

# ALBERT-LÁSZLÓ BARABÁSI

# NETWORK SCIENCE

## RANDOM NETWORKS



## ACKNOWLEDGEMENTS

MÁRTON PÓSFAI  
GABRIELE MUSELLA  
MAURO MARTINO  
ROBERTA SINATRA

SARAH MORRISON  
AMAL HUSSEINI  
PHILIPP HOEVEL

# INDEX

Introduction	1
The Random Network Model	2
Number of Links	3
Degree Distribution	4
Real Networks are Not Poisson	5
The Evolution of a Random Network	6
Real Networks are Supercritical	7
Small Worlds	8
Clustering Coefficient	9
Summary: Real Networks are Not Random	10
Homework	11
ADVANCED TOPICS 3.A	
Deriving the Poisson Distribution	12
ADVANCED TOPICS 3.B	
Maximum and Minimum Degrees	13
ADVANCED TOPICS 3.C	
Giant Component	14
ADVANCED TOPICS 3.D	
Component Sizes	15
ADVANCED TOPICS 3.E	
Fully Connected Regime	16
ADVANCED TOPICS 3.F	
Phase Transitions	17
ADVANCED TOPICS 3.G	
Small World Corrections	18
Bibliography	19
Figure 3.0 (cover image)	
Erdős Number	
The Hungarian mathematician Pál Erdős authored hundreds of research papers, many of them in collaboration with other mathematicians. His relentless collaborative approach to mathematics inspired the <i>Erdős Number</i> , which works like this: Erdős' Erdős number is 0. Erdős' coauthors have Erdős number 1. Those who have written a paper with someone with Erdős number 1 have Erdős number 2, and so on. If there is no chain of coauthorships connecting someone to Erdős, then that person's Erdős number is infinite. Many famous scientists have low Erdős numbers: Albert Einstein has Erdős Number 2 and Richard Feynman has 3. The image shows the collaborators of Pál Erdős, as drawn in 1970 by Ronald Graham, one of Erdős' close collaborators. As Erdős' fame rose, this image has achieved an iconic status.	
	
This work is licensed under a Creative Commons: CC BY-NC-SA 2.0. PDF V32, 08.09.2014	

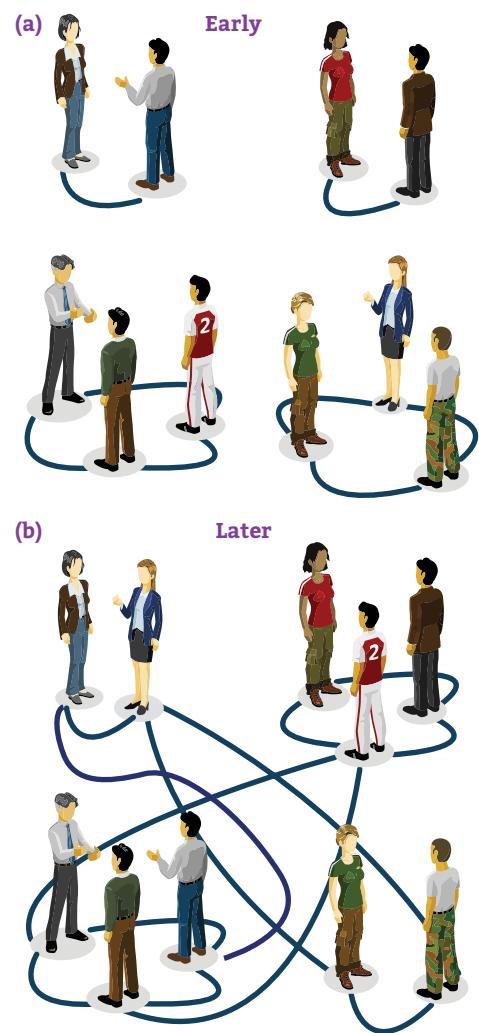
# INTRODUCTION

Imagine organizing a party for a hundred guests who initially do not know each other [1]. Offer them wine and cheese and you will soon see them chatting in groups of two to three. Now mention to Mary, one of your guests, that the red wine in the unlabeled dark green bottles is a rare vintage, much better than the one with the fancy red label. If she shares this information only with her acquaintances, your expensive wine appears to be safe, as she only had time to meet a few others so far.

The guests will continue to mingle, however, creating subtle paths between individuals that may still be strangers to each other. For example, while John has not yet met Mary, they have both met Mike, so there is an invisible path from John to Mary through Mike. As time goes on, the guests will be increasingly interwoven by such elusive links. With that the secret of the unlabeled bottle will pass from Mary to Mike and from Mike to John, escaping into a rapidly expanding group.

