \frac{\mu^4}{24}n.$$

The actual average number of triangles Δ_n is even larger because we computed a lower bound. However, the order of magnitude is correct because we could derive Δ_n similarly: fixing the location of i and computing the areas in which j and k should be in order to form a triangle. These areas are larger than the area of our square, but they all are proportional to r^2.

We conclude that Δ_n in $\mathrm{Geom}_n(\mu/\sqrt{n})$ grows linearly in n. This matches our intuition about, say, social networks: every new person in a network makes several friends and by that also participates in several triangles.

Looking at the formulas we can exactly pinpoint the place where sparse $\mathrm{ER}_n(\lambda/n)$ random graphs differ from the sparse $\mathrm{Geom}_n(\mu/\sqrt{n})$ random graphs. In $\mathrm{ER}_n(\lambda/n)$, in formula (17) we multiplied the probabilities of all three edges because the edges are independent. But in $\mathrm{Geom}_n(\mu/\sqrt{n})$, in formula (21), we multiplied probabilities only of two edges ij and ik, while edge jk appeared automatically, because j and k are so close to i, that they are automatically close to each other. This is exactly the mathematical representation of triangle closure.

What we learned and what's next

The three models and the four properties

Figure 12 schematically shows the four properties of real-life networks and how we can model them with random graphs. The Erdős–Rényi random graph is the simplest model where all edges have the same probability and are independent. This is already

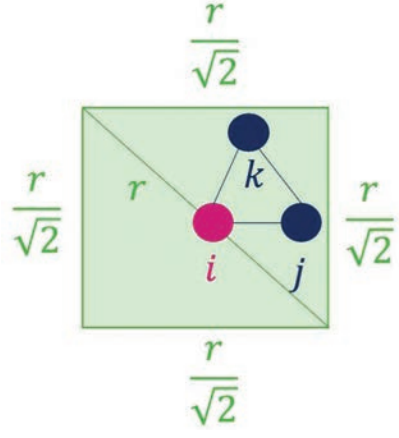


Figure 11 Vertex i is in the center of a square with diagonal r. All vertices inside this square are connected to each other.

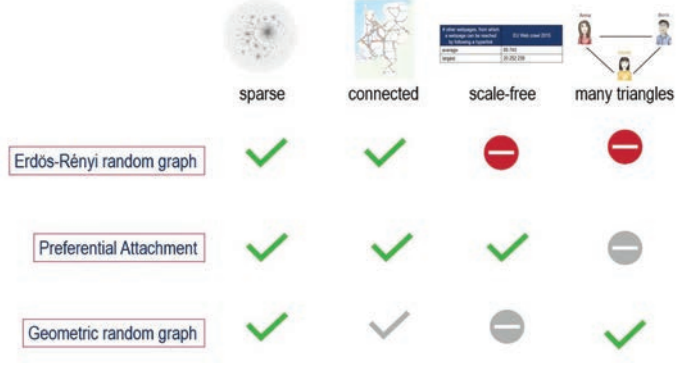


Figure 12 Mathematical models for the properties of real-life networks. Green check mark: the property holds, and we discussed it in the article. Red stop sign: the property does not hold and we discussed it in this article. Gray check mark: the property holds, but we did not discuss it in th article. Gray stop sign: the property does not hold, but we did not discuss it in the article.

sufficient to model *sparse* networks because sparseness concerns only the average number of edges per vertex. And if we choose the probability of an edge above the critical threshold, then the random graph is also *connected*. This model however is very homogeneous: all vertices and edges have exactly the same properties, and the differences between them are only due to random fluctuations. Turns out, this is insufficient to create a *scale-free* network. Also, there are not many *triangles* because the model doesn't have a mechanism to enforce triangle closure.

In the Preferential Attachment model vertices appear one by one. Since every vertex comes with m connections, the Preferential Attachment random graph is *sparse*. It is also *connected* because new vertices attach to the existing connected graph. But, importantly, the vertices are not homogeneous: older vertices have higher chance to gain high degrees, and once they have a high degree, they are more likely to get new connections. This rich-get-richer mechanism results in the *scale-free* property, that is, huge differences between degrees of vertices. We didn't discuss the number of triangles in the Preferential Attachment model. The stop sign in gray indicates that the number of triangles is small because, similarly to the Erdős–Rényi random graph, the Preferential Attachment model does not stir towards triangle closure.

