Random Network Models

Introduction to Network Science

Instructor: Michele Starnini — https://github.com/chatox/networks-science-course

Universitat Pompeu Fabra Barcelona

Contents

- The ER model
- Degree distribution under the ER model

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

Modelling?

Quantify Phenomenon

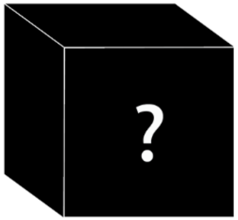
Inform

Assumptions

Test

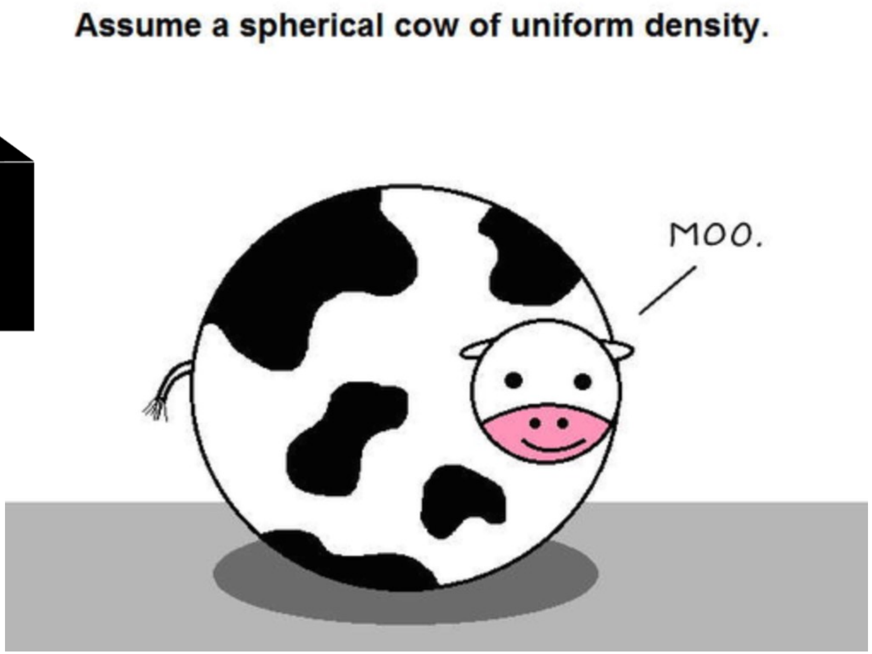
- Reproduce
- Forecast

Input



Output

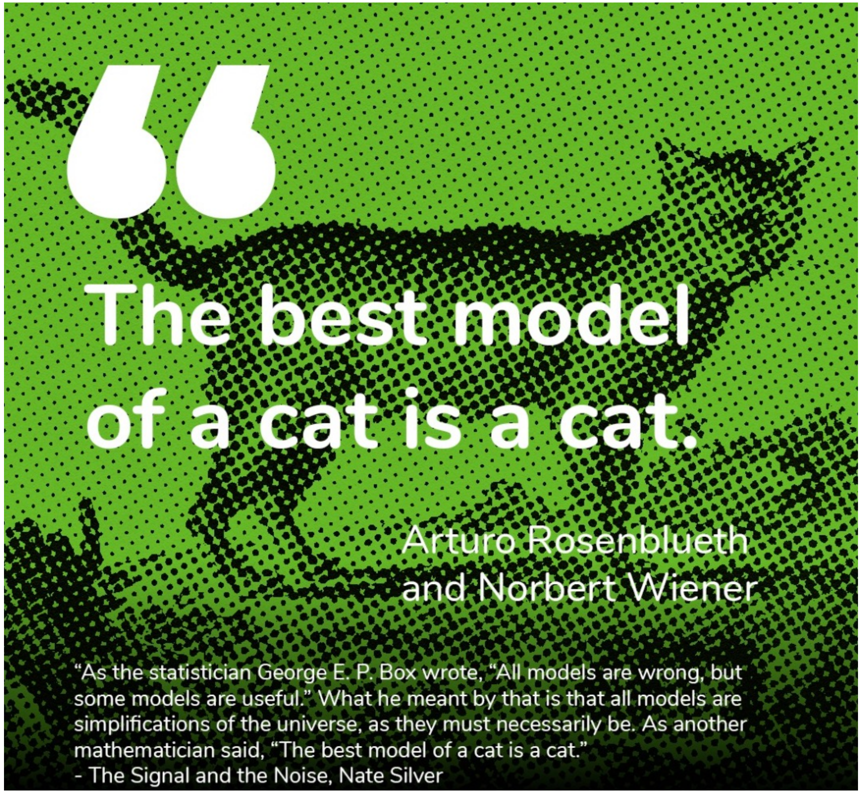
"All models are wrong, but some are useful"

