You have two hours to complete, scan, and re-upload the exam. The exam is open notes, and all course materials are fair game. However, you are required to work alone.

Problem 1. In this question, assume all networks are undirected.

a. What is the diameter of the ring network (also known as the circle or cycle) with n nodes?

b. What is the diameter of a square lattice with L edges (or $L+1$ nodes) on each side? Next, what is the diameter of a d-dimensional hypercubic lattice with L edges on each side? (This lattice is the graph whose vertices correspond to points $(m_1, \ldots, m_d)$ where $m_j \in \{1, \ldots, L+1\}$ for each dimension $j \in \{1, \ldots, d\}$, and two vertices are linked if and only if their coordinates are the same in $(d-1)$ dimensions and differ by 1 in the remaining dimension.) Hence, what is the diameter of a d-dimensional hypercubic lattice with $n = (L+1)^d$ nodes?

c. Consider a tree network where each node except the leaves (the "end nodes") have k neighbors (assuming that $k \geq 3$). How many nodes can be reached from the root node in d steps? Hence, what is the diameter of the tree as a function of k and the number of nodes, n?

d. Say that a network with n nodes exhibits *small worlds* if its diameter, $D(n)$, satisfies $\lim_{n\to\infty} D(n) / \log n < \infty$. In one or two sentences, explain the real-world phenomenon that this definition is attempting to capture, and explain why the definition corresponds to this phenomenon. Which of the networks in parts (a), (b), and (c) exhibit small worlds, and why?

Solution.

a. The diameter is $(n-1)/2$ if n is odd, and $n/2$ if n is even.

b. The diameter of the square lattice is $2L$. The diameter of the d-dimensional lattice is dL. The side-length of a d-dimensional lattice with $n = (L+1)^d$ nodes is $L = n^{1/d} - 1$, so the diameter of this lattice is $d\left(n^{1/d} - 1\right)$.

c. The number of nodes that can be reached from the root in d steps is

$$\min\left\{k(k-1)^{d-1}, n-1\right\}.$$

So the distance from the root to a leaf is the smallest number d such that $k(k-1)^{d-1} \geq n-1$, or

$$\left\lfloor 1 + \frac{\log(n-1) - \log k}{\log(k-1)} \right\rfloor.$$

The diameter is the distance between two leaves that are connected through the root, which equals twice this number. While this is the exact solution, any answer approximately equal to

$$\frac{2\log n}{\log k}$$

is acceptable.

d. The small worlds phenomenon states that, even in large populations, it is often the case that all individuals are connected to each other by a small number of steps. This says that the diameter of the network $D(n)$ is “much” smaller than n, which here we model by saying that, when n is large, $D(n)$ is smaller than $\log n$.

In part (a), $D(n) \approx n/2$, so $\lim_{n\to\infty} D(n)/\log n = \infty$. No small worlds.

In part (b), $D(n) = d\left(n^{1/d} - 1\right)$, so $\lim_{n\to\infty} D(n)/\log n = \infty$. No small worlds.

In part (c), $D(n) \approx 2\log n/\log k$, so $\lim_{n\to\infty} D(n)/\log n < \infty$. Yes small worlds.

Problem 2. Consider the Morris contagion model: There is a finite network (N, E). Each node $i \in N$ takes an action $a_i \in \{0, 1\}$. Each node prefers to take action 1 if and only if at least fraction q of their neighbors also take action 1, where $q \in (0, 1)$ is a fixed parameter.

a. Show that a sufficient condition for never having a contagion from any group of m nodes is to have at least $m + 1$ disjoint sets of nodes that are each more than $(1 - q)$ cohesive.

b. Consider a variant of the Morris contagion model where in period $t = 0$ some nodes play $a = 0$ and others play $a = 1$ (arbitrarily), and subsequently in each period t each node i plays $a = 1$ if and only if at least $q = 0.5$ of its neighbors played $a = 1$ in period $t - 1$. (The difference from the model in lecture is that now nodes can switch from $a = 1$ to $a = 0$ in addition to switching from $a = 0$ to $a = 1$.) Give an example where this process cycles forever.

