Networks theoretical practice problems no.1.

1. During the lectures we have learned that the **adjacency matrix** provides a natural mathematical representation for a network or a graph. Based on A_{ij} we can calculate quite a few quantities describing various properties of the studied network. Here are two simple examples:

   a) Provide a formula giving the total number of links as a function of the adjacency matrix, (treating the directed and undirected cases separately).

   The non-zero entries of the matrix are corresponding to the links, thus,

   $$M = \frac{1}{2}\sum_{i,j} A_{ij} \quad \text{for undirected networks,}$$
   $$M = \sum_{i,j} A_{ij} \quad \text{for directed networks.}$$

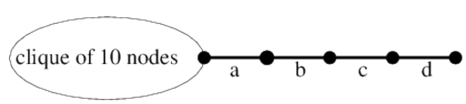
   b) Provide a formula giving the total number of paths of length ℓ between nodes i and j as a function of the adjacency matrix.

   The intuitive meaning of matrix multiplication in case of the adjacency matrix is „walking" on the links. Therefore, the number of paths length ℓ between nodes i, j can be expressed as

   $$\left[\mathbf{A}^{\ell}\right]_{ij}.$$

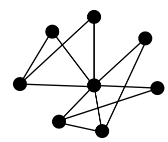
2. We have learned how to characterize nodes (or links) in a network by different measures such as centralities, clustering coefficient, etc. Let us practice the evaluation of these quantities with the help of the following exercises.

   a) Which of the shown links has the highest betweenness centrality? Why?



   The link a has the highest betweenness, since all shortest paths between the members of the clique and the 4 nodes shown that are not part of the clique have to pass through a. In contrast, only 3 of these shortest paths have to pass through b, only 2 have to pass through c, and only 1 through d.

b) Calculate the clustering coefficient for the central node:



The degree of the central node is 7*, and there are altogether* 5 *links between its neighbors, thus, its clustering coefficient is given by* $C = 5 \cdot 2/(7 \cdot 6) = 5/21$.

c) What is the average clustering coefficient in an undirected Erdős–Rényi graph with $N = 100$ nodes and $M = 990$ links?

The linking probability of this E-R graph can be expressed as

$$p = \frac{2M}{N(N-1)} = \frac{2 \cdot 990}{100 \cdot 99} = \frac{1}{5}.$$

Since the average clustering coefficient in the E-R model is equivalent to p*, we arrive to* $\langle C \rangle = p = 1/5$.

3. We have learned that the degree distribution is one of the most important distributions characterising the statistical properties of a network. What is the degree distribution for the following networks?

a) an undirected chain of 100 nodes,
There are 98 nodes in the chain with degree 2, whereas the 2 endpoints have degree 1, thus,

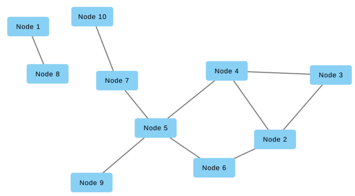
$$\begin{aligned} p(k=0) &= 0 \\ p(k=1) &= 0.02 \\ p(k=2) &= 0.98 \\ p(k \geq 3) &= 0 \end{aligned}$$

b) a balanced undirected binary tree of 127 nodes.
(In this tree we find always two „children" linked under a given node, except for the leafs at the bottom of the tree, which have no descendants at all.)
The root has degree 2, the 64 leafs have degree 1, and the rest of the nodes have degree 3, thus

$$\begin{aligned} p(k=0) &= 0 \\ p(k=1) &= 64/127 \\ p(k=2) &= 1/127 \\ p(k=3) &= 62/127 \\ p(k \geq 4) &= 0 \end{aligned}$$

4. Draw a small network having the following degree distribution:

$$
\begin{aligned}
p(k=1) &= 0.4 \\
p(k=2) &= 0.3 \\
p(k=3) &= 0.2 \\
p(k=4) &= 0.1
\end{aligned}
$$



5. The closeness centrality C_c we discussed on the lectures is not robust against the network falling to seprate (unconnected) components: The distance between node i and any other node j in a different component becomes $l_{ij} = \infty$, and thus, the closeness of i becomes $C_c(i) = 0$.

   Motivated by this, cook up an alternative definition of the closeness with the following properties:

   - it should give higher values for nodes intuitively closer to the rest of the network, i.e., if for node i the average distance to the rest of the nodes is smaller than for node j, then this implies $C_c(i) > C_c(j)$.
   - it should give a meaningful result for $C_c(i)$ even when there are unreachable nodes from i, (for which $l_{ij} = \infty$).

   A simple solution is to define C_c as

$$C_c(i) = \sum_{j \neq i} \frac{1}{l_{ij}}.$$

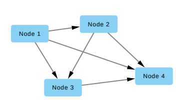
Another possibility is to use the original definition, but include only the nodes reachable from i when calculating the average distance, and then weight the result by the size of the component of i. This can be formulated as

$$C_c(i) = \frac{|S(i)|}{\sum\limits_{j \in S(i)} l_{ij}} \frac{|S(i)|}{N},$$

where $S(i)$ denotes the set of nodes reachable from i, (i.e., the component of i), and $|S(i)|$ is simply the cardinality of $S(i)$. The first fraction is corresponding to the inverse of the average distance of the other nodes in the component of i, whereas the second fraction gives the relative size of this component with respect to the total number of nodes in the network.

6. What is the maximum possible number of strongly connected components in a directed network of N nodes, where at least one link exists between every node pairs? (Note that even a single node can be considered as a strongly connected component). Please draw a figure of the network with the max. number of strongly connected nodes in case of $N = 4$ nodes.

   There maximum possible number of strongly connected components is N, where we have only a single link between any pair of nodes, pointing from the node with lower id. towards the node with higher id.



7. What is the probability distribution for the **number of links** in an undirected Erdős-Rényi graph with N nodes and p linking probability? (Equivalently:

what is the probability for having exactly M links in an Erdos-Rényi graph with N nodes and p linking probability?)

There are $\binom{N}{2} = N(N-1)/2$ *„places" to put the links, and they are introduced independently with uniform probability* p*. Thus, the distribution is binomial, and can be written as*

$$p(M) = \binom{\frac{N(N-1)}{2}}{M} p^M (1-p)^{\frac{N(N-1)}{2} - M}$$