Applied economic networks - III

Ugo Bolletta

RITM, Université Paris-Saclay

October 4, 2021

List of papers for presentations

Down below you can find the link to the spreadsheet to book your slot:

`https://docs.google.com/spreadsheets/d/1PonIAbB4KMLvL2-CCHzDV2uPaQG0vWNZJup8Nzh0mAo/edit?usp=sharing`

- Presentations are **15 minutes each**.
- The main goal of the presentation is that you convey the main message of the paper: everyone should exit the room with something learned through you.
- Time is limited so focus on central aspects, detect and report the details necessary for the overall understanding of the paper.
- Write me an email if you cannot find a slot (I counted 10 students, but I might have miscounted!)

Feel free to write me an email if you need any assistance during the presentation!

Today's class

Models of random network formation:

- Erdos-Renyi
- Albert-Barabasi
- Beyond random network formation

The random network model

Disclaimer

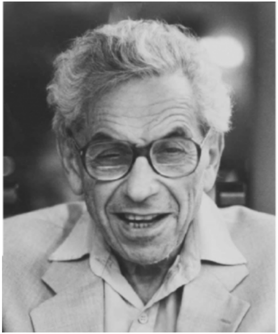
This lecture is inspired by ALBERT-LÀSZLÒ BARABÀSI's book:

http://networksciencebook.com/

https://barabasi.com/

The workhorse model: Erdos-Renyi

Some historical background:



- **Pál Erdos**: more papers than any other mathematician in history. Over 500 co-authors. *Erdos number*.
- **Alfréd Rényi**: brilliant mathematician with contributions in several disciplines, even outside mathematics.

The workhorse model: Erdos-Renyi

Random networks

A random network $G(N, p)$ consists of N nodes where each node pair is connected with probability p.

The workhorse model: Erdos-Renyi

Random networks

A random network $G(N, p)$ consists of N nodes where each node pair is connected with probability p.

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 number of links

Denote the number of links with L. We are interested in the expected number of links in a random network with fixed N and p.

The probability that the network has exactly L links

The product of 3 terms:

- 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.
- The probability that the remaining $N(N-1)/2 - L$ attempts have not resulted in a link, which is $(1-p)^{N(N-1)/2-L}$.
- The factor $\binom{\frac{N(N-1)}{2}}{L}$ counting the number of different ways we can place L links among $N(N-1)/2$ node pairs.

The number of links - II

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

The number of links - II

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

Which is a binomial distribution. Binomial

The number of links - II

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

Which is a binomial distribution. Binomial

Expected number of links

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

The probability that two nodes are connected times the total number of pairs.

Average degree

The average degree in the network $G(N, p)$ is:

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

Summary

The expected number of links is determined by N and p.
Increasing p:

- The average number of links increases linearly from $\langle L \rangle = 0$ to $L_{max} = N(N-1)/2$
- The average degree of a node increases from $\langle k \rangle = 0$ to $N-1$, the maximum number of links a single node can have.

Degree distribution

Outside the realm of expectations, in a specific realization some nodes can have lot of links and some none.

Degree distribution

The probability that a specific node has degree k.

Degree distribution

Outside the realm of expectations, in a specific realization some nodes can have lot of links and some none.

Degree distribution

The probability that a specific node has degree k.

Binomial or Poisson

While the exact form of the degree distribution is binomial, it depends on two parameters N, p, the *Poisson* distribution depends on one parameter only, the average degree $\langle k \rangle$.

