Random Graphs (Part II)
CptS 591: Elements of Network Science, Spring 2021
Lecture Notes

Assefaw Gebremedhin

February 11, 2021

In the last lecture we introduced the notion of random graphs, and discussed things such as random variables and expectation, graph invariants in random graphs, and phase transition. In this lecture we look at how Erdos-Renyi random graphs behave in terms of: (i) degree distribution, (ii) average path length and (iii) clustering coefficient. These notes are more of an outline; some of the details that will be covered in class are left out. You are encouraged to take your own notes in class to fill in for those parts.

Number of edges in random graph $G(n, p)$

Consider the Erdos-Renyi random graph $G(n, p)$. The probability that $G(n, p)$ has exactly m edges (links) can be given by the expression:

$$p_m = \binom{\binom{n}{2}}{m} \cdot p^m \cdot (1-p)^{(n(n-1)/2-m)} \tag{1}$$

Note that the middle term accounts for the present edges, the last term accounts for the "absent" edges, and the first term to all the various combinations we can have.

Equation (1) is a binomial distribution. Great, but what exactly is a binomial distribution?

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 that of the other is $1-p$. The binomial distribution has the form:

$$p_x = \binom{N}{x} \cdot p^x \cdot (1-p)^{N-x} \tag{2}$$

The mean (expected value) of the distribution is:

$$\bar{x} = E(x) = \sum_{x=0}^{N} x p_x = Np \tag{3}$$

(notation: we denote in these notes average quantities with a bar over the variable name)

And the variance is:

$$Var(x) = Np(1-p) \tag{4}$$

Coming back to our network $G(n,p)$, the mean number of edges in it (now that we know it is a binomial distribution) is:

$$\bar{m} = E(m) = \sum_{m=0}^{n(n-1)/2} m p_m = p \cdot n(n-1)/2 \tag{5}$$

Interpretation: Equation (5) is telling us that the average number of edges in the E-R random graph is simply p times the maximum number of edges we could have, which makes intuitive sense.

Using (5), the average degree in the random network is:

$$\bar{d} = 2\bar{m}/n = p(n-1) \tag{6}$$

This again can be interpreted in a similar fashion as in (5): the average degree is p times the maximum we could have.

Concluding notes:
– the number of edges (links) in an E-R random network is not fixed, but varies between realizations.
– Its expected value, however, is determined by n and p.

Degree distribution

The probability that node i has exactly k links is given by:

$$p_k = \binom{n-1}{k} \cdot p^k \cdot (1-p)^{n-1-k} \tag{7}$$

The average degree is then

$$\bar{d} = \bar{k} = E(k) = (n-1)p \tag{8}$$

We just recovered the same expression we had seen earlier (in (6)).

Most real networks are sparse ($\bar{k} \ll n$), so the binomial distribution in (7) can be approximated by the *Poisson distribution*;

$$p_k = e^{-\bar{k}} \cdot \frac{\bar{k}^k}{k!} \tag{9}$$

The binomial and Poisson distribution describe the same quantity; they share several common properties:

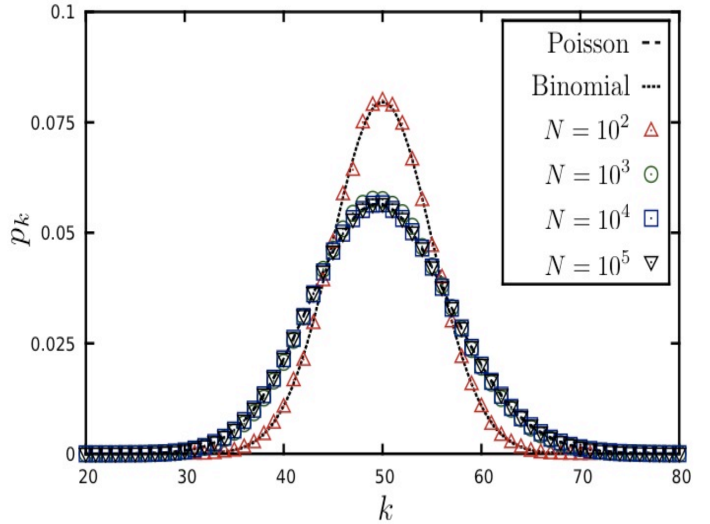
- both distributions have a peak around $\bar{k}$
- the width (dispersion) of the distribution is controlled by p or $\bar{k}$. In particular, the width in a binomial distribution is $\sigma_k = p(1-p)(n-1)$ and that in Poisson distribution is $\sigma_k = (\bar{k})^{1/2}$.

