

Random Networks (ER Model)

Social Networks Analysis and Graph Algorithms

Prof. Carlos “ChaTo” Castillo — <https://chato.cl/teach>



Universitat
Pompeu Fabra
Barcelona

Contents

- The ER model
- Degree distribution under the ER model

Sources

- A. L. Barabási (2016). Network Science – Chapter 03
- Data-Driven Social Analytics course by Vicenç Gómez and Andreas Kaltenbrunner
- URLs cited in the footer of specific slides

Network Models

Network models

- Networks of many different types have similar properties:
 - Short paths
 - Many triangles
 - Skewed degree distributions
- Where do such properties come from?
- How do nodes connect to each other? How are triangles formed?
- We will study **network models**, i.e., sets of instructions to create networks

Why studying network models?

- Our models will be **stochastic**, i.e., randomized
- Running **stochastic network models** can let us check if they generate networks that **look like real ones**
- Almost invariably, the generated networks will be similar to actual networks in some ways, but **different in other ways**

The “Random Network”

Erdős-Rényi (ER) Model

Sounds like “ERDOSH and REGN”



Paul Erdős
(1913-1996)



Alfred Rényi
(1921-1970)

Meeting people at a party

- You pick a random person
- Talk to that person for a while, if there are good vibes, you are connected
- Then pick another person
 - And repeat
- The result is what we call a **random network**



Formalization (Erdős-Rényi or ER)

Sounds like “ERDOSH and REGN”

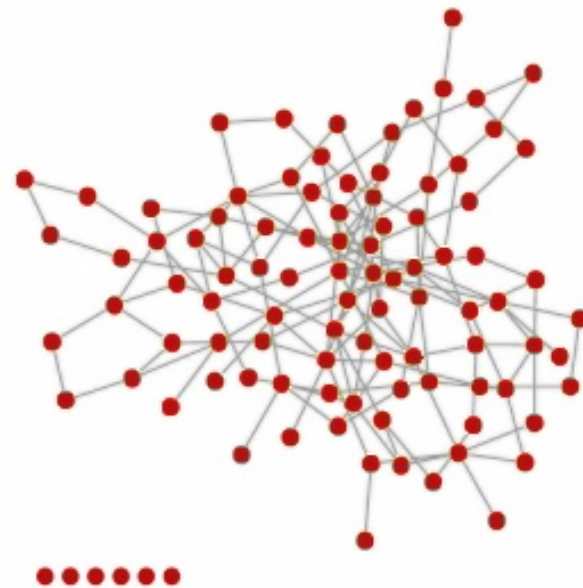
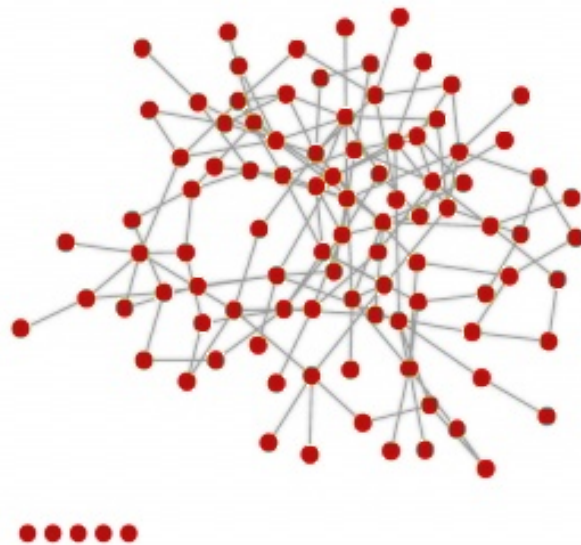
- For each pair of nodes in the graph
 - Perform a **Bernoulli trial** with probability p
 - “Toss a biased coin with probability p of landing heads”
 - If the trial succeeds, **connect** those nodes
 - “If the coin lands heads, connect those nodes”
- Repeat for all pairs $\frac{N(N-1)}{2}$

Example

(3 networks, same parameters)

$$N = 100, p = 0.03$$

Nodes at the bottom ended up isolated

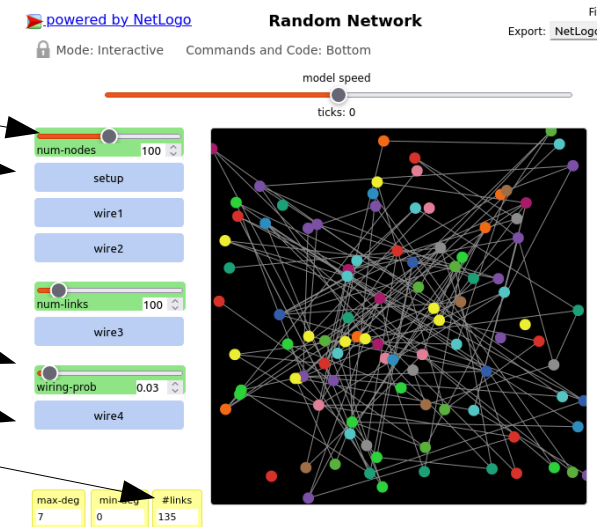


Exercise

Guess formula for $E[L]$

Go to netlogoweb.org/launch and select:
“IABM Textbook / chapter 5 / Random Network”