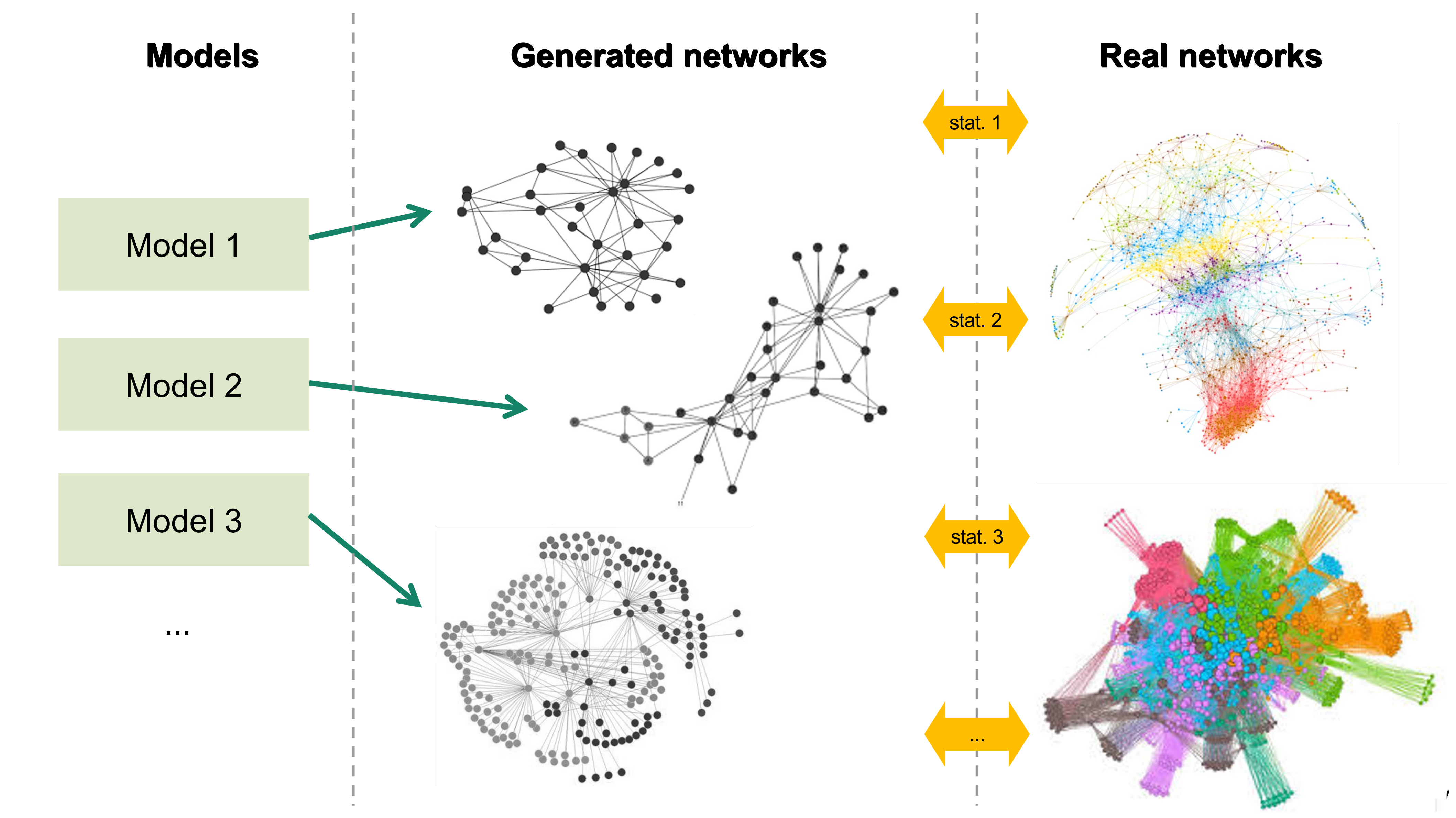


Modelling

- **Null models**
  - Preserve some properties, randomize ALL the rest
  - Compare data with null hypothesis: statistical test.
  - Example: homophily. Comparison with “random network”. Statistically significant?
- **Realistic models**
  - Develop a model that “explains” some observed property
  - Compare with data: How good is the model?
  - Example: homophily. Assume some mechanisms leading to homophily.

Modelling Cats




Models
Generated networks
Real networks
Model 1
Model 2
Model 3
...
stat. 1
stat. 2
stat. 3
...

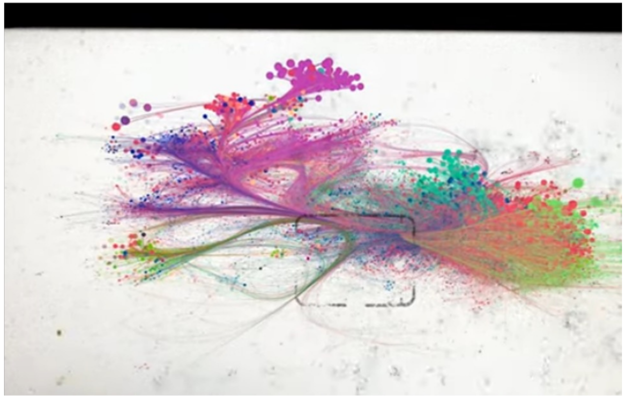
The “Random Network”
Erdös-Rényi (ER) Model

Sounds like “ERDOSH and REGN”

Paul Erdös
(1913-1996)

Alfred Rényi
(1921-1970)

Video (01:20-02:26) by Albert-László Barabási (cont.)



https://www.youtube.com/watch?v=RfgjHoVCZwU
Until "… in a random network, the average dominates."

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 (Erdös-Rényi or ER)

