CS-E5740 Complex Networks, Answers to exercise set 1

Alex Herrero Pons, Student number: 918697

September 21, 2020

Problem 1

a) The adjacency matrix is a square matrix that is used as a way to represent binary relationships. In the graph theory it's used to represent a graph showing which vertices are directly connected.

For the graph $G = (V, E)$ the adjacency matrix is

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

Having, for example, a 1 in position $(1, 4)$ means that there is an edge between nodes 1 and 4 and 0 otherwise.

b) The edge density of a graph is the ratio between the number of edges $|E|$ and the maximum number of possible edges.

For undirected simple graphs the edge density is

$$d = \frac{|E|}{\frac{|V|(|V|-1)}{2}} = \frac{2|E|}{|V|(|V|-1)},$$

hence, for the graph $G = (V, E)$ the edge density is

$$d = \frac{2|E|}{|V|(|V|-1)} = \frac{2 \cdot 9}{8 \cdot (8-1)} = \frac{18}{56} = 0.321428571.$$

c) The degree of a node is the number of edges connected to that node.

In graph $G = (V, E)$ the degree of each node would be

Node i	1	2	3	4	5	6	7	8
Degree k_i	1	1	2	5	3	3	2	1

Table 1: The degree k_i of each node $i \in V$ of $G = (V, E)$.

d) The mean degree of a graph $G = (V, E)$ is the average number of edges per node in the graph. That is, the number of edges $|E|$ divided per the number of nodes $|V|$. For the graph shown in this exercise the degree would be

$$\langle k \rangle = \frac{\sum_{i \in V} k_i}{|V|} = \frac{|E|}{|V|} = \frac{18}{8} = 2.25.$$

e) The diameter of a graph is the maximum distance between any pair of nodes. The diameter is defined as

$$d = \max_{v \in V} \epsilon(v)$$

where $\epsilon(v)$ is the eccentricity of v and it's defined as

$$\epsilon(v) = \max_{u \in V} dist(v, u).$$

where $dist(v, u)$ is the distance defined as the minimum path between v and u.

$$dist(u, u) = 0$$
$$dist(u, v) = dist(v, u)$$

node	1	2	3	4	5	6	7	8
1	0	2	2	1	2	2	3	4
2	2	0	2	1	2	2	3	4
3	2	2	0	1	1	2	3	4
4	1	1	1	0	1	1	2	3
5	2	2	1	1	0	1	2	3
6	2	2	2	1	1	0	1	2
7	3	3	3	2	2	1	0	1
8	4	4	4	3	3	2	1	0

Table 2: Distances between all nodes of graph $G = (V, E)$.

In the graph $G = (V, E)$ we have that the distances between nodes are the ones shown in Table 2. So, we can observe that diameter of $G = (V, E)$ is

$$d = dist(1, 8) = dist(2, 8) = dist(3, 8) = dist(8, 1) = dist(8, 2) = dist(8, 3) = 4$$

f) The clustering coefficient of a node quantifies how much clustered (or interconnected) it is with its neighbors. In other words, it quantifies how close its neighbours are to being a clique.

The clustering coefficient of a node i in an undirected graph is given by the proportion of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them. Therefore, we have that,

$$C_i = \frac{|\{e_{jk} : v_j v_k \in N_i, e_{jk} \in E\}|}{\frac{k_i(k_i-1)}{2}} = \frac{2|\{e_{jk} : v_j v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)},$$

where $N_i = \{v_j : e_{ij} \in E \vee e_{ji} \in E\}$ defines the neighbourhood (immediately connected neighbours) for a vertex v_i and k_i is the degree of v_i.

So, for the graph $G = (V, E)$ we have the following (Table 3) clustering coefficients for each node i that has degree $k_i > 1$.

$$C_3 = \frac{2|\{e_{jk} : v_j v_k \in N_3, e_{jk} \in E\}|}{k_3(k_3 - 1)} = \frac{2 \cdot 1}{2 \cdot (2-1)} = 1$$

$$C_4 = \frac{2|\{e_{jk} : v_j v_k \in N_4, e_{jk} \in E\}|}{k_4(k_4 - 1)} = \frac{2 \cdot 2}{5 \cdot (5-1)} = \frac{1}{5}$$

$$C_5 = \frac{2|\{e_{jk} : v_j v_k \in N_5, e_{jk} \in E\}|}{k_5(k_5 - 1)} = \frac{2 \cdot 2}{3 \cdot (3-1)} = \frac{2}{3}$$

$$C_6 = \frac{2|\{e_{jk} : v_j v_k \in N_6, e_{jk} \in E\}|}{k_6(k_6 - 1)} = \frac{2 \cdot 1}{3 \cdot (3-1)} = \frac{1}{3}$$

$$C_7 = \frac{2|\{e_{jk} : v_j v_k \in N_7, e_{jk} \in E\}|}{k_7(k_7 - 1)} = \frac{2 \cdot 0}{2 \cdot (2-1)} = 0$$