- Execute the “Random Network” program in Netlogo Web
 - Select num-nodes N (e.g., 100)
 - Click “setup”
 - Select wiring-prob p (e.g., 0.03)
 - Click “wire4”
 - Write down “#links” L somewhere
 - Repeat various times
- Guess a formula for $E[L]$ as a function of N and p



Pin board: <https://upfbarcelona.padlet.org/chato/84a1nj59pkqpxvh3>

Degree distribution

A key characteristic of a network: its degree distribution

- One of the most evident characteristics of a network is its **degree distribution**
 - Is this distribution very skewed? Or every node is close to some average? Is there a “typical” degree?
 - Does it look like the degree distribution predicted by a network formation model?
- We will spend a fair amount of time studying the degree distribution under various models

The binomial distribution

- The distribution of the probability of obtaining x successes in n independent trials, in which each trial has probability of succeeding p

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

$$\langle x \rangle = \sum_{x=0}^n x p_x = np$$

Degree distribution in ER model

- Simply a Binomial distribution
- Note that the maximum number of “successes” (links) of a node is $N-1$, hence:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

$$\langle k \rangle = p(N-1)$$

Expected number of links

- Expected number of links

$$\langle L \rangle = p \cdot L_{\max} = p \frac{N(N-1)}{2}$$

- Average degree

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

Exercise [B. 2016, Ex. 3.11.1]

Expected number of links and average degree

- Consider an ER graph with $N=3,000$ $p=10^{-3}$
 - 1) What is the expected number of links $\langle L \rangle$?
 - 2) What is the average degree $\langle k \rangle$?

$$\langle L \rangle = p \cdot L_{\max} = p \frac{N(N-1)}{2}$$

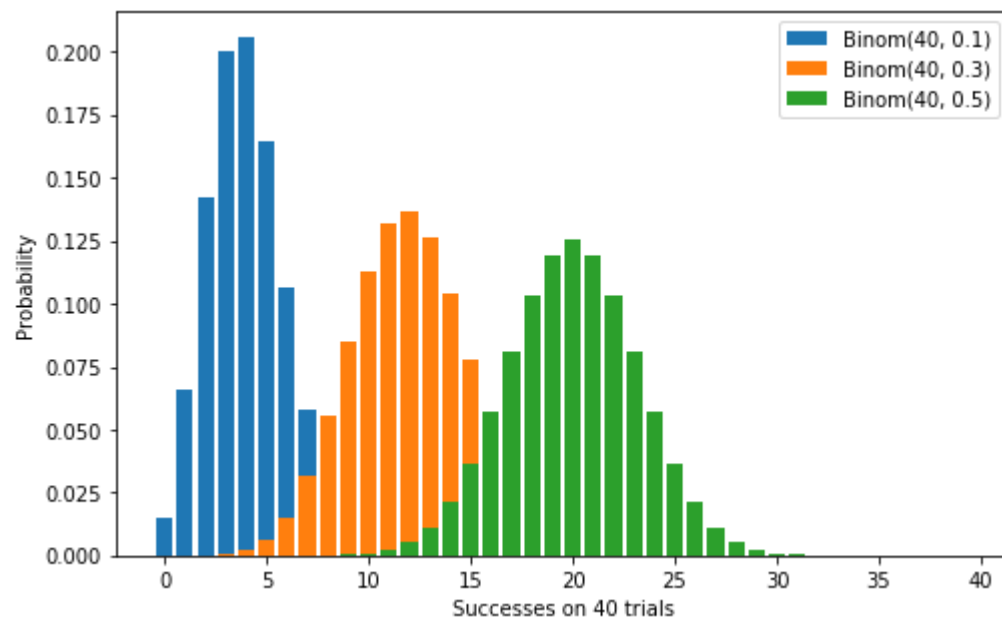
$$\langle k \rangle = \frac{2 \langle L \rangle}{N} = p(N-1)$$