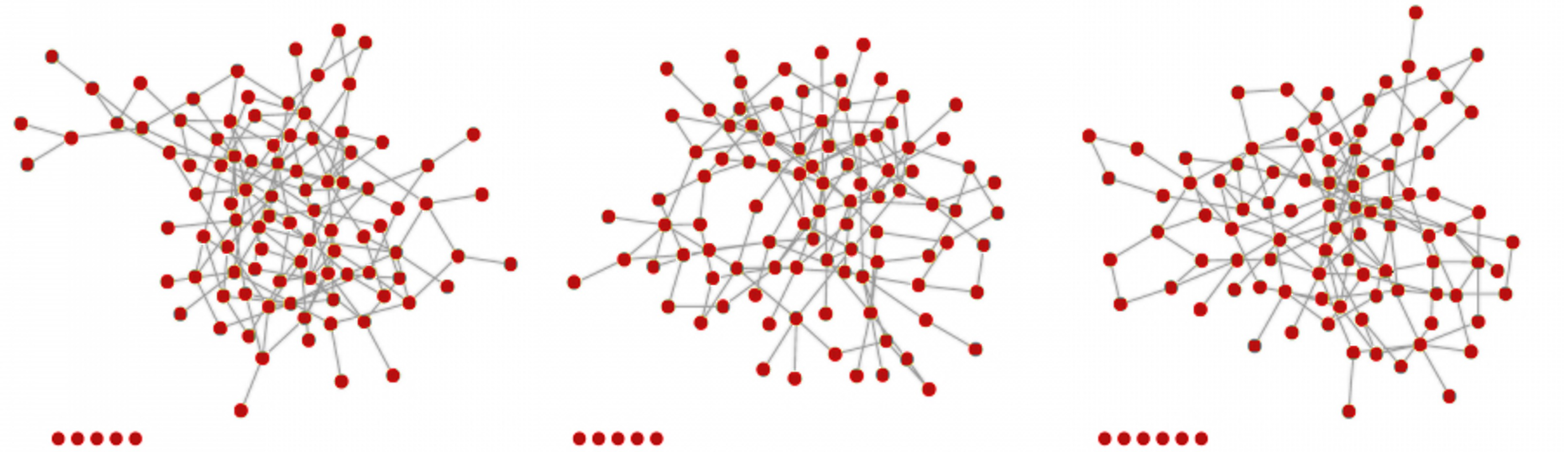
Sounds like "ERDOSH and REGN"

- 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 a formula for $\langle L \rangle$ as a function of N and p

Actual number of links in ER networks is variable!

The **expected** number of links is <L>

Remember the network model has only two parameters: N and p.

Actually, the model explicitly considers all possible links: $N(N-1)/2$.

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

Exactly x "YES"

The order is not relevant!
How many sequences with x "YES" and n-x "NO"?

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

Exactly n-x "NO"

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

$$\Rightarrow$$

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

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

Degree distribution in ER model

- Probability of finding a node with degree k
- Max number of “successes” (links) of a node is N-1
- Each possible link is present with prob p

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

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

Exercise: Prove it!

Links & average degree

- Expected number of links

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

- Average degree

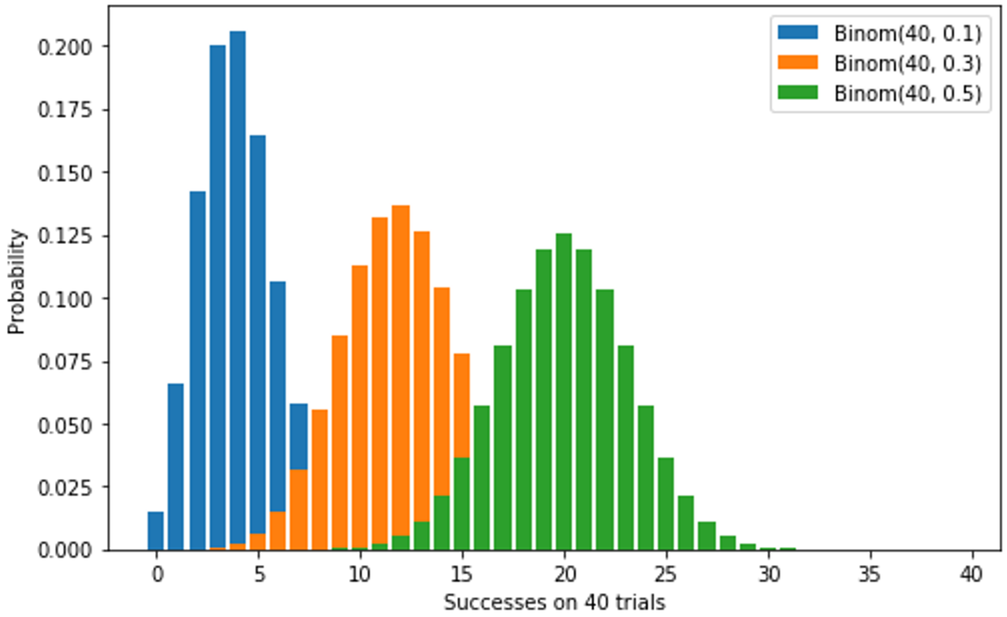
$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$

