University of Utrecht
Mathematics Department
Introduction to Complex Systems
November 2021
DEADLINE: 11 December

Project 2B

Complex Networks

In the present project you will try to analyse the *behaviour* of real networks and of the Watts-Strogatz model.

1. Real Networks

In the webpage `https://snap.stanford.edu/data/index.html` of the *Stanford Network Analysis Project* there several examples of real networks: for instance there are examples of social networks, communication networks, citation networks, collaboration networks, road networks etc...

For each network it is possible to download a text file with a list of the connections of each node of the network.

i) Choose two **undirected** networks from the list, trying to pick up (using the heuristics given in Lecture 8) one scale-free and one non scale-free. Try to explain the heuristics of your choice.

ii) For each of the networks determine the adjacency matrix A.

iii) Calculate the average clustering coefficient of each network (Hint: the clustering coefficient can be written in terms of A^3. In order to optimize the computing time one could use a fast matrix multiplication algorithm, like the Strassen algorithm, or by using sparse matrix multiplication).

iv) Estimate the degree distribution of the networks and plot the probability mass functions p_m ($p_m := \mathbb{P}(K = m)$, where K is the degree of the node). Plot a log-log plot as well. Discuss your findings.

v) Calculate the average degree of the neighbors of a randomly chosen node in one network of your choice. Compare the result with the average degree of the network. Can you observe *the friendship paradox*, i.e. on average, your friends have more friends than you do?

vi) (BONUS PART) Consider a general graph $G(V, E)$ with no isolated vertices. Pick a random vertex $v_1 \in V$ by first picking uniformly at random a vertex $v_0 \in V$, and then by choosing uniformly at random v_1 among the neighbors of v_0. Show that:

$$\mathbb{E}(k_{v_1}) \geq \mathbb{E}(k_{v_0})$$

where we denote with k_v the degree of the vertex $v \in V$.

vii) (BONUS PART) Assume first that the degree distribution is Poisson. Estimate with the method of Maximum likelihood the parameter λ of the distribution. Then, perform a *goodness of fit* test for testing the hypothesis of being the data Poisson distributed.

viii) (BONUS PART) Assume then that the degree distribution is a power law. Estimate with the method of Maximum likelihood the slope and the intercept of the log of the degree distribution Then, perform a *goodness of fit*: can we reject the hypothesis of the network being *scale-free*?
If you are more picky from a statistical point of view, you might want to estimate the tail exponent with a more refined methodology. From the paper `Scale-Free Networks Well Done`, you can use a non-parametric estimator as the Hill's estimator (formula B7), or the smooth Hill's estimator (formula B9) with optimal κ estimated by *bootstraping* (Appendix C).

2. Watts-Strogatz model

Consider the Watts-Strogatz model introduced in Lecture 8 (see the paper `Collective dynamics of small-world network` for a formal definition of the random graph). We recall that a realization of a Watts-Strogatz random graph $WS(N, 2r, p)$ can be obtained following the algorithm:

1. N nodes arranged in a circle.
2. Each node is linked to its $2r$ neighbors on the circle, r clockwise, r anticlockwise.
3. Going through each node one after the other, each edge going clockwise is *rewired* towards a randomly chosen other node with probability p, with duplicate edges forbidden.

Each vertex i on the circle can be seen as a point in $\{1, 2, , \ldots, N\}$. We use the following metric for measuring distances between two points i and j on the circle:

$$d_2^N(i, j) := \min(|i - j|, N - |i - j|)$$

i) Let us take three realizations of a WS graph with different p: in particular let us take $N = 500$, $r = 5$ and $p \in \{0.2, 0.4, 1\}$.

- For each of the graphs derive and plot the empirical degree distributions $\hat{p}_m(p)$. Can you say *a priori* how much is p_m, for $k \in \{1, 2, 3, 4\}$?
- (BONUS PART) We try now to derive an analytical expression for $p_m(p) := \mathbb{P}_p(K = m)$. Since r of the initial $2r$ (with $p = 0$) connections of each vertex are left untouched by the construction of the WS graph, the degree of each vertex i, can be written as $k_i = r + n_i$, with $n_i \geq 0$. The quantity n_i can also written as $n_i := A_i + B_i$, where $A_i \leq r$ have been left in place, each one with probability $1 - p$, and the other $B_i = n_i - A_i = k_i - m - A_i$ links

have been reconnected from other nodes towards i, each with probability of order p/N, for N large. Show that for large N (assuming valid the Poisson approximation of a Binomial distribution) we have:

$$p_m(p) := \mathbb{P}_p(K = m) \approx \sum_{n=0}^{\min(m-r,r)} \binom{r}{n} (1-p)^n p^{r-n} \frac{(rp)^{m-r-n}}{(m-r-n)!} e^{-pr}, \quad \text{for } m \geq r. \tag{1}$$

- Draw the three degree distributions $p_m(p)$ of the previous point and discuss their agreement with $\hat{p}_m(p)$.
- By using the expression in (1), show that

$$\lim_{p \to 1} \mathbb{P}_p(K = m) = \frac{r^{m-r}}{(m-r)!} e^{-r}$$

which is a Poisson distribution for the variable $K-r$. Draw this Poisson degree distribution and discuss its agreement with the empirical $\hat{p}_m(1)$ obtained in the first bullet. Draw also the Poisson degree distribution for an Erdos-Renii random graph with the same average distribution of a WS(500,10,1). Discuss your findings.

ii) For a graph $WS(N, 2r, 0)$ (i.e. the probability of *rewiring* p is zero), prove that the clustering coefficient C_0 is:

$$C_0 = \frac{3r-3}{4r-2}$$

iii) For a general graph $WS(N, 2r, p)$, with $p \neq 0$, the calculation of the clustering coefficient C_p is a bit more involved. However, it can be proven that:

$$C_p = C_0(1-p)^3 \tag{2}$$

Could you explain the probabilistic intuition behind (2)? (Hint: which is the probability that a triangle existing at $p = 0$ is present after the rewiring?)

iv) We want to find the *mean shortest path distance* ℓ of the Watts-Strogatz graph. Let us denote with $\ell(i, j)$ the distance $d_2^N(i, j)$ between two vertices i and j belonging to a $WS(N, r, p)$. Hence:

$$\ell := \frac{2}{N(N-1} \sum_{i,j} \ell(i.j)$$

A *mean-field* treatment of the model showed that this quantity can be approximated by:

$$\ell \approx \frac{N}{r} f(Nrp)$$

with

$$f(x) = \frac{1}{2\sqrt{x^2+2x}} \tanh^{-1} \sqrt{\frac{x}{x+2}}$$

By using the identity $\tanh^{-1}(x) = \frac{1}{2} \log \frac{1+x}{1-x}$, show that for N large enough, there is a regime of the parameter p where the WS graph has *small world properties* (i.e. $\ell \sim \log(N)$)