Complex Networks

Master of Science in Electrical Engineering

Erivelton Geraldo Nepomuceno

Department of Electrical Engineering
Federal University of São João del-Rei

April 30, 2019

3. Random Graphs

- The term random graph refers to the disordered nature of the arrangement of links between different nodes.
- The systematic study of random graphs was initiated by Erdös and Rényi in the late 1950s.
- We focus our attention on the shape of the degree distributions and on how the average properties of a random graph change as we increase the number of links.

3.1 Erdös and Rényi (ER) Models

- A random graph is a graph in which the edges are randomly distributed.
- In the late 1950s, two Hungarian mathematicians, Paul Erdös and Alfréd Rényi came up with a formalism for random graphs that would led to modern graph theory.
- We shall deal with undirected graphs.
- This method is called Erdös and Rényi (ER) random graphs.

290

On random graphs I.

Dedicated to O. Varga, at the occasion of his 50th birthday.

By P. ERDŐS and A. RÉNYI (Budapest).

Let us consider a "random graph" $\Gamma_{n,N}$ having n possible (labelled) vertices and N edges; in other words, let us choose at random (with equal probabilities) one of the $\binom{\binom{n}{2}}{N}$ possible graphs which can be formed from

Figure 1: [1] P. Erdos and A. Rényi, "On random graphs I," Publ Math, vol. 6, pp. 290–297, 1959. GS: 14388.

Definition 1 (ER model A: uniform random graphs)

Let $0 \leq K \leq M$, where $M = N(N-1)/2$. The model, denoted as $G^{ER}_{N,K}$, consists in the ensemble of graphs with N nodes generated by connecting K randomly selected pairs of nodes, uniformly among the M possible pairs. Each graph $G = (\mathcal{N}, \mathcal{L})$ with $|\mathcal{N}| = N$ and $K = |\mathcal{L}|$ is assigned the same probability.

- Consider N nodes and picks at random the first edge among the Mpossible edges, so that all these edges are equiprobable. Then one chooses the second edge at random with a uniform probability among the remaining $M - 1$ possible edges, and continues this process until the K edges are fixed.

Steps to construct a random network[1]

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.

[1] Source: Network Science Textbook - Albert-László Barabási. This is a textbook for network science, is freely available under the Creative Commons license.

- How many different graphs with N nodes and K edges are there in the ensemble?
- This number is equal to the number of ways we can select K objects among M possible ones, namely:

$$\frac{M!}{K!(M-K)!}$$

.

- This is known as binomial coefficient denoted as

$$C_M^K = \binom{K}{M}$$

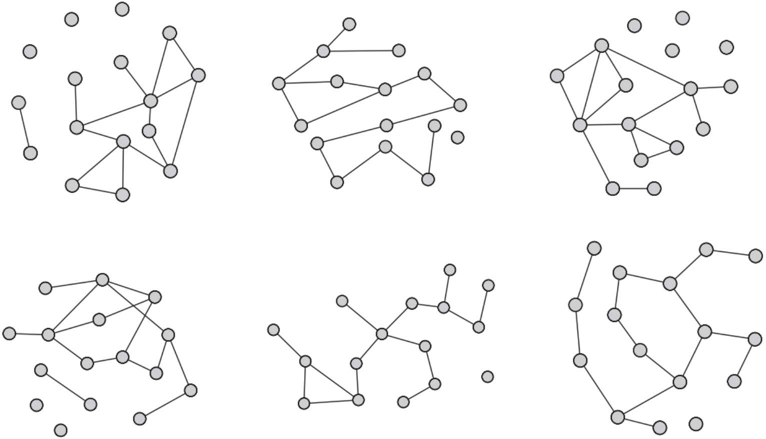

Figure 2: Six different realisations of model A with $N = 16$ and $K = 15$. $C_{120}^{15} \approx 4.73 \times 10^{18}$.

Definition 2 (ER model B: binomial random graphs)

Let $0 \leq p \leq 1$. The model, denoted as $G^{ER}_{N,p}$, consists in the ensemble of graphs with N nodes obtained by connecting each pair of nodes with a probability p. The probability P_G associated with a graph $G = (\mathcal{N}, \mathcal{L})$ with $|\mathcal{N}| = N$ and $|\mathcal{L}| = K$ is $P_G = p^K(1-p)^{M-K}$, where $M = N(N-1)/2$.