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()
```



python™

Exercise [B. 2016, Ex. 3.11.1]

Expected number of links and average degree

- Consider an ER graph with N=3,000 $p=10^{-3}$

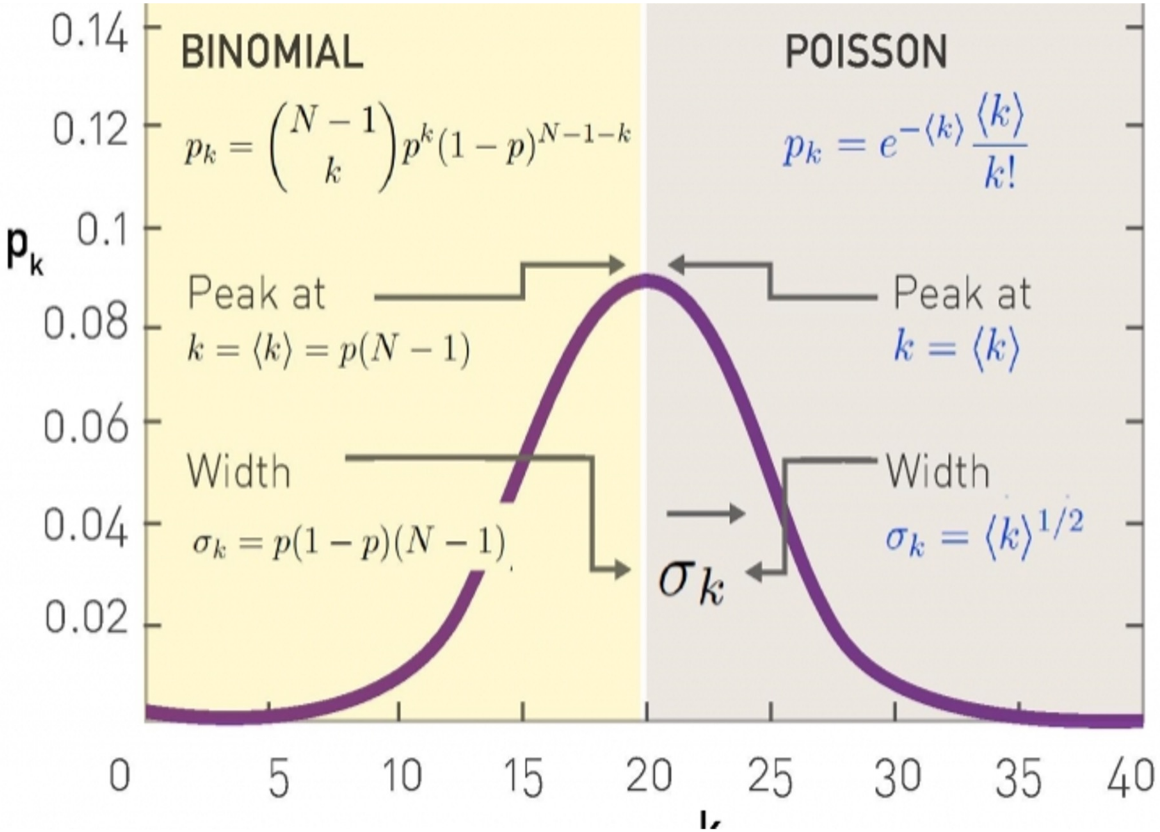
1)What is the expected number of links ⟨L⟩?

2)What is the average degree ⟨k⟩?

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

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

Approximation: Poisson distribution



Valid if

$$\langle k \rangle \ll N$$

$$\sigma_x^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2 = \sum_x x^2 p(x) - (\sum_x x p(x))^2 = \langle x^2 \rangle - \langle x \rangle^2$$