Degree distribution examples

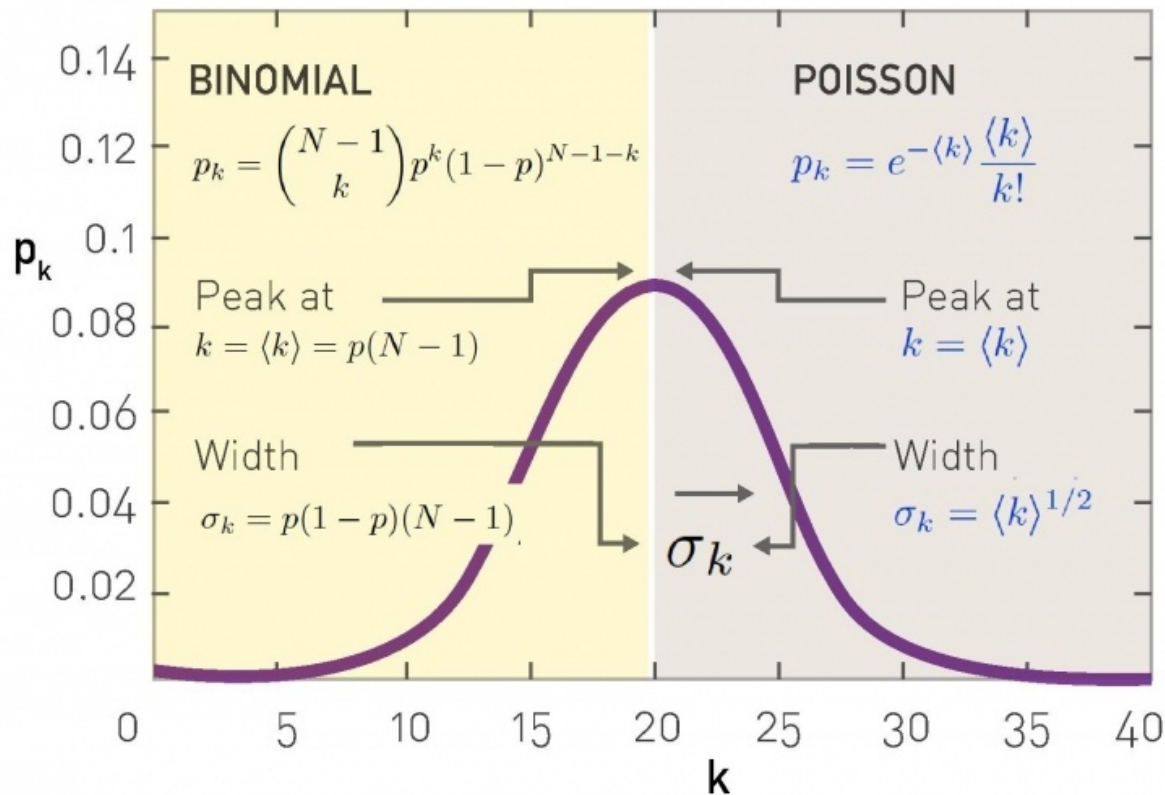
- The peak is always at $\langle k \rangle = p(N - 1)$

```
import numpy as np
from scipy.stats import binom
from matplotlib import pyplot as plt
```

```
x = np.arange(0, 40)
plt.figure(figsize=(8,5))
plt.bar(x, (binom(40, 0.1)).pmf(x), label='Binom(40, 0.1)')
plt.bar(x, (binom(40, 0.3)).pmf(x), label='Binom(40, 0.3)')
plt.bar(x, (binom(40, 0.5)).pmf(x), label='Binom(40, 0.5)')
plt.gca().legend()
plt.xlabel("Successes on 40 trials")
plt.ylabel("Probability")
plt.show()
```

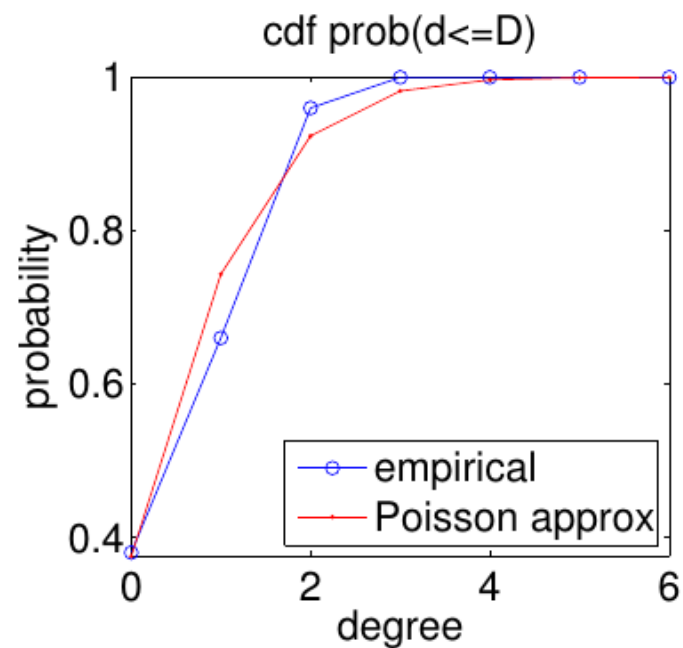
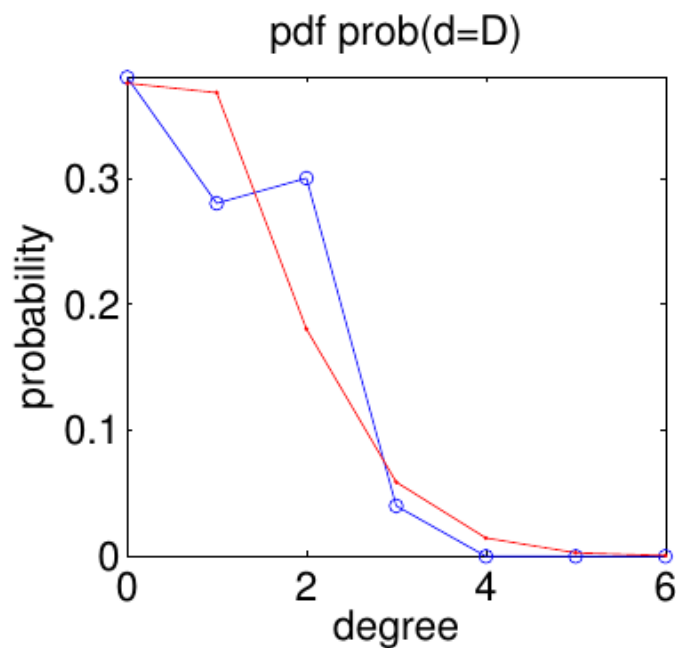
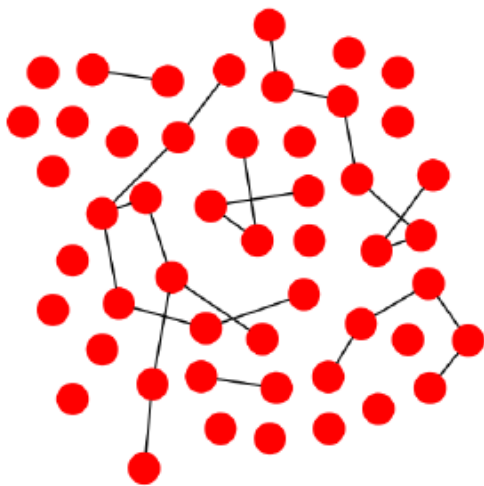