Binomial vs. Poisson

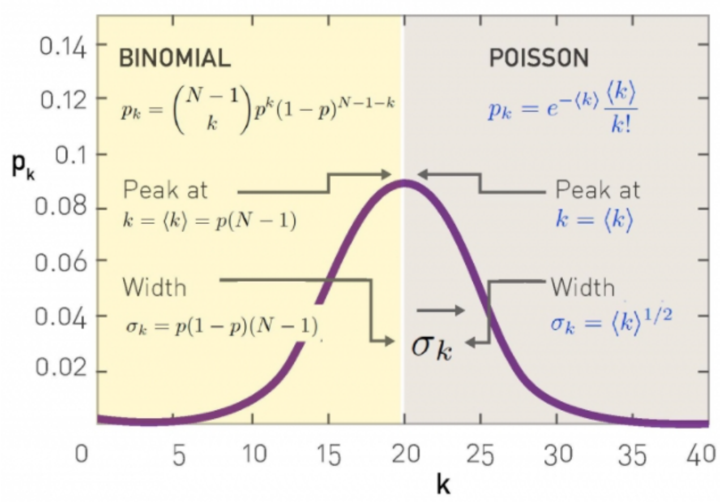


Figure: Source: Barabasi, A.L., *Network science*

Binomial vs. Poisson - II

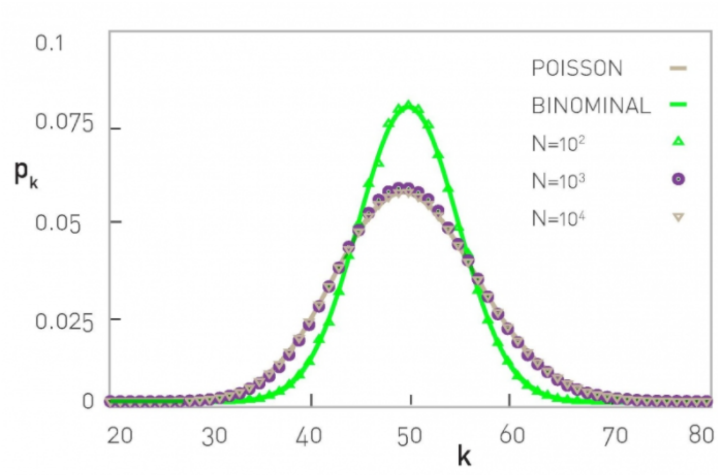


Figure: Source: Barabasi, A.L., *Network science*

Degree distribution of a random network

The main assumption that allows us to approximate the binomial distribution with a Poisson distribution is that $\langle d \rangle << N$.

$$p_k = Prob(\langle k \rangle = k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Degree distribution of a random network

The main assumption that allows us to approximate the binomial distribution with a Poisson distribution is that $\langle d \rangle << N$.

$$p_k = Prob(\langle k \rangle = k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Important remark

The Poisson distribution does not depend on N.
The features and properties of Poisson random networks are the same independently on how large they are, but only depend on the average degree.

The evolution of a random network

Now that we know the properties of random networks, we are interested in seeing how the random network change by gradually increasing p.
In particular we are interested in the **size of the largest component** N_G.

Two extreme cases:

- If $p = 0$, then $\langle k \rangle = 0$. The largest component has size $N_G = 1$, and $N_G/N \to 0$ with large N.
- If $p = 1$, then $\langle k \rangle = N - 1$. The network is **complete** and $N_G = N$. Also, $N_G/N = 1$.

Important thresholds

Sub-critical regime

$0 < \langle k \rangle < 1 \iff p < 1/N$

In the sub-critical regime the size of the largest cluster is still 0.

Important thresholds

Sub-critical regime

$0 < \langle k \rangle < 1 \iff p < 1/N$
In the sub-critical regime the size of the largest cluster is still 0.
If $p > 1/N^{3/2}$ the network has a component with at least 3 links.

Important thresholds

Sub-critical regime

$0 < \langle k \rangle < 1 \iff p < 1/N$
In the sub-critical regime the size of the largest cluster is still 0.
If $p > 1/N^{3/2}$ the network has a component with at least 3 links.

Critical regime

$\langle k \rangle = 1 \iff p = 1/N$
At the critical regime the network has a giant component.
$N_G/N \to N^{-1/3}$ as $N \to \infty$ (the growth rate of the largest component is still very slow).

Important thresholds - II

After the critical thresholds things go very fast.

Super-critical regime

$\langle k \rangle > 1 \iff p > 1/N$

Lots of small components (trees), cycles and loops in the giant component. The super-critical regime lasts until ALL nodes are absorbed in the giant component.

Connected regime

$\langle k \rangle > \ln N \iff p > \ln N/N$

The average degree required for the connected component depends on N. The network is still relatively sparse.