To be sure, when all guests had gotten to know each other, everyone would be pouring the superior wine. But if each encounter took only ten minutes, meeting all ninety-nine others would take about sixteen hours. Thus, you could reasonably hope that a few drops of your fine wine would be left for you to enjoy once the guests are gone.

Yet, you would be wrong. In this chapter we show you why. We will see that the party maps into a classic model in network science called the random network model. And random network theory tells us that we do not have to wait until *all* individuals get to know each other for our expensive wine to be in danger. Rather, soon after each person meets at least one other guest, an invisible network will emerge that will allow the information to reach all of them. Hence in no time everyone will be enjoying the better wine.



**Figure 3.1**  
**From a Cocktail Party to Random Networks**

The emergence of an acquaintance network through random encounters at a cocktail party.

(a) Early on the guests form isolated groups.

(b) As individuals mingle, changing groups, an invisible network emerges that connects all of them into a single network.

# THE RANDOM NETWORK MODEL

Network science aims to build models that reproduce the properties of real networks. Most networks we encounter do not have the comforting regularity of a crystal lattice or the predictable radial architecture of a spider web. Rather, at first inspection they look as if they were spun randomly (Figure 2.4). Random network theory embraces this apparent randomness by constructing and characterizing networks that are *truly random*.

From a modeling perspective a network is a relatively simple object, consisting of only nodes and links. The real challenge, however, is to decide where to place the links between the nodes so that we reproduce the complexity of a real system. In this respect the philosophy behind a random network is simple: We assume that this goal is best achieved by placing the links randomly between the nodes. That takes us to the definition of a random network (BOX 3.1):

*A random network consists of  $N$  nodes where each node pair is connected with probability  $p$ .*

To construct a random network we follow these steps:

- 1) Start with  $N$  isolated nodes.
- 2) Select a node pair and generate a random number between 0 and 1. If the number exceeds  $p$ , connect the selected node pair with a link, otherwise leave them disconnected.
- 3) Repeat step (2) for each of the  $N(N-1)/2$  node pairs.

The network obtained after this procedure is called a *random graph* or a *random network*. Two mathematicians, Pál Erdős and Alfréd Rényi, have played an important role in understanding the properties of these networks. In their honor a random network is called the *Erdős-Rényi network* (BOX 3.2).

## BOX 3.1

### DEFINING RANDOM NETWORKS

There are two definitions of a random network:

#### $G(N, L)$ Model

$N$  labeled nodes are connected with  $L$  randomly placed links. Erdős and Rényi used this definition in their string of papers on random networks [2-9].

#### $G(N, p)$ Model

Each pair of  $N$  labeled nodes is connected with probability  $p$ , a model introduced by Gilbert [10].

Hence, the  $G(N, p)$  model fixes the probability  $p$  that two nodes are connected and the  $G(N, L)$  model fixes the total number of links  $L$ . While in the  $G(N, L)$  model the average degree of a node is simply  $\langle k \rangle = 2L/N$ , other network characteristics are easier to calculate in the  $G(N, p)$  model. Throughout this book we will explore the  $G(N, p)$  model, not only for the ease that it allows us to calculate key network characteristics, but also because in real networks the number of links rarely stays fixed.

## BOX 3.2

### RANDOM NETWORKS: A BRIEF HISTORY

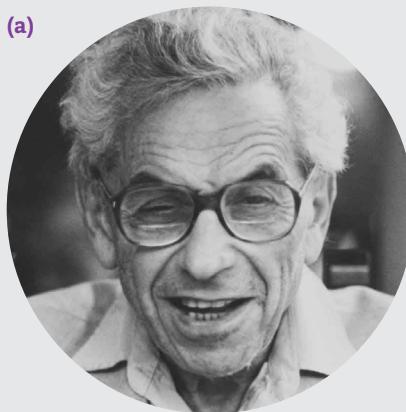


Figure 3.2

#### (a) Pál Erdős (1913-1996)

Hungarian mathematician known for both his exceptional scientific output and eccentricity. Indeed, Erdős published more papers than any other mathematician in the history of mathematics. He co-authored papers with over five hundred mathematicians, inspiring the concept of *Erdős number*. His legendary personality and profound professional impact has inspired two biographies [12, 13] and a documentary [14] ([Online Resource 3.1](#)).

#### (b) Alfréd Rényi (1921-1970)

Hungarian mathematician with fundamental contributions to combinatorics, graph theory, and number theory. His impact goes beyond mathematics: The Rényi entropy is widely used in chaos theory and the random network theory he co-developed is at the heart of network science. He is remembered through the hotbed of Hungarian mathematics, the Alfréd Rényi Institute of Mathematics in Budapest.