Approximation with a Poisson distribution for $\langle k \rangle \ll N$



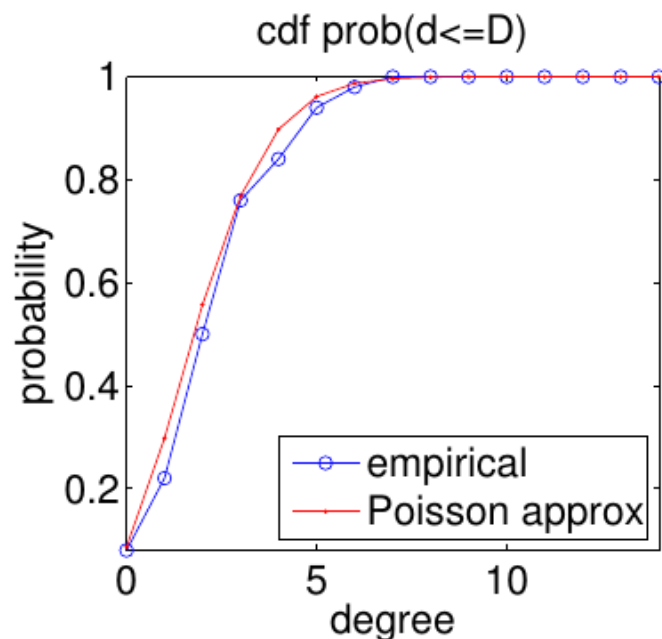
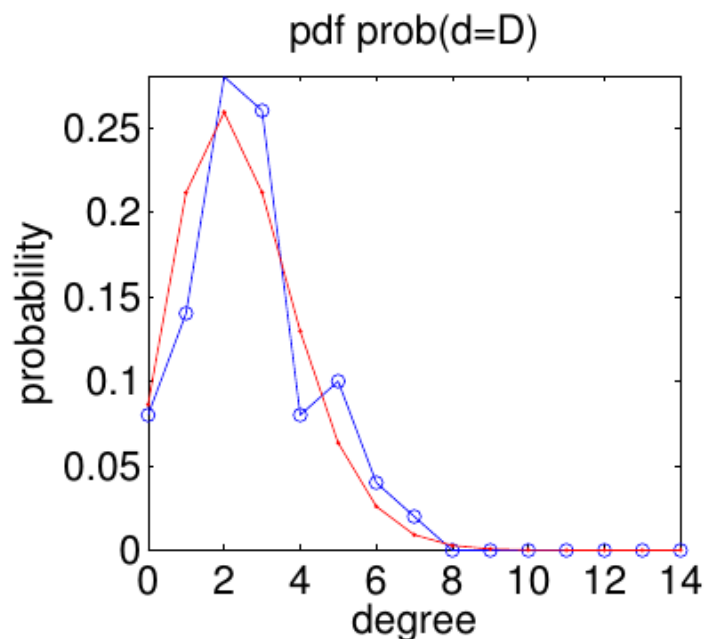
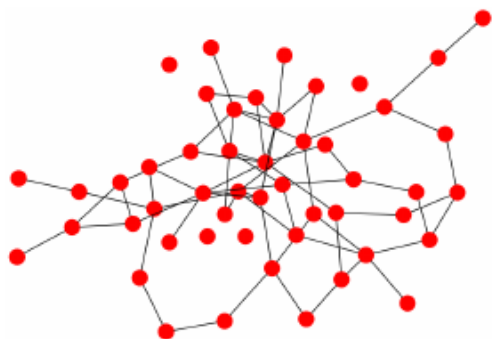
More examples (1/6)