Node i	3	4	5	6	7
Clust. coeff. C_i	1	0.2	0.66	0.33	0

Table 3: The clustering coefficient C_i of each node $i \in V$ that has degree $k_i > 1$ of $G = (V, E)$.

Problem 2

a) For loading the Karate Club Network we used the `nx.read_weighted_edgelist()` function.

We can observe the split of the club in the elongated shape of the network shown in Figure 1.

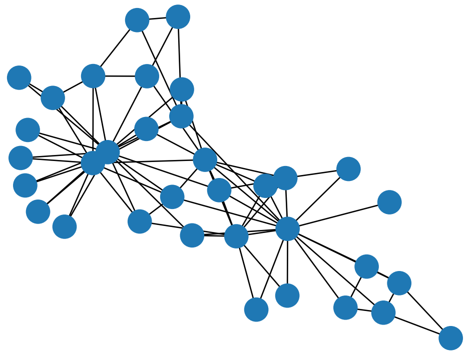

Figure 1: Visualization of the Karate Club Network.

b) To calculate the edge density of the graph we used the formula mentioned in the exercise 1b) as:

`E = nx.number_of_edges(network)`

`V = nx.number`$_o f_n odes(network)$

`D = 2*E/(V*(V-1))`

As we can see in Figure 2 both results (self-written algorithm and NetworkX function) are exactly the same.

```
~/.../ComplexNetworks/ES1 >>> python compute_network_properties.py
D from self-written algorithm: 0.13903743315508021
D from NetworkX function: 0.13903743315508021
<l> from NetworkX function: 2.408199643493761
C from NetworkX function: 0.5706384782076824
```

Figure 2: Output generated by executing `compute_network_properties.py`.

c) We used the `nx.average_shortest_path_length()` function to calculate the average shortest path length $\langle l \rangle$. The result is shown in the Figure 2.

d) For calculating the average clustering coefficient of the network we used the function `nx.average_clustering()`. The result is also shown in the Figure 2.

e) The results of the degree distribution $P(k)$ and the complementary cumulative degree distribution 1-CDF(k) of the network are shown in the Figure 3 and Figure 4 respectively.

   In both plots we can observe how most of the nodes have a low degree due to the split of the Karate Club. We can also observe how there are two nodes that have a very high degree, both will probably be the leaders of the club.

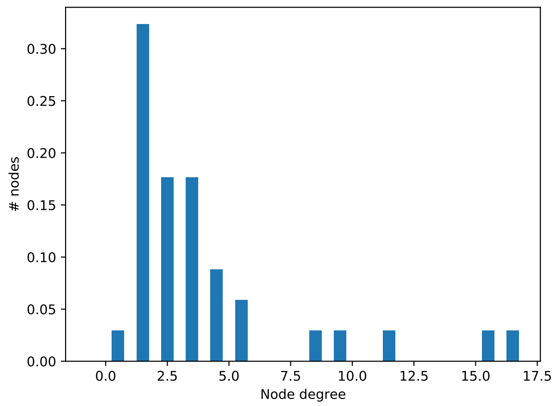

Figure 3: Degree distribution $P(k)$ of the network.

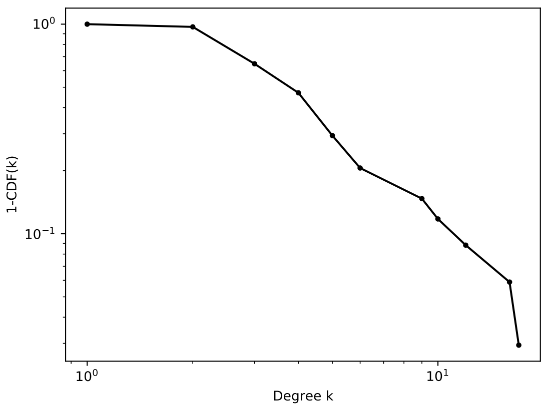

Figure 4: Complementary cumulative degree distribution 1-CDF(k) of the network.