The plot in Figure 1 contrasts binomial distribution with Poisson. Although only an approximation, the Poisson distribution is the most commonly used distribution to describe large E-R random graphs in the literature on network theory. For small n, however, the degree distribution of a random network deviates significantly from a Poisson distribution as the condition for the Poisson approximation $n \gg \bar{k}$ is not satisfied.

DEGREE DISTRIBUTION OF A RANDOM GRAPH

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

<k>=50



K

Network Science: Random Graphs

DEGREE DISTRIBUTION OF A RANDOM NETWORK

Exact Result
-binomial distribution-

Large N limit
-Poisson distribution-

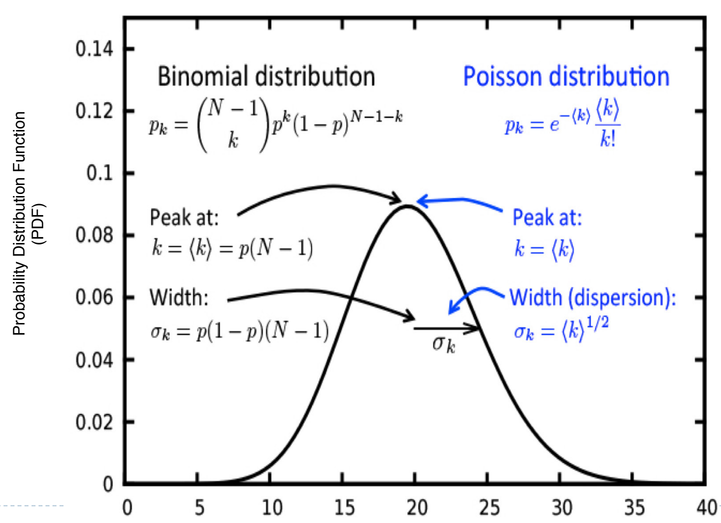


Figure 1: Binomial vs Poisson Distribution

Real networks vis-a-vis Poisson degree distribution

How big are the differences between node degrees in a particular realization of an E-R random graph?

In large enough a random graph, assuming the average degree $\bar{k}$ is about 1000, it can be shown, by mathematically analyzing the Poisson distribution, that the maximum degree k_{max} is approximately 1200 and the minimum degree is roughly 800. These three quantities are strikingly close to one another. In other words, in a random network, nodes have a comparable number of neighbors. Calculations show that in a large random network, the degree of most nodes is in the narrow vicinity of $\bar{k}$.

Real-world networks exhibit a categorically different behavior from random networks in terms of degree distribution. Most real networks instead typically have a Power-law degree distribution. Figure 2 compares Poisson degree distribution with a degree distribution of a typical real network.