$$N = 50, p = 0.02, \langle k \rangle \approx 1$$



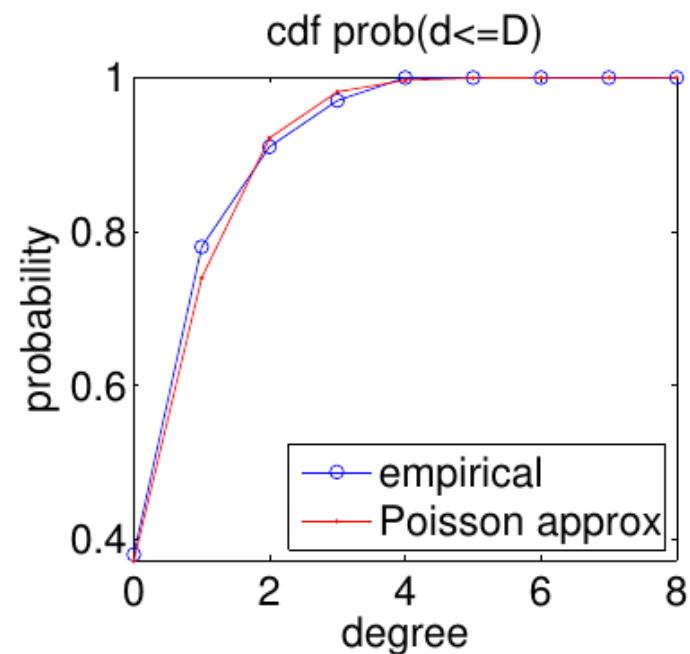
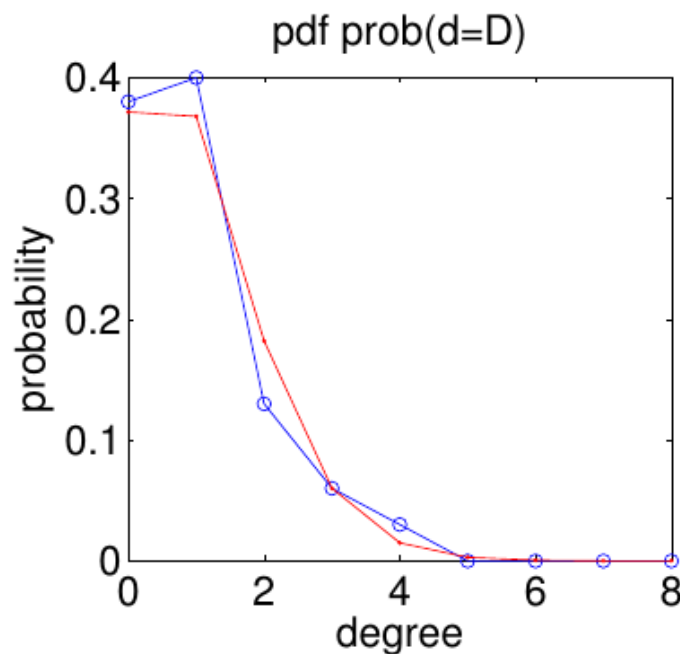
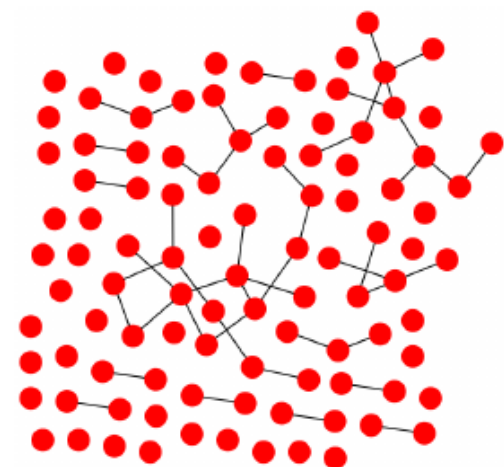
More examples (2/6)