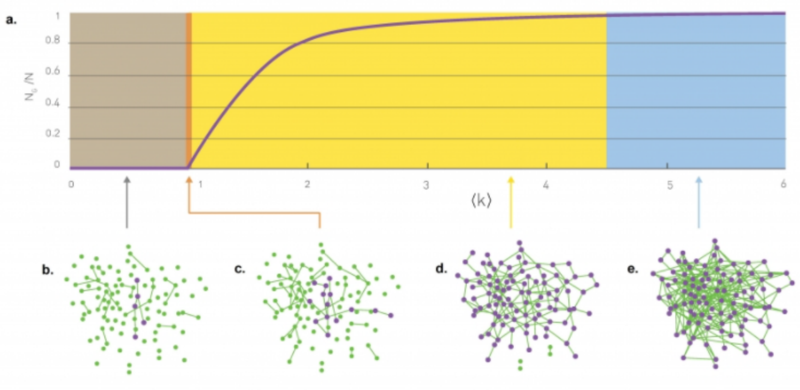


Figure: Source: Barabasi, A.L., *Network science*

Poisson distribution vs. real networks

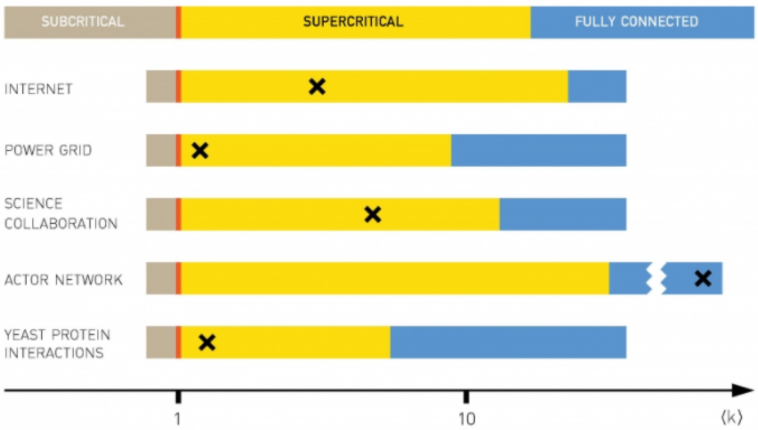


Figure: Source: Barabasi, A.L., *Network science*

Poisson distribution vs. real networks

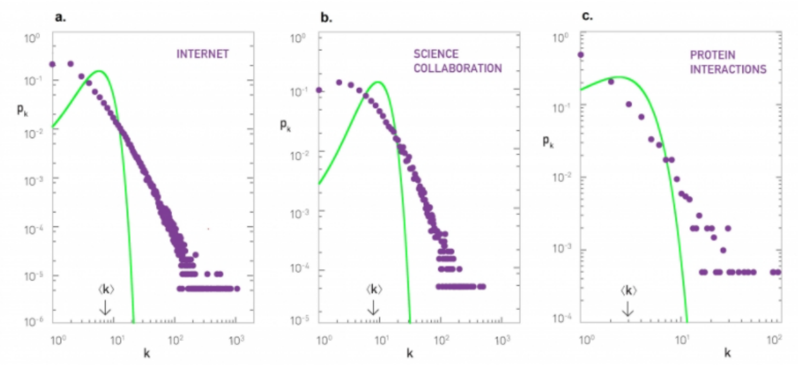


Figure: Source: Barabasi, A.L., *Network science*

What are we missing?

Main characteristics of real networks

- Small world (low diameter)
- High clustering
- Hubs (scale-free networks)

What are we missing?

Main characteristics of real networks

- Small world (low diameter)
- High clustering
- Hubs (scale-free networks)

Diameter in random networks

Consider a random network with average degree $\langle k \rangle$.

- $\langle k \rangle$ nodes at distance 1,
- $\langle k \rangle^2$ nodes at distance 2,
- $\ldots$
- $\langle k \rangle^d$ nodes at distance d

Diameter in random networks - II

So, the number of nodes at distance d are roughly:

$$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 + \cdots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$$

Diameter in random networks - II

So, the number of nodes at distance d are roughly:

$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 + \cdots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$
The maximal distance in the network is by definition

$N(d_{max}) \approxeq N$

Diameter in random networks - II

So, the number of nodes at distance d are roughly:

$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 + \cdots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$
The maximal distance in the network is by definition

