Exercise Sheet 2: Social and Graph Data Management

Pablo Mollá Chárlez

Contents

1 Exercise 1: Graph Measures

Consider the graph G in the following figure:

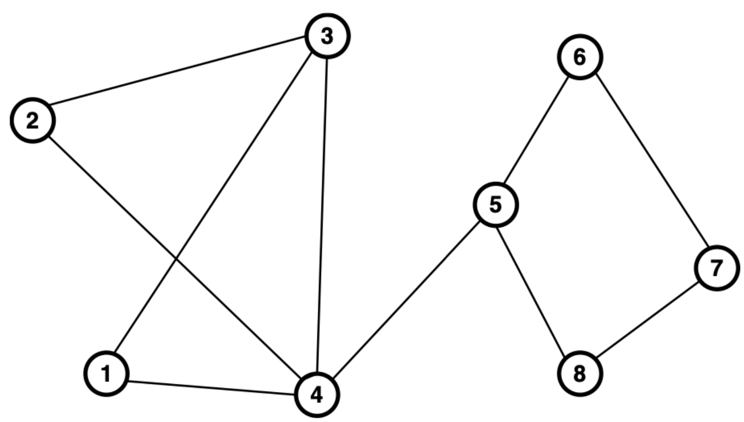

Figure 1: Graph G

- **Question 1:** Represent the graph as an adjacency matrix.
- **Question 2:** Write down the degree distribution of G, and the average degree $\langle k \rangle$.
- **Question 3:** Compute the clustering coefficient of node 4 in G. Explain how it is computed.
- **Question 4:** Compute the diameter d_{max} of G, and show a path of length dmax in G.
- **Question 5:** Assume that the graph was computed using a random network model with parameter p. What is the value of p? Explain how you found it.
- **Question 6:** Compute the Jaccard coefficient score for each of the node pairs $(1, 2)$ and $(3, 6)$.
- **Question 7:** Compute the inverse distance score for each of the node pairs $(1, 2)$ and $(3, 6)$.

1.1 Answers

Let's tackle each question step by step:

- **Question 1:** Represent the graph as an adjacency matrix.

  The graph has 8 nodes, and the adjacency matrix A is an 8×8 matrix where $A[i][j] = 1$ if there is an edge between nodes i and j, otherwise $A[i][j] = 0$, then :

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

- **Question 2:** Write down the degree distribution of G, and the average degree $\langle k \rangle$.

  The degree distribution can be extracted by counting how many $1's$ appear per row in the adjacency matrix A:

$$k_1 = 2; \quad k_2 = 2; \quad k_3 = 3; \quad k_4 = 4; \quad k_5 = 3; \quad k_6 = 2; \quad k_7 = 2; \quad k_8 = 2$$

  Therefore, the degree distribution of G is:

$$p_0 = 0; \quad p_1 = 0; \quad p_2 = \frac{5}{8}; \quad p_3 = \frac{2}{8} \quad p_4 = \frac{1}{8}$$

  Then, the average degree is:

$$\langle k \rangle = 0 \times p_0 + 1 \times p_1 + 2 \times p_2 + 3 \times p_3 + 4 \times p_4 = \frac{20}{8}$$

- **Question 3:** Compute the clustering coefficient of node 4 in G. Explain how it is computed.

  The clustering coefficient of a node i is computed as follows:

$$C_i = \frac{2 \cdot e_i}{k_i \cdot (k_i - 1)}$$

  Where e_i denotes the number of edges between neighbours of i and k_i is the degree of node i. Therefore, in our situation:

$$C_4 = \frac{2 \cdot e_4}{k_4 \cdot (k_4 - 1)} = \frac{2 \cdot 2}{4 \cdot (4 - 1)} = \frac{1}{3}$$

  Here, $e_1 = 2$, because the neighoubours of node 4 are $\{1, 2, 3, 5\}$ and only 2 edges are connecting any one of them directly (between nodes 2 and 3 and nodes 1 and 3 there is an edge, but between nodes $1, 2, 3$ and 5 there).

- **Question 4:** Compute the diameter d_{max} of G, and show a path of length d_{max} in G.

  The d_{max} is the longest shortest path between any two nodes in the graph, therefore the $d_{max} = 4$. One path with such length could be for instance the path that connects node 2 and node 7, which is:

$$2 \underbrace{\longrightarrow}_{1} 4 \underbrace{\longrightarrow}_{1} 5 \underbrace{\longrightarrow}_{1} 6 \underbrace{\longrightarrow}_{1} 7$$

  Another path with maximum distance could be:

$$3 \underbrace{\longrightarrow}_{1} 4 \underbrace{\longrightarrow}_{1} 5 \underbrace{\longrightarrow}_{1} 8 \underbrace{\longrightarrow}_{1} 7$$

- **Question 5:** Assume that the graph was computed using a random network model with parameter p. What is the value of p? Explain how you found it.

  The theory studied during the lectures tells us that the average degree $\langle k \rangle$ in a random network with size N nodes, is described as:

  $$\langle k \rangle = p \cdot (N-1) \longleftrightarrow p = \frac{\langle k \rangle}{N-1} = \frac{\frac{20}{8}}{8-1} = \frac{20}{56} = \frac{5}{14}$$

- **Question 6:** Compute the Jaccard coefficient score for each of the node pairs $(1,2)$ and $(3,6)$.

  The Jaccard coefficient for two nodes u and v is:

  $$J(u,v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

  where $N(u)$ and $N(v)$ are the neighbors of u and v, respectively. Therefore, with the given pairs:

  1. Pair $(1,2)$
     - **Neighbors of 1:** $\{3,4\}$
     - **Neighbors of 2**: $\{3,4\}$
     - **Intersection:** $\{3,4\}$ $(|N(1) \cap N(2)| = 2)$
     - **Union:** $\{3,4\}$ $(|N(1) \cup N(2)| = 2)$

     $$J(1,2) = \frac{2}{2} = 1.0$$

  2. Pair $(3,6)$
     - **Neighbors of 3:** $\{1,2,4\}$
     - **Neighbors of 6:** $\{5,7\}$
     - **Intersection:** $(|N(3) \cap N(6)| = 0)$
     - **Union**: $\{1,2,4,5,7\}$ $(|N(3) \cup N(6)| = 5)$

     $$J(3,6) = \frac{0}{5} = 0.0$$

- **Question 7:** Compute the inverse distance score for each of the node pairs $(1,2)$ and $(3,6)$.

  The inverse distance score for two nodes u and v is:

  $$S(u,v) = \frac{1}{d(u,v)}$$

  where $d(u,v)$ is the shortest path distance between u and v. For pair $(1,2)$, the shortest path distance $d(1,2) = 2 \implies S(1,2) = \frac{1}{2} = 0.5$. For pair $(3,6)$, the shortest path distance $d(3,6) = 3 \implies S(3,6) = \frac{1}{3} \approx 0.333$.