$$\sigma_k^2 = \langle k \rangle$$

Exercise: Prove it!

Poisson distribution

– It does not depend on N (valid only for large N)

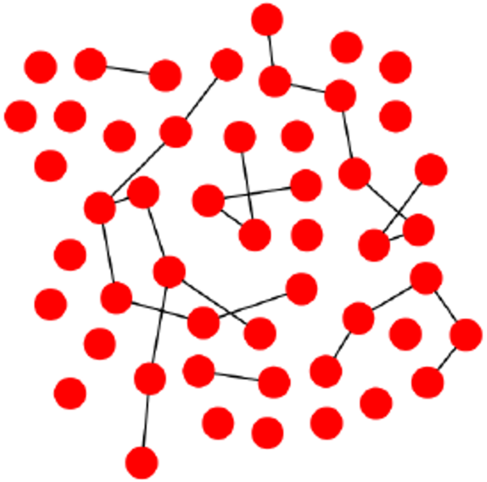
– Completely described by a single parameter <k>

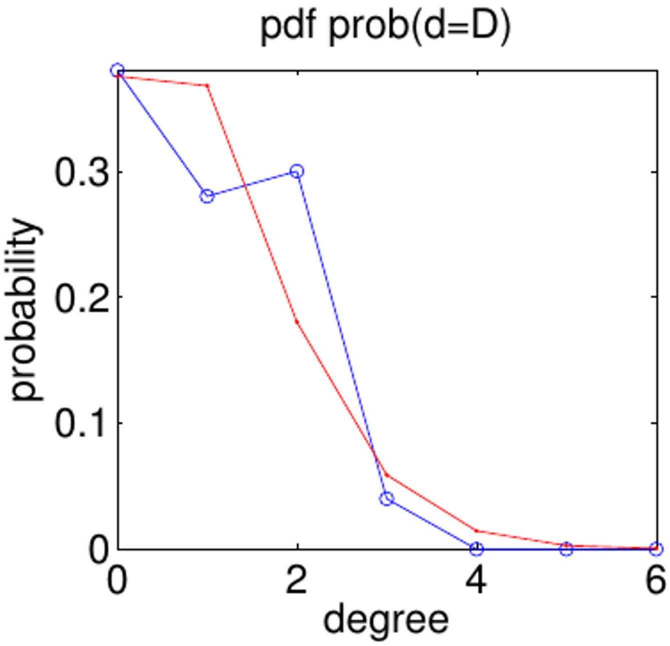
– Can be derived by the binomial distribution by applying <k> << N (try it!)

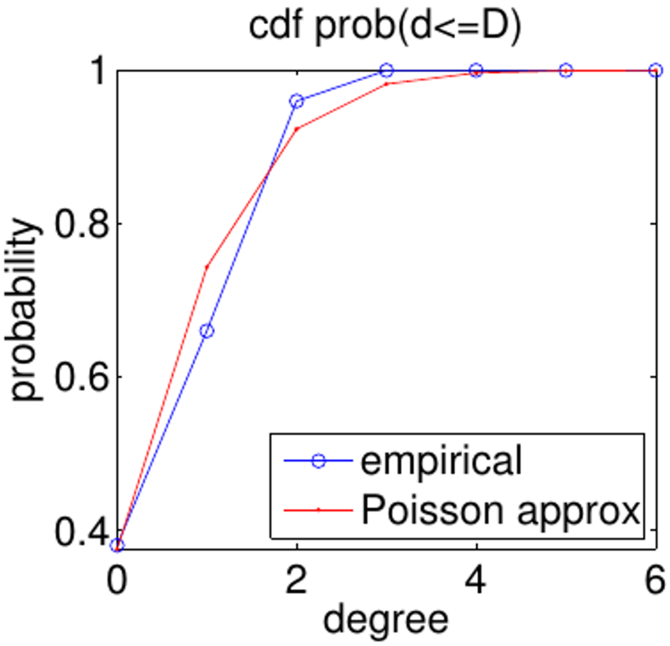
– <k> << N, p << N/(N-1), p << 1 (and large N)

More examples (1/6)

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



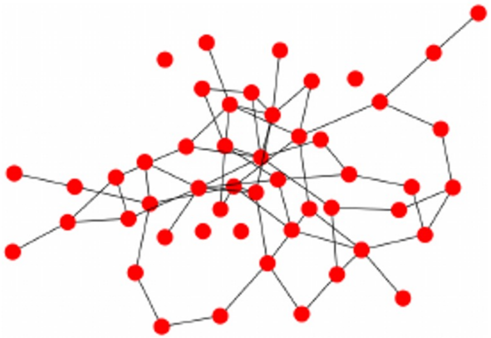


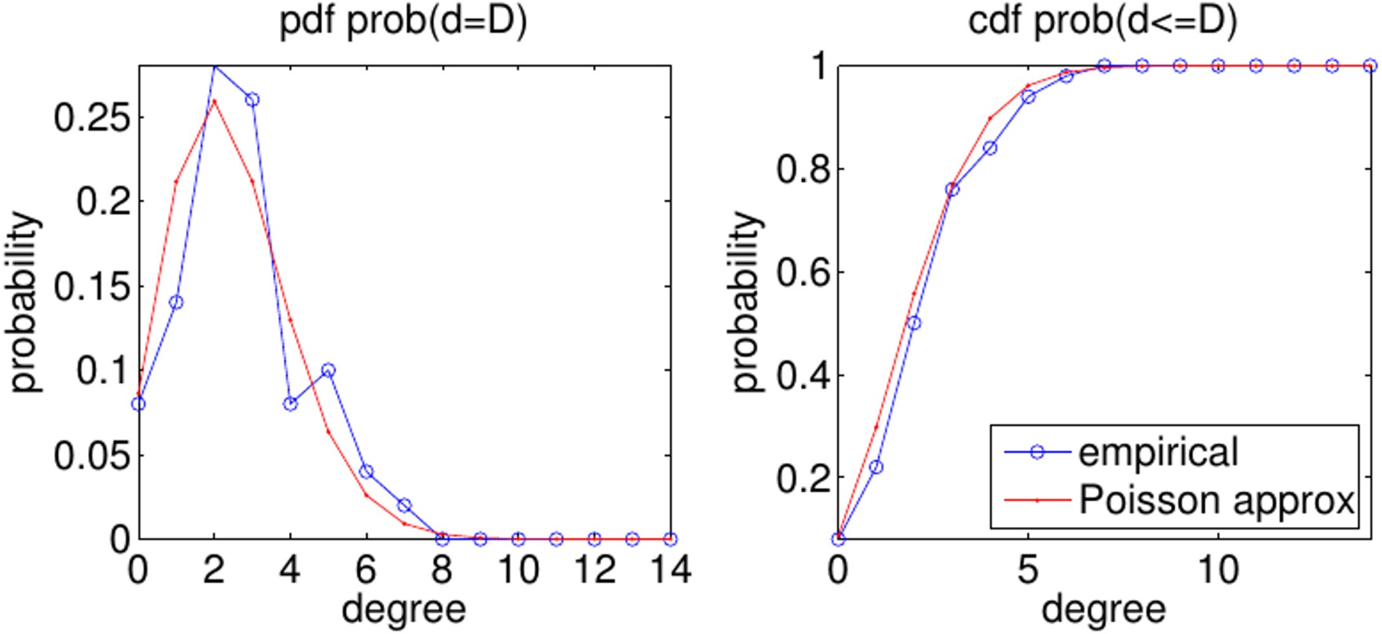