$N(d_{max}) \approxeq N$
and assuming that $\langle k \rangle >> 1$

$\langle k \rangle^{d_{max}} \approx N$

Diameter in random networks - II

So, the number of nodes at distance d are roughly:

$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 + \cdots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$
The maximal distance in the network is by definition

$N(d_{max}) \approxeq N$
and assuming that $\langle k \rangle >> 1$

$\langle k \rangle^{d_{max}} \approx N$
so the diameter of random networks is approximately

$d_{max} \approx \frac{\ln N}{\ln \langle k \rangle}$

Scale-free property

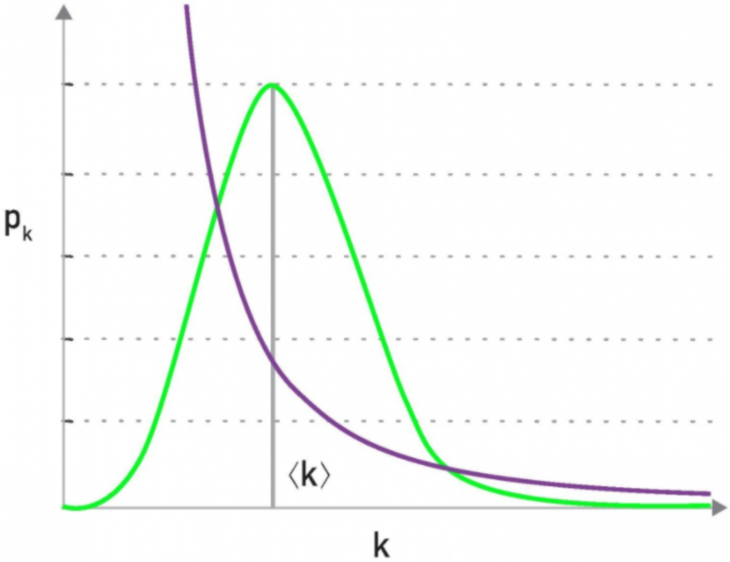
Random networks have a scale

In Poisson networks the network's nodes have a degree in the range $k = \langle k \rangle \pm \sigma_k$, where $\sigma_k = \langle k \rangle^{1/2}$

Scale-free networks

Scale free networks follow a power-law (fat tails). They are not "scaled" by the underlying distribution.
The node can have very small degree or very large degree.

Scale-free vs. Random networks



Albert-Barabasi model

Looking at the World Wide Web, the authors realized that random networks cannot represent real networks effectively.

Two new features

- **Growth**: at each time-step a new node is added, and it forms m new links (with older nodes).
- **Preferential attachment**: the probability that one of the m links is formed with a node i depends on the degree of node i: $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$

Albert-Barabasi model

Looking at the World Wide Web, the authors realized that random networks cannot represent real networks effectively.

Two new features

- **Growth**: at each time-step a new node is added, and it forms m new links (with older nodes).
- **Preferential attachment**: the probability that one of the m links is formed with a node i depends on the degree of node i: $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$

If two old nodes have degree 2 and 4, respectively, the probability to attach to the node with 4 is twice the probability to attach to the node with 2.

Evolution with preferential attachment

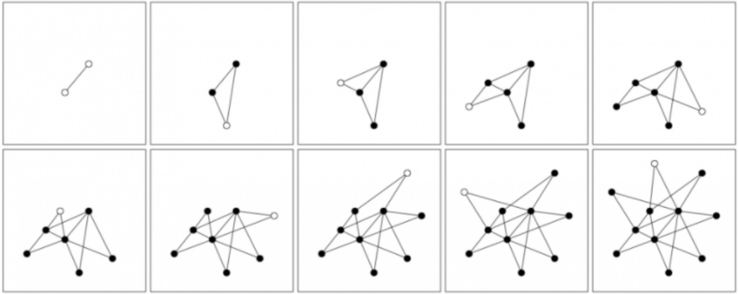


Figure: Source: Barabasi, A.L., *Network science*

Evidence of preferential attachment

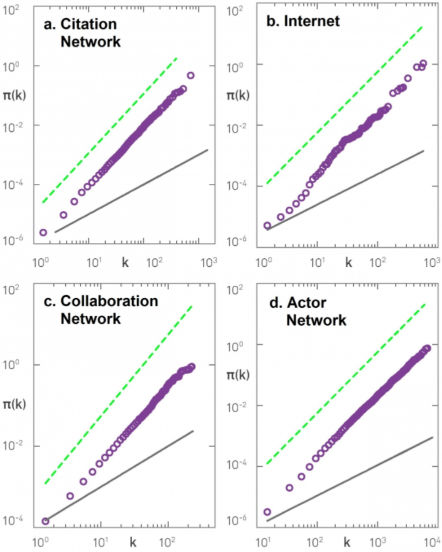


Figure: Source: Barabasi, A.L., *Network science*

Watts-Strogatz model

Another feature missing from Erdos-Renyi random networks is *clustering*.

