MIT 6.3200 Problem Set 1

Joshua Pereira

February 14, 2024

1 Problem 1

1.A Part a

Given the adjacency matrix, $\mathbf{g}$, and the vector $\mathbf{1}$, we can create an n x n vector, $\mathbf{d}$, corresponding to the degrees of each node as:

$$\mathbf{d} = \mathbf{g}\mathbf{1}$$

1.B Part b

To calculate the number of edges, we can take the inner product of the vector $\mathbf{1}'$ and $\mathbf{d}$ and divide the result by 2. This will count the degrees of all nodes in the network that is twice the number of edges in the graph, by the handshake lemma. Therefore, we have that:

$$m = \frac{1}{2}\mathbf{1}'\mathbf{d} = \frac{1}{2}\mathbf{1}'\mathbf{g}\mathbf{1}$$

1.C Part c

For some undirected network, the common neighbors between two nodes i and j can be found by taking the inner-product of their rows in the adjacency matrix. In other words, row i in $\mathbf{g}$ will tell us the neighbors of i. Therefore, if we take the inner-product between both vectors, the result will be a scalar that is the sum of the number of nodes listed as neighbors by both vectors. Therefore, we can take the number of shared neighbors for any two nodes as:

$$\mathbf{N} = \mathbf{g}'\mathbf{g}$$

1.D Part d

If we take the cube of the adjacency matrix, $\mathbf{g}$, this will give us the number of paths of length 3 beginning and ending at a particular pair of nodes as row i, column j. Therefore, because a triangle is inherently defined as a length-three self-loop, we can take the trace of the adjacency matrix divided by $\binom{3}{2}$ to extract the number of triangle cycles:

$$\Delta(\mathbf{g}) = \frac{\text{trace}(\mathbf{g}^3)}{\binom{3}{2}}$$

This is because the trace of $\mathbf{g}$ is the number of all length-3 paths beginning and ending at the same node i. The nodes in a singular triangle can be partitioned as a pair of two nodes $\binom{3}{2}$ times. Therefore, we divide by this factor to prevent over-counting the same triangle more than once.

2 Problem 2

2.A Part a

Betweenness centrality is given as:

$$B_k \equiv \sum_{(i,j):i\neq j,k\neq i,j} \frac{\frac{P_k(i,j)}{P(i,j)}}{(n-1)(n-2)}$$

Where $P_k(i,j)$ is the number of shortest paths through k that connect family i and family j and $P(i,j)$ is the number of shortest paths connecting family i and family j.

Therefore, for some node k in our network, for any arbitrary one of the attached regions $n_r = i$ and $n_q = j$, there are exactly $n_r \cdot n_q$ different shortest paths between family i and family j. However, because this is also a tree, then $P_k(i,j) = P(i,j)$, because only one such shortest path may exist. So, we have that:

$$B_k \equiv \sum_{m=1}^{d} \frac{n_m(n-n_m-1)}{(n-1)(n-2)}$$

The numerator here is counting the number of paths for each n_m that originate in the n_m and terminate in any of the other nodes, excluding k, because of the tree-structure of the graph. We can calculate the compliment as the number of paths for each n_m that originate in the n_m and terminate at k or within the same n_m. In other words:

$$B_k \equiv \sum_{m=1}^{d} \frac{n_m(n-n_m-1)}{(n-1)(n-2)} = 1 - \sum_{m=1}^{d} \frac{n_m(n_m-1)}{(n-1)(n-2)}$$