https://www.upf.edu/web/iis/DataDrivenSocialAnalytics

More examples (2/6)

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

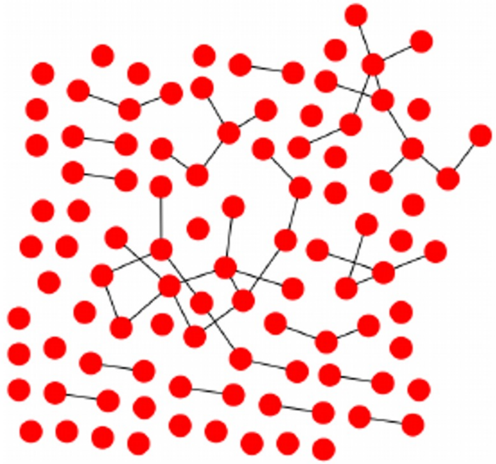


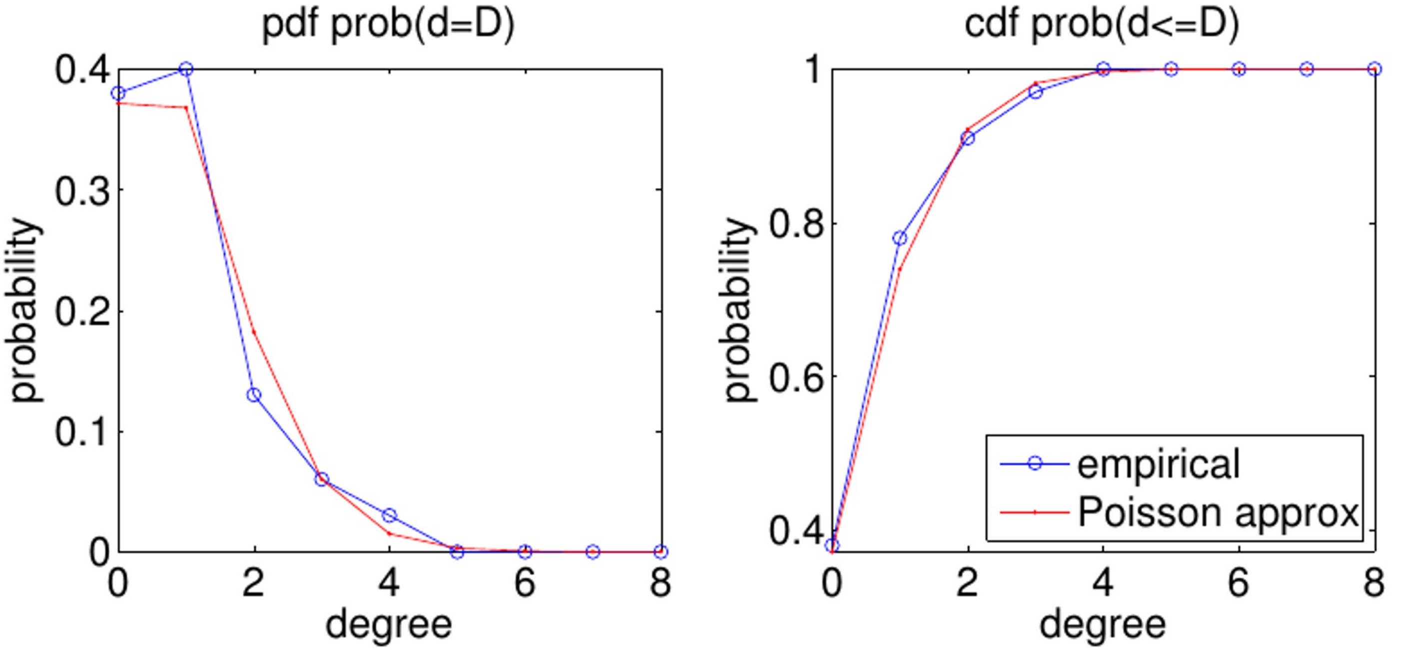


https://www.upf.edu/web/iis/DataDrivenSocialAnalytics

More examples (3/6)

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

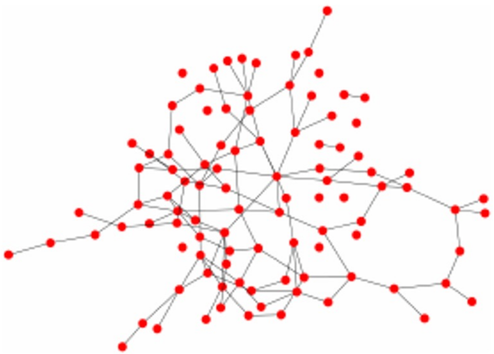


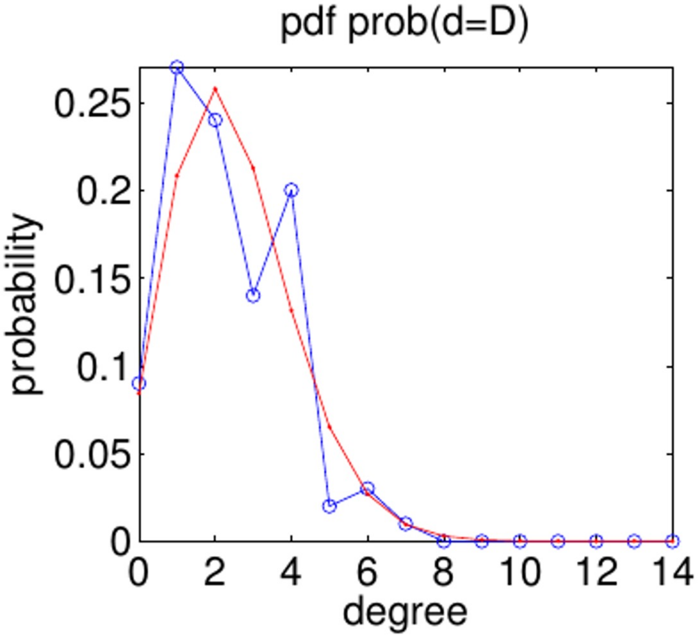


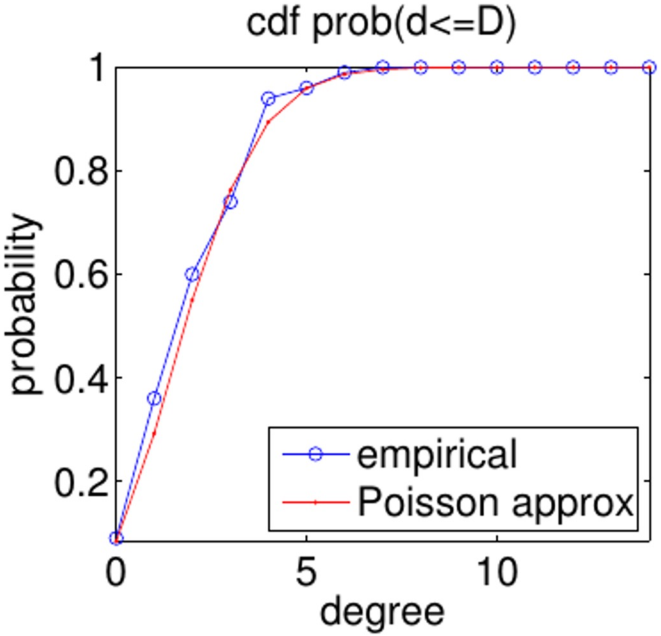
https://www.upf.edu/web/iis/DataDrivenSocialAnalytics

More examples (4/6)

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



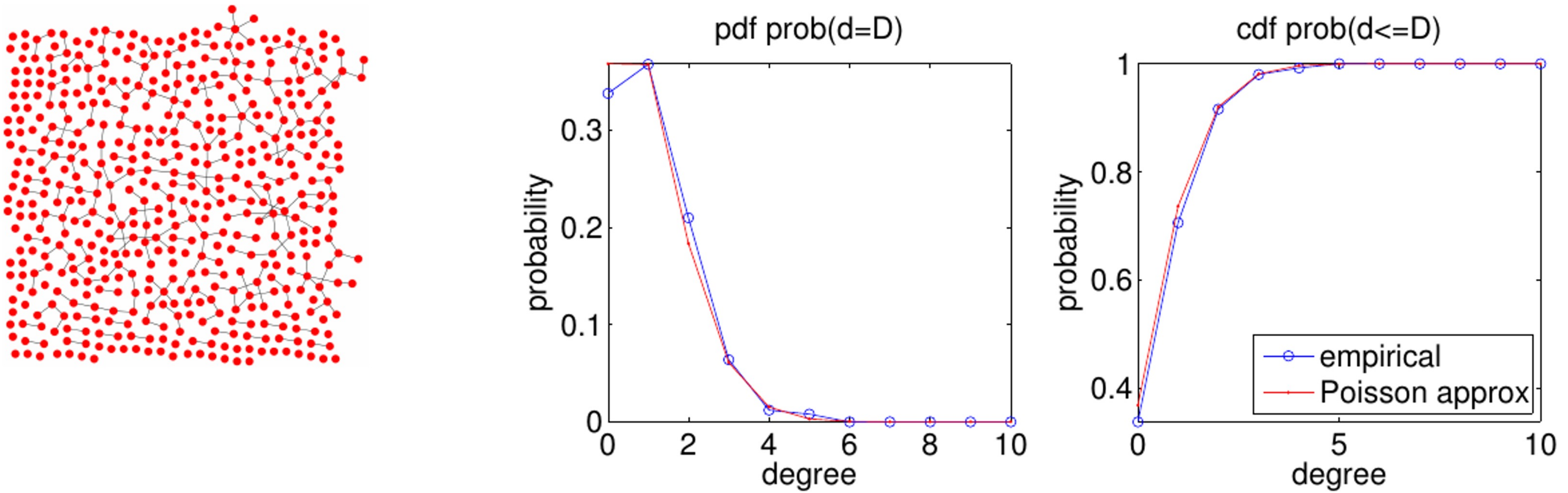




