1 Networks

1.1 Key Concepts

1. A graph $G = (V, E)$ is a mathematical model that contains a set of vertices $V$ (or nodes) and a set of edges $E$. Graphs are mathematical models of networks, and represent relationships between objects. An edge in a graph may be undirected or directed.

2. Real world networks exhibit heavy-tailed degree distributions, small worlds properties, and high edge clustering.

3. We can check that a distribution is heavy-tailed by noting linearity of the log-log plot of degree versus frequency (i.e. power-law distribution).

4. A network exhibits small worlds properties, if the average shortest-path length between all pairs in the network grow slowly with the number of nodes (e.g. $O(\log n)$). Think Milgram’s 6 degrees of separation.

5. We define the clustering coefficient of graph $G$ to be the average clustering coefficient of its nodes, with the node’s clustering coefficient defined as the fraction of its neighbors that are also neighbors. $G$ has high edge clustering when the clustering coefficient is greater than the graph’s edge density $\frac{m}{\binom{n}{2}}$, if $G$ has $m$ edges. (This means that the average node’s neighbors are more likely than chance to be neighbors.)

1.2 Practice

1. Network Basics
   Assume that each link in the graph represents a mutual friendship between the two lettered nodes.

(a) Who has the most friends of distance one?

(b) Who has the most friends of distance two or less?
(c) How many cycles exist in this graph?
(d) How many components does this graph have? This means it is (blank).
(e) How many triangles exist in this graph? Is this typical of a friendship network?
(f) What is the neighborhood overlap of $B$ and $D$?

2. Small World
Suppose we live in a world with 7 billion people. Everyone has 50 friends, from a random subset of the other 7 billion people.

(a) How many unique “acquaintances” do you have who are two edges or fewer away from you on the graph? (Feel free to make assumptions).
(b) What is the diameter of this graph (i.e. the maximum length of the shortest path separating two agents in the model)?
(c) In the real world, why would we expect you to have fewer acquaintances than you calculated in part (a)?

2 Network Formation Games

2.1 Key Concepts
1. Random graph models view graphs as the realization of a stochastic process. Three important models are

(a) Erdos Renyi: every possible edge edge forms independently with probability $p$; only has small worlds property
(b) preferential attachment: start with a connected graph and grow graph one vertex at a time, new vertices form $q$ edges with existing vertices where each existing vertex’s probability of forming an edge is proportional to their degree so that new vertices preferentially form edges with vertices that already have a high degree; has small worlds property and heavy-tailed degree distribution
(c) copying: for each new vertex, choose a random subset of the existing vertices and the neighborhood of that subset, the new vertex independently forms an edge to each vertex of the random subset with some probability and to each vertex of the neighborhood of the set with another probability; has small worlds property, heavy-tailed degree distribution, and high edge clustering

2. Strategic models view graphs as formed by game-theoretic agents at the nodes. We study three examples:

(a) stay-connected game: an agent’s cost is its distance to every other agent plus the cost of any edges it sponsors; since the distance to a vertex it is not connected to is infinite, agents want to stay connected; either agent can sponsor an edge to another agent, a both agents get the benefit (i.e. edges are undirected)
(b) directed connections game: an agent’s utility is the number of other agents it can reach less the number of edges it sponsors; edges are directed.
(c) bilateral-connections: an agent’s utility is the sum of a constant discounted exponentially by its distance to every other user less the cost of sponsoring any edges; edges are undirected, but both agents must sponsor the edge between them for the edge to exist

2.2 Practice
1. Random graph models

(a) Briefly define the Erdos-Renyi, preferential attachment, and the copying model.
(b) Intuitively, why do the Erdos-Renyi, preferential attachment, and copying graph models lead to small worlds graphs?
(c) Why do preferential attachment and copying additionally exhibit heavy-tailed degree distribution while Erdos-Renyi does not?
(d) Why does copying exhibit edge clustering while Erdos-Renyi and preferential attachment does not?

2. Undirected-reachability game
Recall the directed-reachability game in which the benefit of a connection flows only in one direction. We now consider the undirected-reachability game, which is modeled as an undirected graph and has the following utility function:

\[ u_i(a) = \sum_{j \neq i} \mathbb{I}[d_{ij} < \infty] - c \sum_{j \neq i} a_{ij}. \tag{1} \]

The edge-formation rule in the undirected version of the game is \( g_{ij} = a_{ij} \lor a_{ji} \), so that the graphs formed are undirected and the benefit of an edge flows in both directions.

Consider the two graphs in Figure 2 where arrows denote sponsorship (but the edges are undirected, i.e. both vertices of an edge receive the benefit of the edge). For what values of \( c \) is the left graph a Nash graph? For what values of \( c \) is the right graph a Nash graph?