- From $P_G = p^K(1-p)^{M-K}$, we can calculate that the probabilities of having an empty or a complete graph with N nodes are respectively equal to $(1-p)^M$ and p^M.
- However, while all the graphs in the first ensemble have 15 edges, the number of links in the second ensemble.

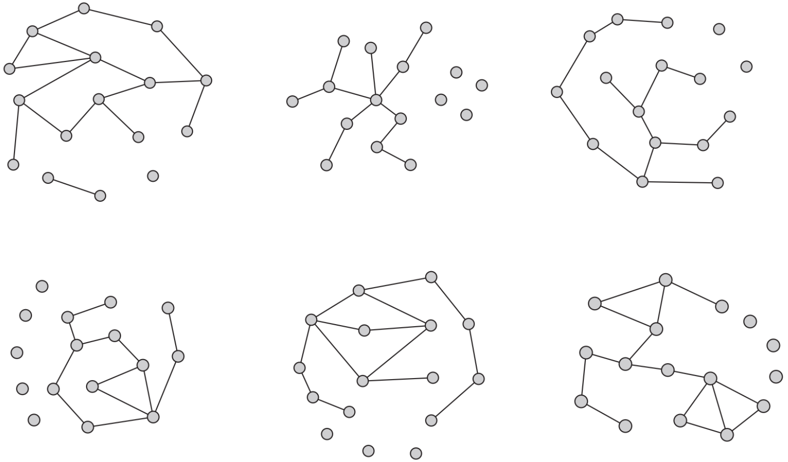

Figure 3: Six different realisations of model B, with $N = 16$ and $p = 0.125$. We have appropriately chosen the value of p, namely $pM = 15$. Thus, we have $G^{ER}_{N,p} \equiv G^{ER}_{N,K}$.

Box 1 (Binomial Distribution)

In statistics, the binomial distribution is valid for processes where there are two mutually exclusive possibilities. The binomial distribution is the discrete probability distribution of the number k of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p:

$$Bin(n, p, k) \equiv \binom{n}{k} p^k (1-p)^{n-k}$$

where k can take thevalues $1, 2, 3, \ldots, n$. A simple example is tossing ten coins and counting the number of heads.The distribution of this random number is a binomial distribution with $n = 10$ and $p = 1/2$ (if the coin is not biased). As another example, assume 5 per cent of a very large population to be green-eyed. You pick 100 people randomly.The number of green-eyed people you pick is a random variable which follows a binomial distribution with $n = 100$ and $p = 0.05$.

Example 1

Suppose a biased coin comes up heads with probability 0.3 when tossed. What is the probability of achieving $0, 1, \ldots, 6$ heads after six tosses?[a]

- $\Pr(0 \text{ heads}) = f(0) = \Pr(X = 0) = \binom{6}{0} 0.3^0 (1-0.3)^{6-0} = 0.117649$
- $\Pr(1 \text{ heads}) = f(1) = \Pr(X = 1) = \binom{6}{1} 0.3^1 (1-0.3)^{6-1} = 0.302526$
- $\Pr(2 \text{ heads}) = f(2) = \Pr(X = 2) = \binom{6}{2} 0.3^2 (1-0.3)^{6-2} = 0.324135$
- $\Pr(3 \text{ heads}) = f(3) = \Pr(X = 3) = \binom{6}{3} 0.3^3 (1-0.3)^{6-3} = 0.185220$
- $\Pr(4 \text{ heads}) = f(4) = \Pr(X = 4) = \binom{6}{4} 0.3^4 (1-0.3)^{6-4} = 0.059535$
- $\Pr(5 \text{ heads}) = f(5) = \Pr(X = 5) = \binom{6}{5} 0.3^5 (1-0.3)^{6-5} = 0.010206$
- $\Pr(6 \text{ heads}) = f(6) = \Pr(X = 6) = \binom{6}{6} 0.3^6 (1-0.3)^{6-6} = 0.000729$

[a] Source: https://en.wikipedia.org/wiki/Binomial_distribution

- Fluctuations of K in model B is

$$\sigma_K^2 = Mp(1-p).$$

- The ratio between the standard deviation σ_K and the average number of links $\bar{K} = pM$ is given by:

$$\frac{\sigma_K}{\bar{K}} = \sqrt{\frac{(1-p)}{pM}}.$$

- This proves that, in large graphs, the fluctuations in the value of K of model B can be neglected.