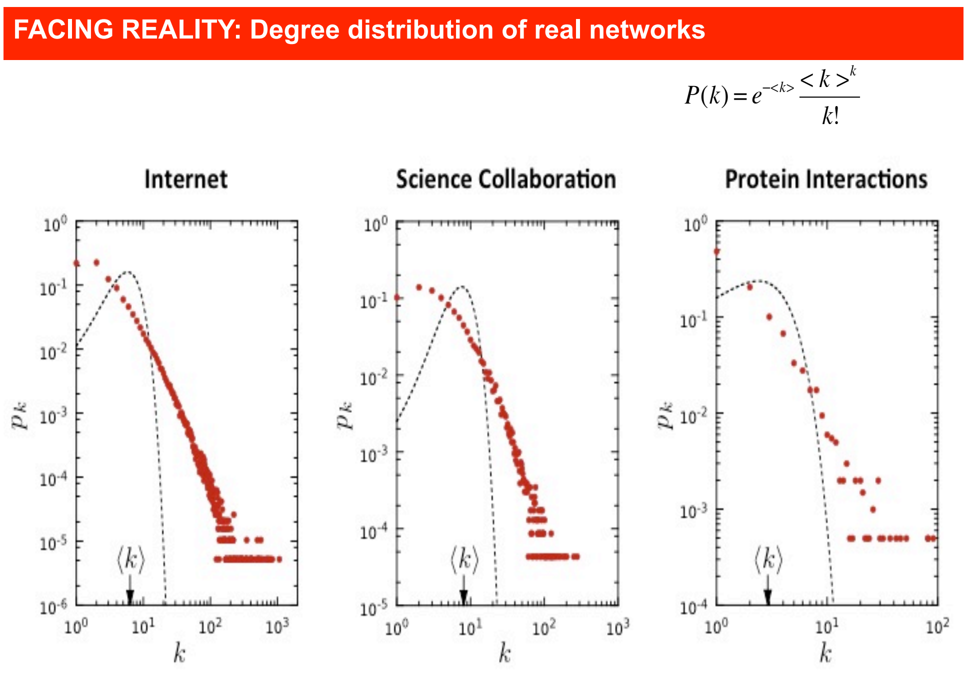


Figure 2: Poisson vs Real-world (Power Law) Degree Distribution

Average path length

Consider the random graph $G(n, p)$, average degree $\bar{k}$.

A node in this network has on average

$$\begin{gathered}\bar{k} \text{ nodes at distance } d = 1 \\ (\bar{k})^2 \text{ nodes at distance } d = 2 \\ \cdots \\ (\bar{k})^d \text{ nodes at distance } d.\end{gathered}$$

The expected number of nodes up to distance d from the starting node is then

$$N(d) \approx 1 + \bar{k} + (\bar{k})^2 + \ldots + (\bar{k})^d = \frac{(\bar{k}^{d+1} - 1)}{\bar{k} - 1} \tag{10}$$

But $N(d)$ can not exceed n, so $N(d_{max}) \approx n$. Assuming $\bar{k} \gg 1$,

$$\bar{k}^{d_{max}} \approx n \tag{11}$$

and

$$d_{max} \propto \log n / \log \bar{k} \tag{12}$$

Two points to make of (12):
(i) the diameter d_{max} increases as $\log n$, as opposed to n, and
(ii) the diameter is inversely related to average degree.
Together these points suggest small-world behavior.

Actually empirical evidences show that (12) predicts the average path length better than the diameter of a network.

Clustering coefficient

The local clustering coefficient c_i measures the density of links in node i's immediate neighborhood. In the extreme cases, $c_i = 0$ means no links at all among i's neighbors and $c_i = 1$ means i's neighbors are all linked to one another.

In an E-R random network, the expected number of links L_i between the k_i neighbors of node i is:

$$E(L_i) = \bar{L_i} = p \cdot \frac{k_i(k_i - 1)}{2} \tag{13}$$

Thus the local clustering coefficient is

$$c_i = \frac{2\bar{L_i}}{k_i(k_i - 1)} = p = \bar{k}/n \tag{14}$$

Interpretation: the local clustering coefficient of a node is independent of its degree. It is simply p for every node in the network. This becomes unsurprising once you recall how the E-R random graph is defined to begin with.

Summary

	E-R random networks	**Real-world networks**
Degree distribution	Binomial $\rightarrow$ Poisson	Power Law
Average path length	Scales as $\log n / \log \bar{k}$	Similar
Clustering coefficient	Independent of node degree	Decreases with degree
	$\bar{C}$ varies with $1/n$	Independent of n

References

(For both this and the last lecture).

Diestel, Graph Theory, Chapter 11.
Newman, Networks, Chapter 12.
Barabasi, Network Science, Chapter 3.

Credits

Figures 1 and 2 are taken from slides of Barabasi, Network Science.