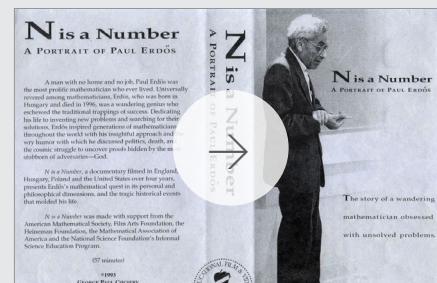
Anatol Rapoport (1911-2007), a Russian immigrant to the United States, was the first to study random networks. Rapoport's interests turned to mathematics after realizing that a successful career as a concert pianist would require a wealthy patron. He focused on mathematical biology at a time when mathematicians and biologists hardly spoke to each other. In a paper written with Ray Solomonoff in 1951 [11], Rapoport demonstrated that if we increase the average degree of a network, we observe an abrupt transition from disconnected nodes to a graph with a giant component.

The study of random networks reached prominence thanks to the fundamental work of Pál Erdős and Alfréd Rényi (Figure 3.2). In a sequence of eight papers published between 1959 and 1968 [2-9], they merged probability theory and combinatorics with graph theory, establishing *random graph theory*, a new branch of mathematics [2].

The random network model was independently introduced by Edgar Nelson Gilbert (1923-2013) [10] the same year Erdős and Rényi published their first paper on the subject. Yet, the impact of Erdős and Rényi's work is so overwhelming that they are rightly considered the founders of random graph theory.

***"A mathematician is a device for turning coffee into theorems"***

Alfréd Rényi (a quote often attributed to Erdős)



#### Online Resource 3.1

##### **N is a Number: A Portrait of Paul Erdős**

The 1993 biographical documentary of Pál Erdős, directed by George Paul Csicsery, offers a glimpse into Erdős' life and scientific impact [14].



# NUMBER OF LINKS

Each random network generated with the same parameters  $N, p$  looks slightly different (Figure 3.3). Not only the detailed wiring diagram changes between realizations, but so does the number of links  $L$ . It is useful, therefore, to determine how many links we expect for a particular realization of a random network with fixed  $N$  and  $p$ .

The probability that a random network has exactly  $L$  links is the product of three terms:

- 1) The probability that  $L$  of the attempts to connect the  $N(N-1)/2$  pairs of nodes have resulted in a link, which is  $p^L$ .
- 2) The probability that the remaining  $N(N-1)/2 - L$  attempts have not resulted in a link, which is  $(1-p)^{N(N-1)/2-L}$ .
- 3) A combinational factor,

$$\binom{\frac{N(N-1)}{2}}{L}, \quad (3.0)$$

counting the number of different ways we can place  $L$  links among  $N(N-1)/2$  node pairs.

We can therefore write the probability that a particular realization of a random network has exactly  $L$  links as

$$p_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}. \quad (3.1)$$

As (3.1) is a binomial distribution (BOX 3.3), the expected number of links in a random graph is

$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L p_L = p \frac{N(N-1)}{2}. \quad (3.2)$$

