CS-E5740 Complex Networks, Answers to exercise set 1

Tommaso Brumani, Student number: 100481325

September 14, 2022

Problem 1

a) The **adjacency matrix** A of a graph is a binary NxN matrix (with N being the number of nodes in the graph), in which the element a_{ij} is set to 1 if and only if the graph contains the directed edge $e_k = (v_i, v_j)$. If the network is undirected, then if $a_{ij} = 1$ also $a_{ji} = 1$.

For the graph in Figure 1, the adjacency matrix is:

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

b) The **edge density** ρ of a graph is the fraction of existing edges out of all the possible edges for the graph. Since a graph with N nodes can contain at most $\binom{N}{2}$ edges, if the graph actually contains m edges then we have:

$$\rho = \frac{2m}{N(N-1)}$$

For the graph in Figure 1, the edge density is:

$$\rho = \frac{2 * 9}{8 * (8 - 1)} = \frac{9}{28}$$

c) The **degree** k_i of a node i in a graph is the number of edges it is incident (directly connected) to, with each directed or undirected edge being counted once.

For the graph in Figure 1, the vector of node degrees $\mathbf{k}$ is:

$$\mathbf{k} = \begin{bmatrix} 1 \\ 1 \\ 3 \\ 5 \\ 2 \\ 3 \\ 1 \\ 2 \end{bmatrix}$$

d) The **average degree** $\langle k \rangle$ of a graph is the average degree of its nodes, obtained with the formula:

$$\langle k \rangle = \sum_{i=1}^{N} \frac{k_i}{N}$$

For the graph in Figure 1, the average degree is:

$$\langle k \rangle = \frac{1+1+3+5+2+3+1+2}{8} = 2.25$$

e) The **diameter** d of a graph is the maximum distance found between two of its vertices, meaning the length of the shortest path between the two.

For the graph in Figure 1, the diameter is the path(s) between nodes 5 and 7, whose length is:

$$d = 4$$

f) The **clustering coefficient** c_i of a node i in a graph is the ratio between the actual number of edges between its neighbors, and the maximum possible amount. It is obtained through the formula:

$$c_i = \frac{E_i}{\binom{k_i}{2}} = \frac{2E_i}{k_i(k_i - 1)}$$

For the graph in Figure 1, the vector of clustering coefficients $\mathbf{c}$ is (nodes with degrees equal to one were left blank):

$$\mathbf{c} = \begin{bmatrix} - \\ - \\ (2*2)/(3*2) \\ (2*1)/(5*4) \\ (2*1)/(2*1) \\ (2*2)/(3*2) \\ - \\ (2*0)/(2*1) \end{bmatrix} = \begin{bmatrix} - \\ - \\ 2/3 \\ 1/10 \\ 1 \\ 2/3 \\ - \\ 0 \end{bmatrix}$$

Problem 2

See Jupyter notebook for more details.

a) The visualized network is reported below

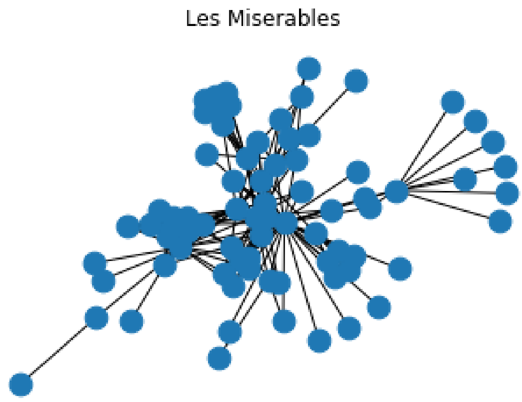

Fig. 1: visualized network.

b) The calculated density for the Les Miserables network is:

$$\rho = 0.08680792891319207$$

Which is the same as the one calculated by NetworkX.

c) The calculated average shortest path length is:

$$\langle l \rangle = 2.6411483253588517$$

d) The calculated average clustering coefficient is:

$$\langle c \rangle = 0.5731367499320135$$

e) The calculated degree and complementary cumulative degree distributions are reported below:

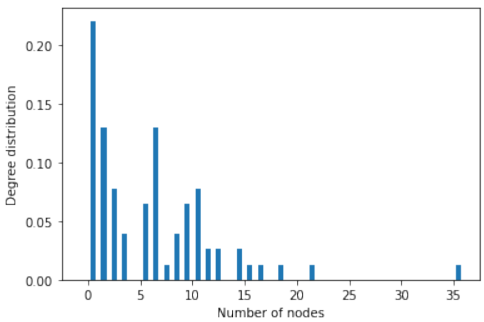

Fig. 2: degree distribution.

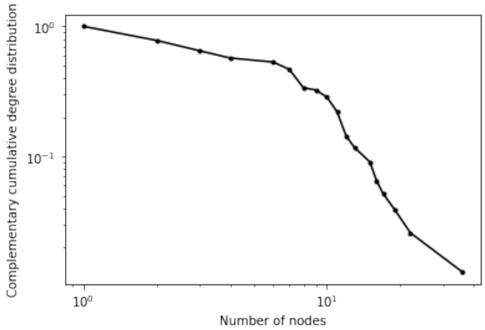

Fig. 3: ccd distribution.

Problem 3

a) The induced subgraph G^* is presented below:

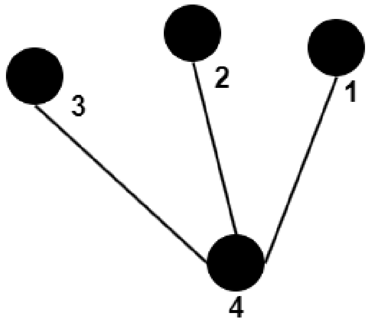

Fig. 4: visualized network.

There are a total of 12 walks of length 2, specifically:

- 1-4-1
- 2-4-2
- 3-4-3
- 1-4-2
- 2-4-1
- 2-4-3
- 3-4-2
- 3-4-1
- 1-4-3
- 4-1-4
- 4-2-4
- 4-3-4

The adjacency matrix A for G^* is:

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Thus, the squared adjacency matrix A^2 is:

$$A^2 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

We can notice that A^2 represents a sort of adjacency matrix for walks of length 2, where an element with non-zero value indicates the number of walks that can lead from the row node to the column node in precisely 2 steps. Summing up all elements in the matrix we obtain the number 12, which validates our previous counting of walks of length 2.

b) There are a total of 3 walks of length 3 leading from node 3 to node 4, and these are:

- 3-4-3-4
- 3-4-2-4
- 3-4-2-1

To compute $(A^3)_{3,4}$, we use row 3 and column 4 from A^2 and A respectively:

$$(A^3)_{3,4} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix} = 0*0+0*0+0*0+1*3 = 3$$

This corresponds to the number of walks of length 3 from node 3 to node 4 that were previously counted, validating the result.

c) To demonstrate that $(A^m)_{i,j}$ indicates the number of walks of length m between nodes i and j, we use mathematical induction.

- for $m = 1$: by definition of adjacency matrix, we have that $(A)_{i,j} = 1$ if and only if there is an edge connecting nodes i and j, which in turn means there exists a walk of length 1 connecting the two.
- assuming that the property holds true for a general m, we have for $m+1$ that $(A^{m+1})_{i,j}$ is given by the multiplication of row i and column j of A^m and A respectively. Since row i of A^m contains in each position the number of walks of length m leading from node i to another node k, and since column j of A contains the number of walks of length 1 leading from each node k to node j, the multiplication of these two vectors will generate non-zero addendums only when a path exists leading from i to k in m steps, and then from k to i in one step. Because in these instances the number of m length walks leading to k will be multiplied by 1, the resulting sum will be the sum of all walks leading from i to j in $m+1$ steps.

The hypothesis is thus proven correct.