Problem 3

a) In Figure 5 we can observe the induced subgraph $G^{>}*$ that is induced by vertices $V^* = \{1 \dots\}$ of the network $G = (V, E)$ from the exercise 1.

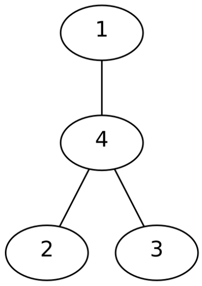

Figure 5: Induced subgraph G^*.

In this subgraph G^* we can find the following walks of length two between all node pairs (i, j), $i, j \in \{1, \dots, 4\}$ in G^*:

$$\{1 \to 4 \to 1\}, \{1 \to 4 \to 2\}, \{1 \to 4 \to 3\},$$
$$\{2 \to 4 \to 1\}, \{2 \to 4 \to 2\}, \{2 \to 4 \to 3\},$$
$$\{3 \to 4 \to 1\}, \{3 \to 4 \to 2\}, \{3 \to 4 \to 3\},$$
$$\{4 \to 1 \to 4\}, \{4 \to 2 \to 4\}, \{4 \to 3 \to 4\}.$$

As we can see, in total we have 12 walks of length two.
In the other hand, computing the A^2 matrix (shown below) we can observe that each element of the matrix $(A^2)_{u,v}$ indicates how many walks of length 2 there are from vertex u to vertex v (or vice versa). So if we sum all the elements of the matrix we obtain the total number of walks of length 2.

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

$$n\_walks(2)^1 = 1+1+1+0+1+1+1+0+1+1+1+0+0+0+0+3 = 12$$

[1] $n\_walks(m)$ represents the number of walks of length m.

b) We obtained the 3 following walks of length 3 from node 3 to node 4.

$$\{3 \to 4 \to 1 \to 4\}, \{3 \to 4 \to 2 \to 4\}, \{3 \to 4 \to 3 \to 4\}$$

By computing the value of $(A^3)_{3,4}$ we can easily obtain the same result.

$$n\_walks(3,4,3)^2 = (A^3)_{3,4} =$$
$$= (A)_{3,1} \cdot (A^2)_{1,4} + (A)_{3,2} \cdot (A^2)_{2,4} + (A)_{3,3} \cdot (A^2)_{3,4} + (A)_{3,4} \cdot (A^2)_{4,4} =$$
$$= 0 \cdot 0 + 0 \cdot 0 + 0 \cdot 0 + 1 \cdot 3 = 3$$

c) Let's consider a general network with adjacency matrix A. Show that the element $(A^m)_{i,j}, m \in N$ corresponds to the number of walks of length m between nodes i and j.
Prove by induction that $\forall m \in N : (A^m)_{i,j} = n\_walks(i,j,m)$.

$$(A^m)_{i,j} = (A)_{i,1} \cdot (A^{m-1})_{1,j} + \cdots + (A)_{i,k} \cdot (A^{m-1})_{k,j} =$$
$$= \sum_{l=1}^{k} (A)_{i,l} \cdot (A^{m-1})_{l,j} = n\_walks(i,j,m)$$

where k is the number of nodes of the network.
Base case $(m = 1)$**:** By definition of the adjacency matrix, the element $A_{i,j}$ represents the number of direct connections between the nodes i and j and that is exactly the same as saying the number of walks of length 1 (m) between nodes i and j.
Inductive step:
Assume $(m = n)$ holds: So that $(A^n)_{i,j} = n\_walks(i,j,n)$ is true.
Show that $(m = n + 1)$ holds:
We have that

$$(A^{n+1})_{i,j} = \sum_{l=1}^{k} (A)_{i,l} \cdot (A^{(n+1)-1})_{l,j} = \sum_{l=1}^{k} (A)_{i,l} \cdot (A^n)_{l,j} =$$

$$= \sum_{l=1}^{k} (A)_{i,l} \cdot n\_walks(l,j,n) = \sum_{l \in V_i} n\_walks(l,j,n)$$

where V_i is the set of nodes that are neighbors of i.
And that is totally correct because finding paths from i to j of length m is exactly the same as finding how many paths from l to j of length $m - 1$ being l a neighbor of i.

$\square$

[2] $n\_walks(i,j,m)$ represents the number of walks between nodes i and j of length m.