https://www.upf.edu/web/iis/DataDrivenSocialAnalytics

More examples (5/6)

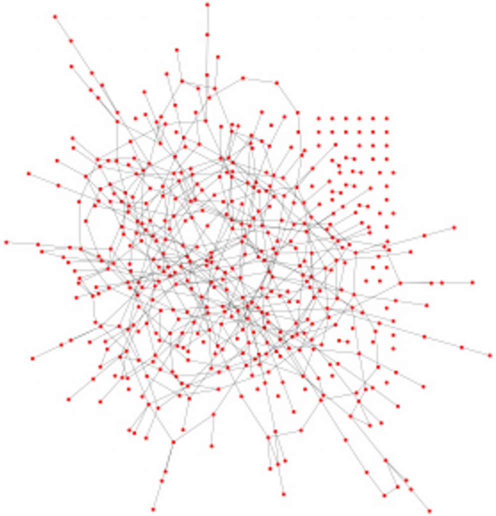
$$N = 500, p = 0.002, \langle k \rangle \approx 1$$

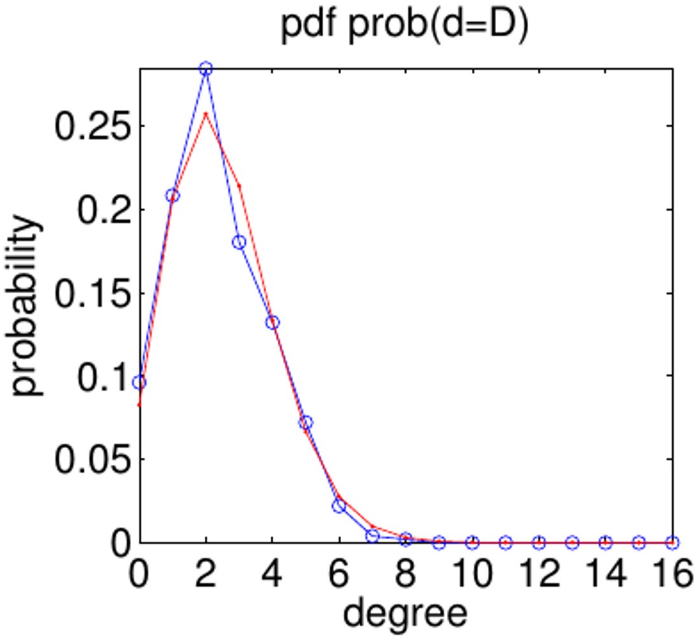


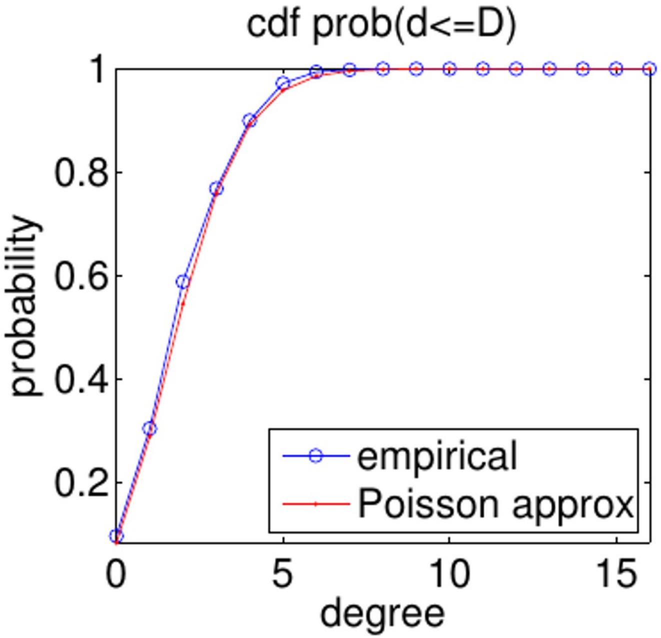
https://www.upf.edu/web/iis/DataDrivenSocialAnalytics

More examples (6/6)

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







https://www.upf.edu/web/iis/DataDrivenSocialAnalytics

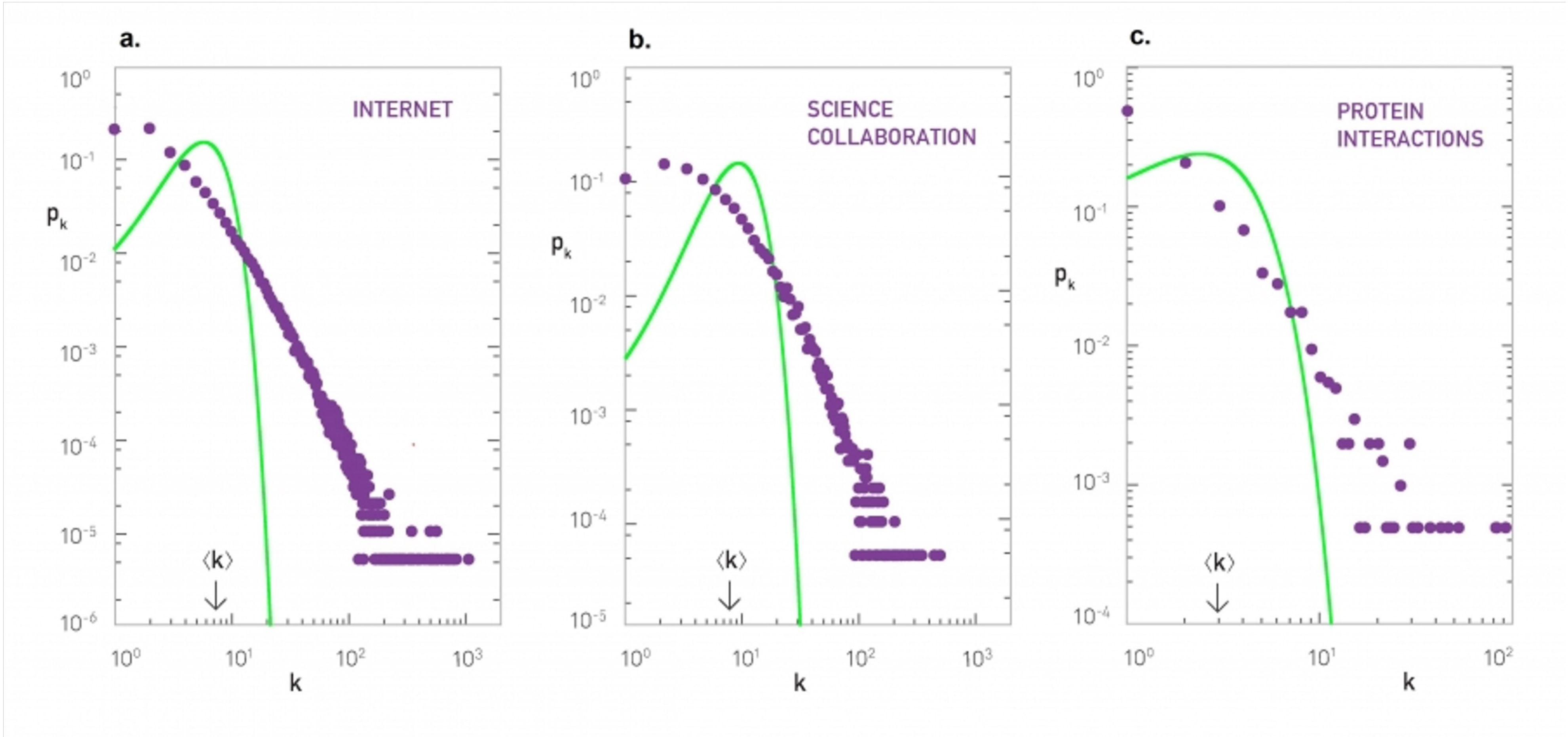
“Back of the envelope” calculations

- Suppose N = 7 x 10^9
- Suppose <k> = 1,000
  - A person knows the name of approx. 1,000 others