$$N = 50, p = 0.05, \langle k \rangle \approx 2.5$$



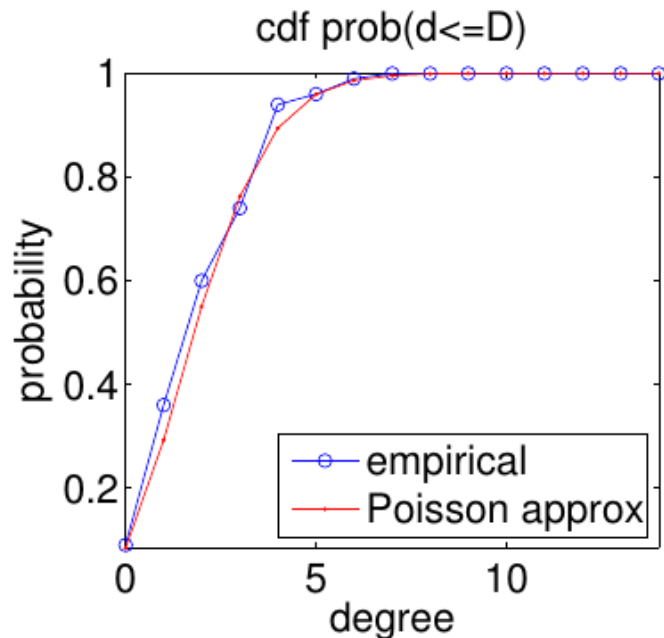
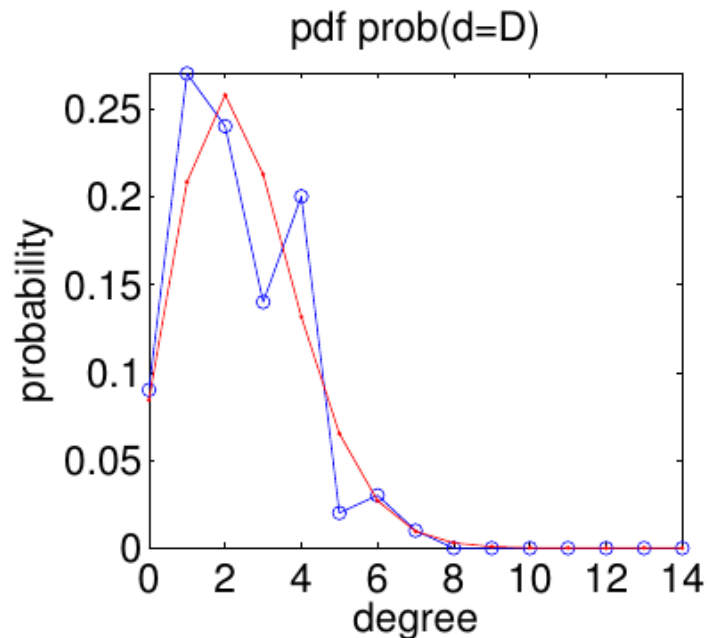
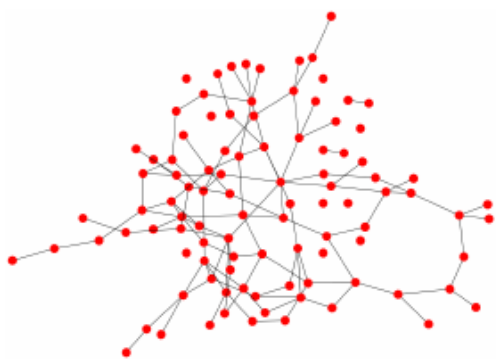
More examples (3/6)

$$N = 100, p = 0.01, \langle k \rangle \approx 1$$



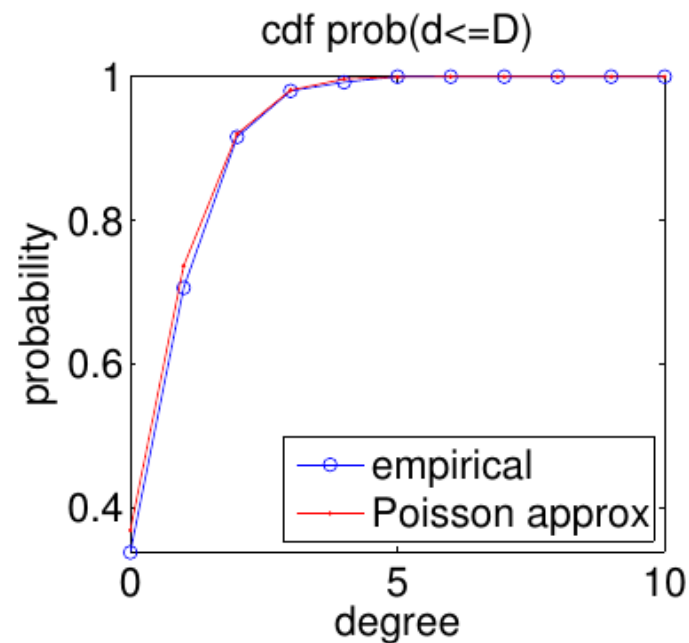
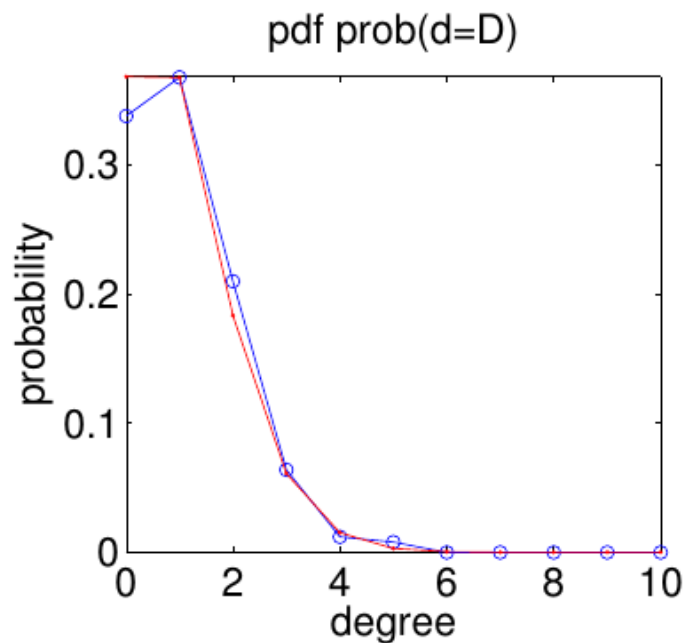
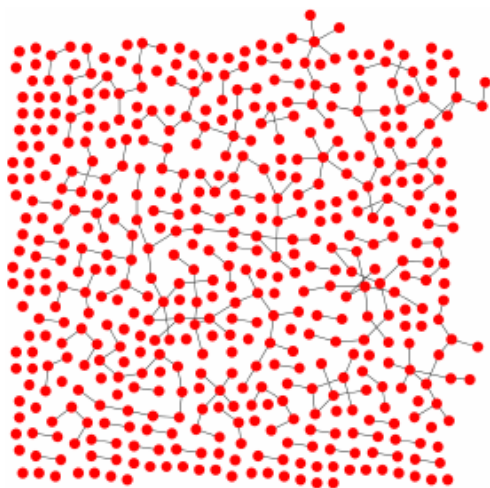
More examples (4/6)