Finally, Geometric random graphs facilitate triangle closure by connecting vertices that are close to each other. Again, I included in gray the properties that we did not address in this article. If we make r large enough, geometric random graph is connected, this is indicated by the gray check mark. Further, sparse Geometric random graphs as in this article, are not scale free, this is indicated by the gray stop sign. Similarly to the Erdős–Rényi model, there is no heterogeneity between the vertices. When the graph is large, the degree of a vertex follows the $Poisson$ distribution with average degree $\pi\mu^2$.

Figure 12 may leave you with impression that it is not possible to model all four properties together. However, this is only because we considered very simple models. Recent state-of-the-art models are richer and closer to reality. For example, a very influential model that combined the scale-free property and triangles was introduced in 2010 [10]. It is in fact a geometric model, but the vertices are placed not in an Euclidean space, but in a hyperbolic space. The hyperbolic space is curved, so the size of a 'circle' around a vertex depends on the vertex's position, thus, there is high heterogeneity between vertices. Since then, many models appeared that include geometry and heterogeneity of vertices at the same time. There are also Preferential Attachment models that are either geometric or explicitly include high probability of triangle closure.

Other properties of real-life networks

There are many other very interesting and common structural properties of real-life networks that I did not mention in this article. For example, we often see *communities*. In social networks, communities can be defined by interests, language, or geography. Another famous property of real-life networks is the 'small world phenomenon': most pairs of vertices are connected by a short path of edges. In social networks, this phenomenon is also known as 'six degrees of separation' stating that "everybody on this planet is separated only by six other people" (John Guare). The communities and the small world phenomenon, of course, too, have been studied using random graphs. The research on random graphs and complex networks is happening right now, and as a mathematician I am very excited to be a part of this collective scientific effort.

I hope that this article gave you insight into mathematical tools we use to think about complex networks, and left you with motivation and curiosity to explore the endless opportunities of this quickly developing branch of modern mathematics.

References

1 Noga Alon and Joel H Spencer, *The Probabilistic Method*, Wiley, 2016.

2 Albert-László Barabasi and Reka Albert, Emergence of scaling in random networks, *Science* 286(5439) (1999), 509–512.

3 Paolo Boldi and Sebastiano Vigna, The WebGraph framework I: Compression techniques, in *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)*, *Manhattan*, ACM Press, 2004, pp. 595–601.

4 Béla Bollobás, Oliver Riordan, Joel Spencer and Gábor Tusnády, The degree sequence of a scale-free random graph process, *Random Structures & Algorithms* 18(3) (2001), 279–290.

5 Anna D. Broido and Aaron Clauset, Scale-free networks are rare, *Nature communications* 10(1) (2019), 1017.

6 Derek J. de Solla Price, The science of science, *Bulletin of the Atomic Scientists* 21(8) (1965), 2–8.

7 Paul Erdős and Alfréd Rényi, On random graphs. I, *Publ. Math. Debrecen* 6 (1959), 290–297.

8 Michalis Faloutsos, Petros Faloutsos and Christos Faloutsos, On power-law relationships of the internet topology, *ACM SIGCOMM Computer Communication Review* 29(4) (1999), 251–262.

9 Remco van der Hofstad, *Random Graphs and Complex Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 43, Cambridge University Press, 2016.

10 Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat and Marián Boguná, Hyperbolic geometry of complex networks, *Physical Review E* 82(3) (2010), 036106.

11 M. ten Thij, T. Ouboter, D. Worm, N. Litvak, H. van den Berg and S. Bhulai, Modelling of trends in twitter using retweet graph dynamics, in *Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings*, Springer, 2014, pp. 132–147.

12 Ivan Voitalov, Pim van der Hoorn, Remco van der Hofstad and Dmitri Krioukov. Scale-free networks well done, *Physical Review Research* 1(3) (2019), 033034.

13 George Udny Yule, A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis F.R.S., *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213(402–410) (1925), 21–87.

14 Folding a torus, https://youtu.be/nkszzWYCs8c. Date accessed: 4 February 2023.

15 Hoe machtig is het superknooppunt? https://www.nrc.nl/nieuws/2019/12/20/hoe-machtig-is-het-superknooppunt-a3984586, 20 December 2019. Date accessed: 26 January 2023.

16 The network pages – the math and algorithms that keep us connected, https://www.networkpages.nl/CustomMedia/Animations/RandomGraph/ERRG/AddoneEdgepATime.html. Date accessed: 24 January 2023.