- <k> ± σ is the range from 968 to 1,032
- Is this realistic?

Survey: how many WhatsApp contacts do you have?

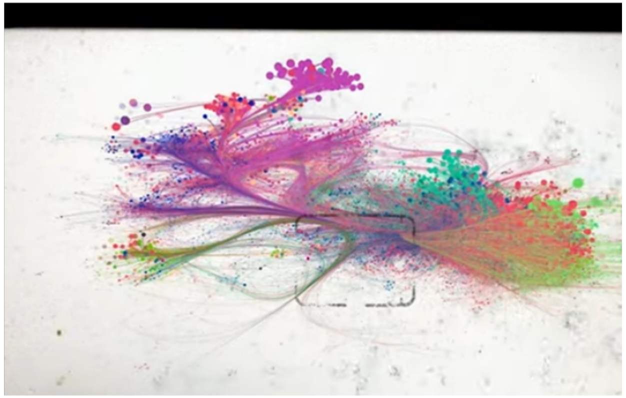
https://forms.gle/9xEYhzv2U5NrPQdH8

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



Video (02:17-03:15)
by Albert-László Barabási (cont.)





https://www.youtube.com/watch?v=RfgjHoVCZwU
From "… in a random network, the average dominates."
To "… does not capture how networks form"

Summary

Things to remember

- The ER model
- Degree distribution in 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

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