Solving:

$$B_k \equiv \sum_{m=1}^{d} \frac{n_m(n-n_m-1)+n_m(n_m-1)}{(n-1)(n-2)} = 1$$

$$B_k \equiv \sum_{m=1}^{d} \frac{n_m(n-2)}{(n-1)(n-2)} = 1$$

$$B_k \equiv \sum_{m=1}^{d} \frac{n_m}{n-1} = 1$$

We yield that the sum $\sum_{m=1}^{d} n_m$ is the sum over all nodes, excluding k, which is $n-1$. Therefore we have that:

$$B_k \equiv \sum_{m=1}^{d} \frac{n_m}{n-1} = \frac{n-1}{n-1} = 1$$

2.B Part b

I will assume that the ith index is indexed from 1. The ith node will have $(i-1)$ nodes to its left and $(n-i)$ nodes to its right. Therefore, the betweeness centrality is given as:

$$B_k \equiv 1 - \frac{(i-1)+(n-i)}{(n-1)(n-2)} = 1 - \frac{1}{(n-2)} = \frac{n-3}{n-2}$$

3 Problem 3

3.A Part a

This process results in biasing towards selecting nodes with higher degrees as they have more edges connected to them (i.e. more odds of being randomly chosen). Therefore, the expectation of the degree D is given as:

$$\mathbb{E}[D] = \sum_{d_i=1} d_i \cdot P_D(d_i)$$

$P_D(d_i)$ must be proportional to the degree of node i and the probability that the node should have such a degree:

$$P_D(d_i) = d_i \cdot P(d_i)$$

We can find the normalization factor:

$$< P_D >= \sum_{d'} d' \cdot P(d')$$

Therefore, we have:

$$\mathbb{E}[D] = \sum_{d_i} d_i \cdot \frac{d_i \cdot P(d_i)}{\sum_{d'} d' \cdot P(d')}$$

3.B Part b

The expectation of X is given by:

$$\mathbb{E}[X] = \sum_{d_i} d_i \cdot P(d_i)$$

Now, we have:

$$\mathbb{E}[X] \leq \mathbb{E}[D]$$

$$\sum_{d_i} d_i \cdot P(d_i) \leq \sum_{d_i} d_i \cdot \frac{d_i \cdot P(d_i)}{\sum_{d'} d' \cdot P(d')}$$

let $< d >$ denote the average degree of the network:

$$< d > \leq \sum_{d_i} d_i \cdot \frac{d_i \cdot P(d_i)}{< d >}$$

$$< d >^2 \leq \sum_{d_i} {d_i}^2 \cdot P(d_i)$$

$$E[X] \leq E[X^2]$$

The square of the expected value is less than or equal to the expected value of the square.

3.C Part c

We begin with the inequality:

$$\frac{1}{|N|} \sum_{i \in N} d_i \leq \frac{1}{|N|} \sum_{i \in N} \delta_i$$

$$\sum_{i \in N} d_i \leq \sum_{i \in N} \delta_i$$

$$\sum_{i \in N} d_i \leq \sum_{i \in N} \frac{\sum_{j \in N(i)} d_j}{d_i}$$

In the smallest upper-bound for the lefthand side of the inequality, we have the situation where each node in the neighborhood of i has degree 1: where $\sum_{j \in N(i)} d_j = d_i$.

We have:

$$\sum_{i \in N} d_i \leq \sum_{i \in N} \frac{d_i}{d_i}$$

$$\sum_{i \in N} d_i \leq \sum_{i \in N} 1$$

Simplifying and with the handshake lemma:

$$2|E| \leq |N|$$

$$|E| \leq \frac{|N|}{2}$$

Which is true, such that the number of edges in a graph cannot exceed half the number of nodes in the tree.

3.D Part d

We see from part a that the square of the expected value is less than or equal to the expected value of the square due to the weighting that occurs as the probability of D actually biases towards degrees with higher nodes as more connections implies a higher chance of being chosen. This is likely how sampling of friendships goes, where choosing a random friendship to poll will lead to more responses from those with more friendships. This is why we often perceive our friends as being more connected than we are because we do not consider that polling a random friendship, selecting one side of it, and determining reality from there actually biases us towards friends with unusually many friends. Indeed, part c suggests that when evaluating the connectedness of a network, considering how connected someone is on top of the popularity of their connections is crucial: i.e. truly popular and well-connected people hangout with other truly popular and well-connected people. In other words, while the number of connections someone has is important, also considering the weight of those connections too is similarly important and we should consider the quality of friendships too: quality over quanitity.