3. Bilateral-connections game
Recall the bilateral-connections game in which the utility of agent \( i \) is

\[ u_i(a) = \sum_{j \neq i} \delta d_{ij} - c \sum_{j \neq i} a_{ij} \]

for some \( \delta \in (0,1) \), and an edge between two players is formed if both players approve, i.e. \( g_{ij} = (a_{ij} \land a_{ji}) \). For this question, refer to Figure 2 where solid lines denote edges and dotted lines denote edges that an agent is considering sponsoring, but do not exist. Recall also that a graph is pairwise stable if for every edge in the graph, both vertices would weakly prefer to have the edge than not have it, and for every edge not in the graph, if one agent strictly prefers to add the edge, then the other strictly prefers to not add it. Prove the following:
(a) for cost $c < \delta - \delta^2$ and $n = 3$, the complete graph is the unique pairwise stable graph.
(b) for $\delta - \delta^2 < c < \delta$ and $n = 3$, the star graph is pairwise stable.
(c) for $c > \delta$ and $n = 2$, the empty graph (right graph) is pairwise stable.

3 Games on Networks

3.1 Key Concepts

1. The public goods game is an example of a network game with strategic substitutes. One example is the HBO-GO game, in which each agent chooses to buy or not buy a subscription and can benefit from neighbors’ subscriptions. Nash equilibria are given by maximally independent sets (why?).

2. The coordination game is an example of a network game with strategic substitutes, in which neighbors get payoff $(q, q)$ for playing $(0, 0)$ and $(1 - q, 1 - q)$ for playing $(1, 1)$ (and 0 payoff otherwise). Then Nash equilibria occur when each person playing 1 has at least fraction $q$ of its neighbors playing 1 (and each playing 0 has at least fraction $1 - q$ playing 0).

3. We define the hard-threshold cascade as follows. Initialize the graph to all 0’s, force a seed set $S$ to play 1, and at each time step allow anyone who wishes to switch to 1 to do so. The cascade will cause everyone to play 1 if and only if $N \setminus S$ does not contain a subset of cohesion greater than $(1 - q)$. (Remember we defined $coh(S) = \min_{i \in S} \frac{|N_i \cap S|}{|N_i|}$, the minimum over all agents in $S$ of the fraction of neighbors of an agent in $S$ who are also in $S$.)

4. We define the independent-cascades model as follows. Initialize the graph to all 0’s, and fix a seed set $S$ to play 1 at all time periods. In each subsequent period $t$, for each agent $j$ who plays 0 in period $t - 1$, for each neighbor $i$ of $j$ who plays 1 for the first time in period $t - 1$, agent $i$ independently influences agent $j$ with probability $p_{ij}$.

5. The influence maximization problem is to find the seed set $S$ of size $k$ such that maximizes the expected number of agents that will become activated. This problem is NP-hard for both the hard-threshold model and the independent cascades model (the proof for this is reduction to the set cover problem).

6. In the independent-cascades model, the maximum influence problem is then to find the seed set $S$ that maximizes $h_G(S) = E(|I(S)|)$, the expected number of nodes activated starting with seed $S$. The greedy algorithm to do this starts with an empty seed set $S$ and for each of $k$ steps adds the node $i$ that maximizes $h_G(S \cup \{i\}) - h_G(S)$.

3.2 Practice

1. Hard-threshold cascades practice
Recall the model for hard-threshold cascades in which the network starts playing 0 entirely, a seed set $S$ is switched to play 1, and thereafter each node plays 1 when fraction $q$ of its neighbors are doing so.
Consider the following network:

(a) Suppose $q = \frac{1}{2}$, $S = \{A, B\}$, which nodes eventually play 1?
(b) Fix $q = \frac{1}{3}$, specify a seed set $S$ such that $|S| \leq 3$ and the entire graph will play 1.
(c) Fix $S = \{A, B\}$, what is the largest value of $q$ for which $S = \{A, B\}$ will lead to a full cascade?

2. Generalized Coordination Game

Recall the coordination game between each pair of neighbors in a graph as giving payoff $(q, q)$ for playing $(0, 0)$; $(1 - q, 1 - q)$ for playing $(1, 1)$; and $(0, 0)$ otherwise. Now suppose we generalize this game so that neighbors win payoff $(a, a)$ for playing $(0, 0)$; $(b, b)$ for playing $(1, 1)$; and $(x, x)$ otherwise, with $x < a, b$.

(a) Is there a threshold $q$ in terms of the three quantities $a, b, x$ such that each node will play 0 if at least fraction $q$ of its neighbors are adopting 0, and it will adopt 1 otherwise?
(b) If this is possible, calculate $q$. Why is the assumption that $x < a, b$ important?