

Resources: This problem set contains 5 parts. All the relevant concepts can be found in the lecture and section notes. For further reading you may wish to consult Chapter 2 of Matt Jackson’s book.

1. Diameter and Average Path Length (Jackson 2.2.2).

- The *distance* between two nodes is the length of a geodesic between them – i.e. the smallest number of links in the graph needed to connect the two nodes. For example, the distance between the nodes b and e in network A in the figure at the bottom is 2. If there is no path between two nodes, we say that the distance between them is infinite.
- The *diameter* of a network is the largest distance between any two nodes in the network.
- The *average path length* between the nodes of a network is the arithmetic mean of the distances between all pairs of nodes in the network. For example, consider the following network:

$$(\{a, b, c\}, \{ab, ac\}).$$

The average path length is:

$$\frac{1}{3} [\text{distance}(a, b) + \text{distance}(a, c) + \text{distance}(b, c)] = \frac{1}{3} [1 + 2 + 1] = \frac{4}{3}.$$

- What is the distance between nodes e and i in network B ?
- What is the diameter of the following network?

$$(\{a, b, c, d, e\}, \{ab, ac, ad, de\})$$

- What is the average path length in the following network?

$$(\{a, b, c, d\}, \{ab, ac, ad\})$$

2. Cliquishness, Cohesiveness and Clustering (Jackson 2.2.3). Note we use the words “network” and “graph” interchangeably.

- A *complete subgraph* of a graph is a nonempty subgraph C with every two nodes in C linked. A *clique* of graph G is¹ a maximal complete subgraph of G . In other words, it is a complete subgraph C of G that is not a subgraph of any larger complete subgraph of G . For example, $(\{a, c\}, \{ac\})$ is a clique of network A but the following subgraph is not:

$$(\{b, d, e, f\}, \{bd, de, df, ef\})$$

- Represent a clique of network B via a picture.

¹Intuitively, cliques are the biggest completely connected pieces of a network.

- The most common way of measuring the cohesiveness of a network is based on transitive triples (or clustering). The most basic clustering measure is to perform the following exercise. Look at all situations in which two edges emanate from the same node (for example, edges ij and ik both involve node i) and ask how often the edge jk is then also in the network. We call this likelihood the *clustering coefficient*. For example, the clustering coefficient of network A is $\frac{3}{5}$. To see this, note that there are five pairs of nodes that are linked to a common node (bf , be , ef , df , de), and only three of these (de , ef , df) are themselves linked. In a friendship network, the clustering coefficient tells us how likely it is that two individuals with a common friend are friends themselves.²
- b. Compute the clustering coefficient of network B .

3. Graph Cuts and Expansion.

- Given a graph $G = (V, E)$, and two disjoint subsets of nodes $S, T \subset V$ with $S \cap T = \emptyset$, a *cut* $C(S, T)$ is the set of all edges that have one end point in S and another endpoint in T . For example, if we have a graph $G = (V, E)$ with set of nodes $\{u, v, w, x, y, z\}$ and the set of edges $\{uw, ux, vx, vz, wx, xy, yz\}$, for $S = u, v$ and $T = w, x, y$ the cut is $C(S, T) = \{uw, ux, vx\}$.
- The *expansion* of a graph $G = (V, E)$ is the minimum, over all cuts we can make of the number of edges crossing the cut divided by the number of vertices in the smaller half of the cut. Formally:

$$\min_{S \subseteq V} \frac{|C(S, V \setminus S)|}{\min\{|S|, |V \setminus S|\}}.$$

Show that the expansion of a graph can also be expressed as α , where:

$$\alpha = \min_{\substack{S \subseteq V, \\ 1 \leq |S| \leq \frac{n}{2}}} \frac{|e(S)|}{|S|}$$

where $e(S)$ is the number of edges leaving the set S .

4. Basics of Erdős-Rényi: Playing with Small Graphs. Consider the set of $n = 4$ nodes $V = \{a, b, c, d\}$. Suppose that an edge (i.e., link) between any two nodes i and j forms with probability $p \in (0, 1)$, independently of all other edges. This is the $G(n, p)$ Erdős-Rényi random graph with $n = 4$.

- For this graph on $n = 4$ nodes, what is the probability that the graph has (i) no edges, (ii) one edge and (iii) more than one edge? The only variable in your answer should be p .
- What is the expected degree of each node in $G(4, 0.5)$?
- Generate a random graph $G(4, 0.5)$ on the nodes $V = \{a, b, c, d\}$ using a coin (as was done in class). Write down your graph (any of the descriptions introduced in Problem Set 1 are okay).
- Bonus question (not mandatory):** Plot the degree distribution of the graph you just made. Plot on top of it the Poisson distribution (see lecture notes for more information) with μ , the mean, equal to the expected degree you computed in (b). Comment on the quality of the approximation.

²Another measure often used is the *average clustering coefficient*. To compute that, we look at each node i and compute i 's clustering coefficient c_i —defined as the clustering coefficient of the subgraph consisting of i , its neighbors, and all links among them in the original graph. Then take the arithmetic mean of c_i across nodes. See Jackson for details.

5. Compound Growth. Fix an integer $k \geq 1$ and consider the sequence $s_n \equiv \left[\left(1 + \frac{1}{k}\right)^k \right]^n = \left[\left(1 + \frac{1}{k}\right)^k \right]^1, \left[\left(1 + \frac{1}{k}\right)^k \right]^2, \dots$.

- a. Show that $\left(1 + \frac{1}{k}\right)^k \geq 2$.
- b. Show that $n \geq 20$ implies $\left[\left(1 + \frac{1}{k}\right)^k \right]^n \geq 10^6$.
- c. If you had one dollar to invest, and were guaranteed that you will earn p percent every day on your money, how long would it take you to become a millionaire (if you aren't one already)?

Figure 1: Network A

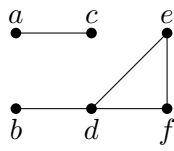


Figure 2: Network B