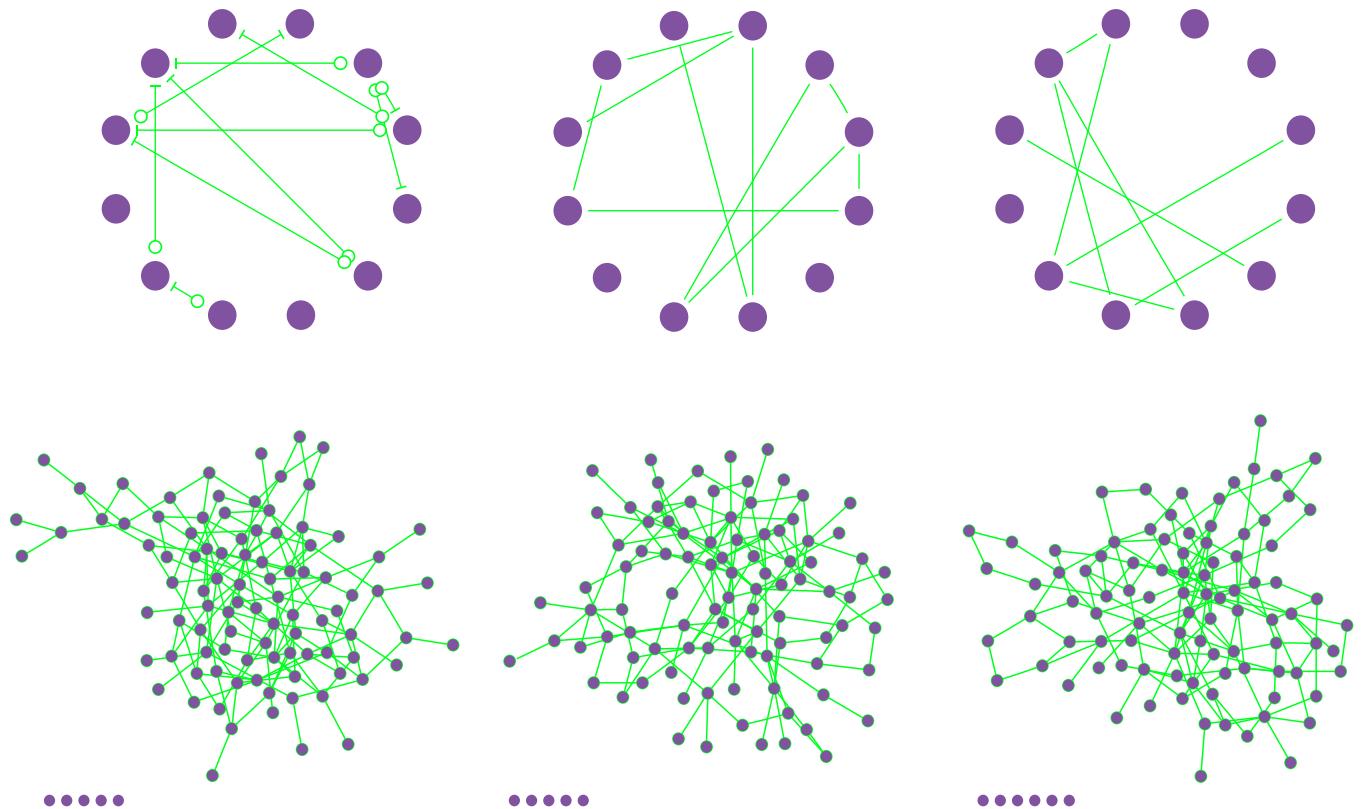
Hence  $\langle L \rangle$  is the product of the probability  $p$  that two nodes are connected and the number of pairs we attempt to connect, which is  $L_{\max} = N(N - 1)/2$  (CHAPTER 2).

Using (3.2) we obtain the average degree of a random network

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N - 1). \quad (3.3)$$

Hence  $\langle k \rangle$  is the product of the probability  $p$  that two nodes are connected and  $(N-1)$ , which is the maximum number of links a node can have in a network of size  $N$ .

In summary the number of links in a random network varies between realizations. Its expected value is determined by  $N$  and  $p$ . If we increase  $p$  a random network becomes denser: The average number of links increase linearly from  $\langle L \rangle = 0$  to  $L_{\max}$  and the average degree of a node increases from  $\langle k \rangle = 0$  to  $\langle k \rangle = N-1$ .



**Figure 3.3**  
**Random Networks are Truly Random**

#### Top Row

Three realizations of a random network generated with the same parameters  $p=1/6$  and  $N=12$ . Despite the identical parameters, the networks not only look different, but they have a different number of links as well ( $L=10, 10, 8$ ).

#### Bottom Row

Three realizations of a random network with  $p=0.03$  and  $N=100$ . Several nodes have degree  $k=0$ , shown as isolated nodes at the bottom.

## BOX 3.3

### BINOMIAL DISTRIBUTION: MEAN AND VARIANCE

If we toss a fair coin  $N$  times, tails and heads occur with the same probability  $p = 1/2$ . The binomial distribution provides the probability  $p_x$  that we obtain exactly  $x$  heads in a sequence of  $N$  throws. In general, the binomial distribution describes the number of successes in  $N$  independent experiments with two possible outcomes, in which the probability of one outcome is  $p$ , and of the other is  $1-p$ .

The binomial distribution has the form

$$p_x = \binom{N}{x} p^x (1-p)^{N-x}.$$

The mean of the distribution (first moment) is

$$\langle x \rangle = \sum_{x=0}^N x p_x = Np. \quad (3.4)$$

Its second moment is

$$\langle x^2 \rangle = \sum_{x=0}^N x^2 p_x = p(1-p)N + p^2 N^2, \quad (3.5)$$

providing its standard deviation as

$$\sigma_x = \left( \langle x^2 \rangle - \langle x \rangle^2 \right)^{\frac{1}{2}} = [p(1-p)N]^{\frac{1}{2}}. \quad (3.6)$$

Equations (3.4) - (3.6) are used repeatedly as we characterize random networks.