$$N = 100, p = 0.025, \langle k \rangle \approx 2.5$$



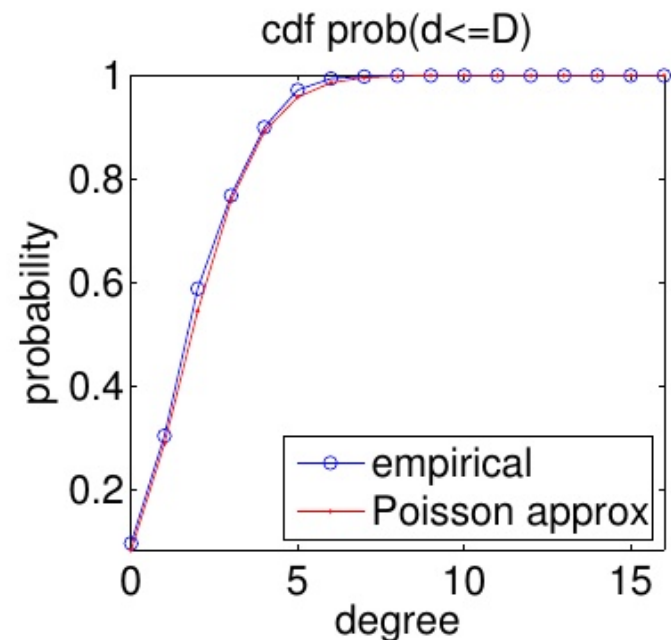
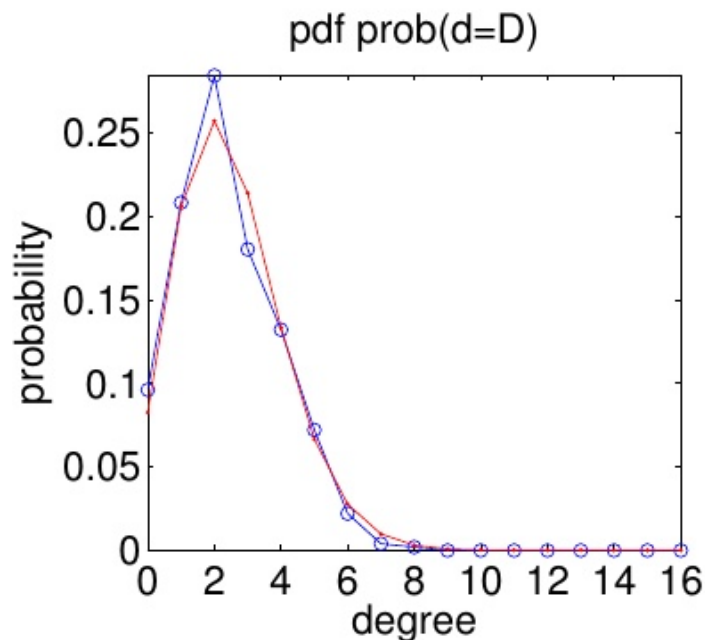
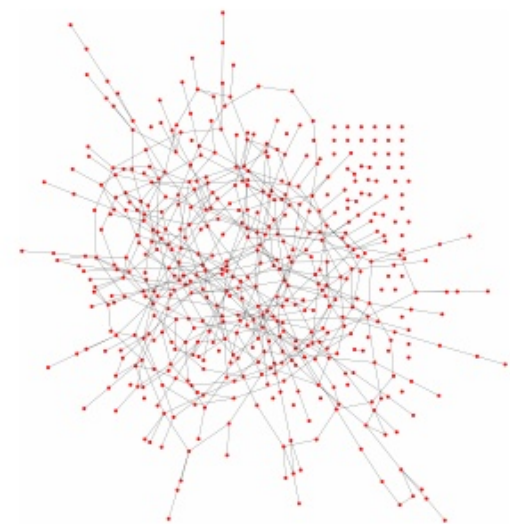
More examples (5/6)