Solution.

a. Suppose there are $m + 1$ disjoint sets $T_1, \ldots, T_{m+1}$, each of which is at least $(1-q)$ cohesive. Let S be a set of m infected nodes. Since the T_i are all disjoint and there are $m + 1 \geq \#S$ of them, at least one of them does not contain any element of S.

   Thus, we have $T_i \subset V \setminus S$ for some $(1 - q)$-cohesive set T_i, implying there is not contagion from S (by the theorem given in lecture).

b. Consider the complete bipartite graph $K_{2,2}$. In particular, suppose there are two disjoint subsets of nodes, A and B, each of size 2, and that each node from A is connected to every node in B and vice versa.

   If A starts out infected at $t = 0$, then each node in B has all its neighbors infected and each node in A has none of its neighbors infected. Thus, at $t = 1$, each node in B is infected and none of the nodes in A are infected.

   By symmetry, at $t = 2$ each node in A will be infected and no nodes from B. Since this is the same as the $t = 0$ state, we can deduce that the process will cycle indefinitely: when t is even, A will be infected and when t is odd, B will be infected.

Problem 3. Consider an arbitrary network in which, in addition to its position in the network, each node i has a real-valued characteristic $x_i \in \mathbb{R}$. For example, x_i may denote the conversational skills of individual i.

a. Suppose we select a node by first randomly sampling an edge and then randomly sampling one of the nodes on the edge. What is the average value of x_i under this sampling procedure?

b. Now consider the configuration model where each node's degree is drawn independently from a distribution $P(d)$, and then each node's characteristic is draw independently from a distribution $Q(x|d)$. Thus, the characteristic x_i may be correlated with d_i. (For example, individuals with better conversational skills may tend to have more friends.) Suppose we select a node by first choosing an arbitrary node and then randomly sampling a neighbor of that node. What is the expected value of x_i under this sampling procedure in the limit as the number of nodes n goes to ∞?

c. Show that the average value of x_i computed in part (b) is greater than the population average value of x_i if and only if $\text{cov}(x_i, d_i) = \mathbb{E}[x_i d_i] - \mathbb{E}[x_i]\,\mathbb{E}[d_i]$ is positive.

Solution.

a. Letting m denote the number of edges, since a node with degree d_i lies at the end of d_i links, the average value of x_i under this procedure is given by

$$\langle x_i \rangle_{\text{edge}} = \frac{1}{2m} \sum_i d_i x_i.$$

b. In the configuration model, a given stub is equally likely to link to every other stub. Hence, for a given realization of the sequences of degrees $(d_1, \ldots, d_n)$ and characteristics $(x_1, \ldots, x_n)$, if we choose an arbitrary node and an arbitrary stub, the stub connects to a node with degree d_i with probability $\frac{1}{2m-1} d_i$. Thus, the average value of x_i equals

$$\frac{1}{2m-1} \sum_i d_i x_i.$$

Therefore, as we take $n \to \infty$, the expected value of x_i converges in probability to

$$\langle x_i \rangle_{\text{edge}} = \frac{n}{2\langle m \rangle} \langle d_i x_i \rangle ,$$

where $\langle - \rangle$ denotes expectation with respect to the degree distribution.

c. Since $2\langle m\rangle = n\langle d_i\rangle$, we have

$$\begin{aligned}\langle x_i\rangle_{\text{edge}} &= \frac{n}{2\langle m\rangle}\langle d_i x_i\rangle \\ &= \frac{1}{\langle d_i\rangle}\left(\operatorname{cov}(x_i, d_i) + \langle x_i\rangle\langle d_i\rangle\right) \\ &= \frac{\operatorname{cov}(x_i, d_i)}{\langle d_i\rangle} + \langle x_i\rangle.\end{aligned}$$

This exceeds $\langle x_i\rangle$ if and only if $\operatorname{cov}(x_i, d_i) > 0$.

14.15 / 6.207 Networks
Spring 2022