Example 2 (Fitting a real network with a random graph)

Suppose you are given a real world network G with N nodes and adjacency matrix $\{aij\}$ and you want to model it with a binomial random graph ensemble. What is the ensemble of graphs $G^{ER}_{N,p}$ that best approximates the real network? We can infer the best value of the parameter p from maximum likelihood considerations. It is useful to work with the logarithm of P_G, the so-called log-likelihood $\mathcal{L}(p)$ that the network G belongs to the ensemble:

$$\mathcal{L}(p) = \log P_G(p) = K \log p + [M - K] log(1 - p)$$

Maximising the log-likelihood with respect to p, i.e. solving the equation $d\mathcal{L}(p)/dp = 0$. The solution is $p = K/M = 2K/(N(N-1))$, where N and K are number of nodes and links in G.

3.2 Degree Distribution

- The nice property of random graphs is that they can be studied analytically.
- The degree distribution of ER random graphs can be easily derived analytically in model B in the following way.
- If p is the probability that there exists an edge between two generic vertices, the probability that a specific node i has degree k_i equal to k is given by the following expression:

$$\text{Prob}_{k_i=k} = C^k_{N-1} p^k (1-p)^{N-1-k} \quad 0 \leq k \leq N-1.$$

- The degree distribution in a random graph is a binomial distribution:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} = Bin(N-1, p, k) \quad k = 0, 1, 2, \ldots, N-1$$

- Average degree of a randomly chosen node in a random graph:

$$\langle k \rangle = \sum_{k=0}^{N-1} k p_k = p(N-1). \tag{1}$$

- The standard deviation around this quantity

$$\sigma_k = \sqrt{p(1-p)(N-1)} \tag{2}$$

- Using the Poisson distribution

$$p_k = \mathrm{Pois}(z, k) = e^{-z} \frac{z^k}{k!} \quad k = 0, 1, 2, \ldots \tag{3}$$

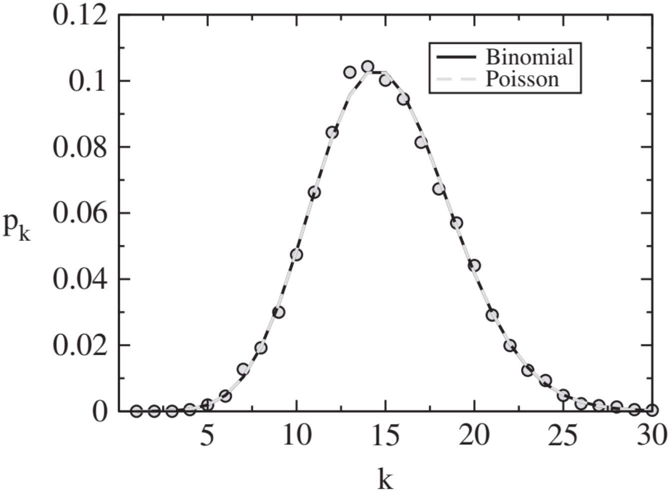

Figure 4: The degree distribution (circles) that results from one realisation of a random graph with $N = 10000$ nodes and $p = 0.0015$ is compared to the corresponding binomial (solid line) and Poisson distributions (dashed line).