The clustering in random networks is in the order of p.

Unless the average degree is very large, clustering tends to 0.

There is *some* clustering, but still incredibly small for large networks.

A simple model for clustering

Watts and Strogatz (1998) come to the rescue and propose the following model:

- We start with the **ring lattice** and then randomly pick some links to rewire.
- Initially the network exhibits high clustering, but low diameter.
- As we rewire *enough* links, we obtain low diameter.
- If we do not rewire too many links, clustering remains high.

Watts-Strogatz model

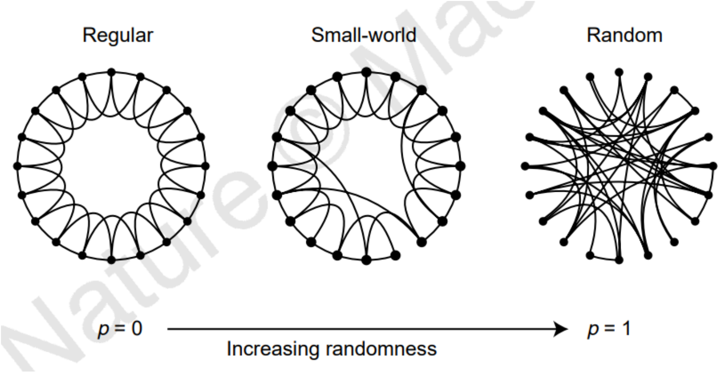


Figure: Source: Watts Strogatz, Nature, 1998

Watts-Strogatz model

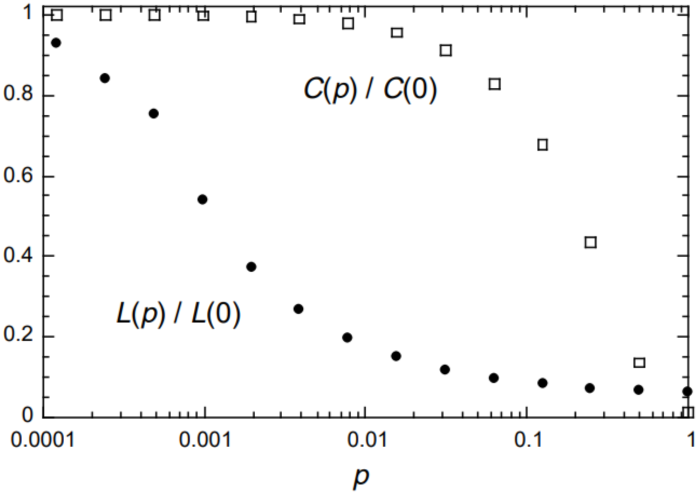


Figure: Source: Watts Strogatz, Nature, 1998

What about human networks?

We leave the floor to your colleague, talking about *Jackson, Rogers* AER 2007.

Thank you for your attention

The binomial distribution

A coin toss: the probability of *heads* (and *tails*) is 0.5.

Definition

The binomial distribution describes the number of successes in N independent trials with two outcomes, one with probability p, one with probability $1-p$.

Moments

- The distribution: $p_x = \begin{pmatrix} N \\ x \end{pmatrix} p^x (1-p)^{N-x}$
- First moment: $\langle x \rangle = \sum_{x=0}^{N} x p_x = Np$
- Second moment: $\langle x^2 \rangle = \sum_{x=0}^{N} x^2 p_x = p(1-p)N + p^2 N^2$
- Standard deviation: $\sigma_x = (\langle x^2 \rangle - \langle x \rangle^2)^{\frac{1}{2}} = [p(1-p)N]^{\frac{1}{2}}$

Go back