Aalto University
CS-E5740 Complex Networks, Autumn 2023
Hiraoka, Kivelä, Korhonen
Exercise set #2

Exercise set #2 (21 pts)

- The deadline for handing in your solutions is September 25th 2023 23:59.
- There are multiple ways for submitting your solutions. Check each question for details.
- Return also one `.zip` file containing all your Python code of the round in MyCourses.
- Check also the course practicalities file in MyCourses for more details on submitting your solutions.

Please submit all your solutions for round 2 to MyCourses as a single pdf report.

1. Degree distribution of Erdős-Rényi networks (5 pts)

The Erdős-Rényi (ER) model is a model for generating random networks where N nodes are randomly connected such that the probability that a pair of nodes is linked is p. In this exercise, we are going to study the degree distribution of an ER network.

a) (1 pt) Create an instance of the ER network with $N = 1000$ nodes and $p = 0.01$ and visualize it.

Hints:

- NetworkX has several functions that generates ER networks, but for a sparse (small p) network, you can use `nx.fast_gnp_random_graph(N, p)` function. You can also pass the Generator to the function as `nx.fast_gnp_random_graph(N, p, seed=rng)`.
- Use `nx.draw()` function for visualization.
- In most cases, you should get a connected network. However, there is a chance that the generated network is not connected. In this case, run the code again until you get a connected network.

b) (2 pt) Plot the degree distribution of the generated ER network. First, get the histogram of the degree and normalize it to transform it into a probability mass function.

c) (2 pt) Plot the Poisson distribution expected for an ER network and the degree distribution of the realized ER network in one plot.

Hints:

- An ER network with large N is expected to have the Poisson degree distribution with mean degree Np.
- Poisson distribution with mean λ is given by the following formula:

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

- You can use the function `scipy.special.factorial(k)` for calculating the factorial $k!$.

Challenge exercise (3 pts) (pen and paper)

Strictly speaking, the degree of an ER network is binomially distributed. Show that the Poisson distribution is the limiting case of the binomial distribution for large N.

2. Erdős-Rényi graph ensemble (5 pts)

The Erdős-Rényi (ER) model is a random graph model, meaning that it generates a network from a set of networks, rather than a single deterministic network. The set of all graphs a model may generate is called a *graph ensemble*. More precisely, a graph ensemble is a distribution of graphs where each graph G_i has a certain probability π_i. We let $G(N, p)$ denote the ensemble of graphs generated by the Erdős-Rényi model with N nodes and edge probability p.

Let's say we have a variable X whose value is a function of network G. We can compute X for each network, but we may be more interested in the expected value of X across the ensemble generated by a model. The expected value, or *ensemble average*, of quantity X is defined as $\langle X \rangle = \sum_i \pi_i X(G_i)$.

Let us define the following quantities for graph G: the average degree $k(G)$, the diameter of the largest connected component $d^*(G)$, and the average clustering coefficient $c(G)$ (assuming that the clustering coefficient equals to 0 for nodes of degree 0 and 1).

a) (2 pt) Calculate, using pen and paper, the formulas for expected average degree $\langle k \rangle$ and expected diameter $\langle d^* \rangle$ for the ER model $G(N = 3, p)$. Remember to simplify the formulas you get as results.

b) (1 pt) For ER models with sufficiently large p, the expected value of the average clustering coefficient $\langle c \rangle$ equals p. Explain why this is the case.

   Hints:

   – You don't need to provide a detailed mathematical proof. A short paragraph of clear reasoning is enough.

   – In practice, this property holds even for a small value of p if N is large.

c) (2 pt) Explain what happens to $\langle c \rangle$, if $N \to \infty$ with $\langle k \rangle$ bounded.

   Hints:

   – Mathematically, if x is 'bounded' then x is always smaller than some possibly big, but finite, number M.

   – Think about the implications of N growing at a significantly faster pace than k.

3. Implementing the Watts-Strogatz model (6 pts)

In this exercise, you will implement the Watts-Strogatz (WS) small-world model, which is a simple network model that yields a small diameter as well as a high level of clustering. In practice, the WS model is a one-dimensional ring lattice where some of the links have been randomly rewired. The model has three parameters: network size N, m (each node on the ring connects to m nearest neighbors to the left and m nearest neighbors to the right), and p, the probability of rewiring one end of each link to a random endpoint node.

a) (3 pt) Implement the WS model and generate a network using two different sets of parameters: $N = 30$, $m = 2$, $p = 0.2$, and $N = 80$, $m = 2$, $p = 0.4$. Visualize the networks using a circular layout algorithm (`nx.draw_circular(G)`), and check that the networks look right. For each network, report the total number of links and the number of rewired links. Note that NetworkX has a ready-made function for the WS model. However, the task is to program your own function, so do not use it (except for checking results, if in doubt).

b) (3 pt) Compute, for WS model networks with $N = 1000$ and $m = 4$, *relative* clustering coefficient $c(p)/c(p = 0)$ and *relative* average path length $l(p)/l(p = 0)$. Plot them as a function of p for $p = 0.001, \ldots, 1$ in one figure. Use a logarithmic x-axis in your plot.

   Then answer the following questions:

   – Are your results in line with the plots in the lecture slides?
   – Why does the clustering coefficient decrease as the probability increases?
   – What happens to the average path length? Why?

   Hint: See the code template on how to generate logarithmically evenly spaced values using `np.logspace()`.

Challenge exercise (2 pts)

You probably observe that your plots in b) are not exactly the same as the one in the lecture slides. Explain why this is the case. Based on your argument, improve your code and plot again $c(p)/c(p = 0)$ and $l(p)/l(p = 0)$ as a function of p.

Feedback (1 pt)

To earn one bonus point, give feedback on this exercise set and the corresponding lecture latest two days after the exercise round's submission deadline. You can find the feedback form in MyCourses.

References