3.3 Trees, Cycles and Complete Subgraphs

- What are the salient features of a typical random graph with N nodes and K edges?
-

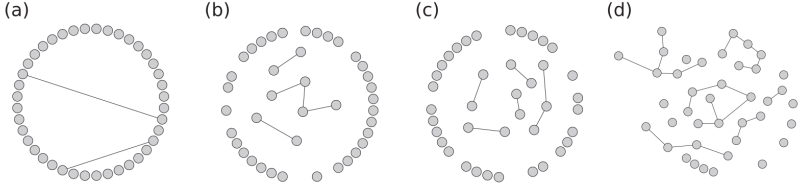

Figure 5: The typical graph of the ensemble $G^{ER}_{40,p}$ for four different values of p, namely $p = 2 \times 10^{-3}, 5 \times 10^{-3}, 8 \times 10^{-3}, 3 \times 10^{-2}$

- A Random Graph growth. [Video-05].
- GraphStream 1.0 - Dynamics graphs. [Video-06].

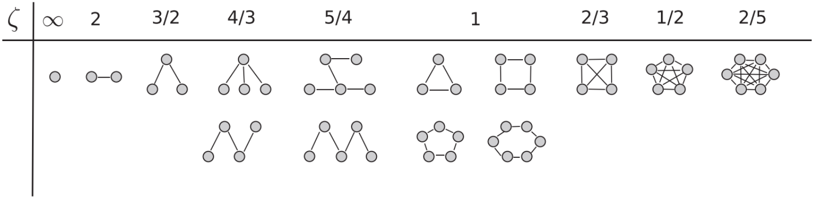

Figure 6: Threshold probabilities $p(N) \sim N^{-\zeta}$ for the appearance of different subgraphs in a random graph.

3.4 Giant Connected Component

- In their works, Erdös and Rényi also discovered a phase transition concerning the order of the largest component in the graph, namely the abrupt appearance of a macroscopic component known as a giant component.

Definition 3 (Giantcomponent)

A giant component in a graph G is a component containing a number of vertices which increases with the order of G as some positive power of N.

- They proved that a giant component appears at a critical probability function $p_c(N) = 1/N$. in model B, or analogously at a critical number of links $K_c(N) = N/2$ in model A.

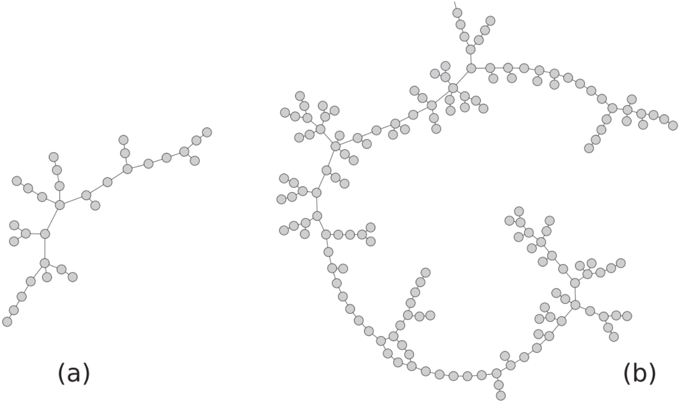

Figure 7: Plot of the largest component g_1 obtained in an ER random graph with $N = 1000$ nodes and with an average degree respectively below, $\langle k \rangle = 0.9$ (a), and above, $\langle k \rangle = 1.1$ (b), the critical value $\langle k \rangle = 1$. The largest component g_1 in the first case has $s_1 = 38$ nodes, while it contains $s_1 = 202$ nodes, i.e. 20 per cent of the graph nodes, in the second case.

Convertion of Vito's file to Pajek format

```
%Write a file.net for Pajek
%Number of Nodes
N=40;
%Probability
p=8e-3;
fileID = fopen('er.net','w');
fprintf(fileID,'*Vertices %d \n',N);
fprintf(fileID,'*Edges\n');
system(sprintf('er_B %d %f > teste.net',N,p))
A=load('teste.net');
E=[A+1 ones(length(A),1)];
fprintf(fileID,'%d %d %d\n',E');
fclose(fileID);
type er.net
```

Result

```
*Vertices 40
*Edges
6 38 1
6 40 1
10 13 1
10 37 1
14 15 1
16 36 1
24 27 1
```

Computational Exercise

1. Use the algorithm in slide 23 to reproduce the results of slide 19.
2. Use the option Fruchterman-Reingold to visualize your graphs.
3. Explore the phase transitions properties to find the giant component.
4. Develop a code to reproduce the results shown in slide 17.