$$N = 500, p = 0.002, \langle k \rangle \approx 1$$



More examples (6/6)

$$N = 500, p = 0.005, \langle k \rangle \approx 2.5$$



“Back of the envelope” calculations

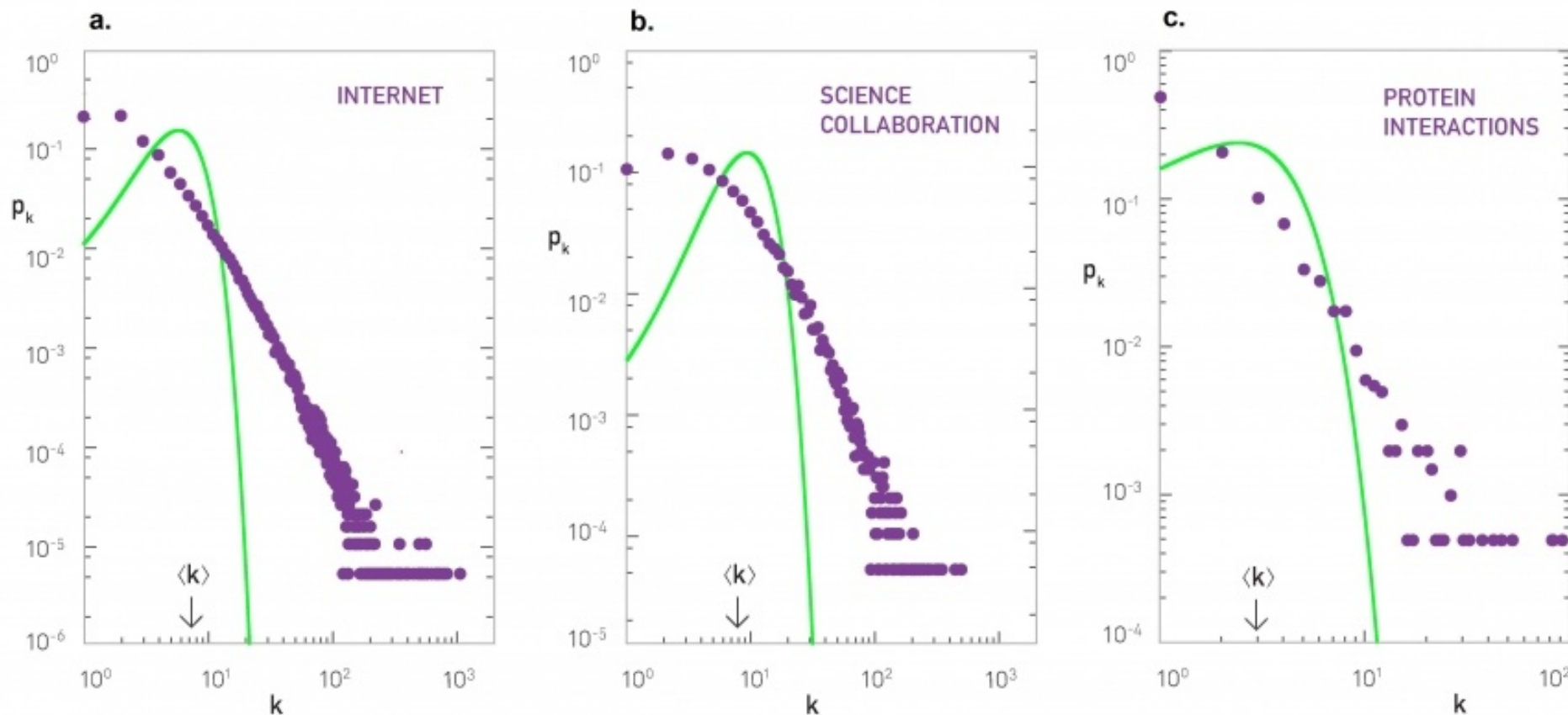
- Suppose $N = 7 \times 10^9$
- Suppose $\langle k \rangle = 1,000$
 - A person knows the name of approx. 1,000 others
- Then on expectation $k_{\max} = 1,185$
- $\langle k \rangle \pm \sigma$ is the range from 968 to 1,032
- Is this realistic?

Survey: how many WhatsApp contacts do you have?



<https://forms.gle/9xEYhzv2U5NrPQdH8>

Real networks (green = $e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$)



Summary

Things to remember

- The ER model
- Degree distribution in the ER model

Practice on your own

- Indicate the expected number of edges of a network with $N=256$, $p=0.25$; then compare your solution with the one on this video:



<https://www.youtube.com/watch?v=2DckiyysQy4>