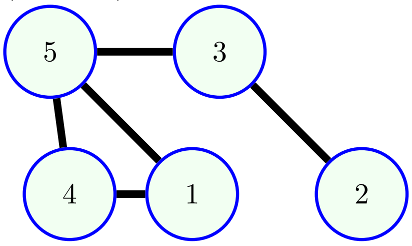
Network Science
Spring 2024
Midterm test solution

There are 3 questions worth a total of 25 points.
Please show your work.

1. (a) (2 marks) Give the adjacency matrix A_{ij} of the following graph .



Solution:

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

(b) (3 marks) What is the global clustering coefficient of the graph?

Solution: There is one triangle in the graph, and six triples, so the global clustering coefficient is $3/6 = 1/2$.

(c) (2 marks)What is the degree distribution of the graph?

Solution: The degree distributions is $p_1 = 1/5, p_2 = 3/5, p_3 = 1/5$.

(d) (3 marks) Suppose the graph given in (a) was generated by the G_{Np} model. What is the probability of generating this realisation in terms of p, where p is the probability of a link being placed between each distinct pair of nodes?

Solution: There are 5 nodes, and therefore $5(5-1)/2 = 10$ Bernoulli trials. 5 trials were successful and therefore 5 were not. So the probability of generating this realisation is $p^5(1-p)^5$.

2. Consider the following modification to the Katz centrality for directed connected graphs with N nodes and $N > 1$:

$$x_i = \alpha \sum_{j=1}^{N} \frac{A_{ij}}{max(k_j^{out}, 1)} x_j + 1. \tag{1}$$

Here, x_i is the centrality of node i, $\mathbf{A}$ is the adjacency matrix for the graph, k_j^{out} is the out-degree of node j, and α is a positive real number.

(a) (2 marks) Provide a concise explanation of the reasoning behind the modification of the Katz centrality when given a directed graph.

Solution: The centrality of a node is now measured by the centrality of its neighbouring nodes, where each neighbour's centrality is scaled with the out-degree of said neighbour. Consequently, a neighbouring node contributes more to a nodes centrality the less out-links it has, and is therefore deemed more important than neighbouring nodes with more out-links.

(b) (3 marks) Now explain the differences between the PageRank centrality given by

$$x_i = \sum_{j=1}^{N} \frac{(1-m)A_{ij}}{max(k_j^{out}, 1)} x_j + \frac{m x_j}{N}, \tag{2}$$

and Equation (1).

Solution: Rather than giving a base-line contribution of 1 to all nodes, the PageRank centrality gives a base-line contribution of the average of all centralities, scaled with m. m puts a weight on the importance of the average of the centralities to the centrality of node i, in relation to the importance of the centralities of the neighbours of node i. When $m > 0.5$ the former is deemed more important to the centrality, and when $m < 0.5$ the latter is deemed more important. When m is close to 1, there is a tendency for the nodes to have the same centrality.

(c) (5 marks) Consider a simple directed N-node graph where node 1 has a directed link to every other N, such that all other nodes have an in-degree $k^{in} = 1$. Compute the centrality vector $\mathbf{x}$ in (2) when $m = 0.5$, giving your answer in terms of N.

Solution: From the symmetry of the graph, we can assume that nodes with $k^{in} = 1$ have the same centrality. Define this as x_s, and define the centrality for node 1 as x_1. Letting $m = 0.5$ gives

$$x_i = \sum_{j=1}^{N} \frac{A_{ij}}{2(max(k_j^{out}, 1))} x_j + \frac{x_j}{2N}.$$

Now, $A_{ij} = 1$ when there is a link from node j to node i, therefore $A_{1j} = 0$ for all j. Consequently,

$$x_1 = \sum_{j=1}^{N} \frac{x_j}{2N} = \frac{(N-1)x_s + x_1}{2N} \tag{3}$$

For nodes $i > 1$, $A_{i1} = 1$, and $A_{ij} = 0$ for $j > 1$. Therefore,

$$x_s = \frac{x_1}{2(N-1)} + \frac{(N-1)x_s + x_1}{2N}. \tag{4}$$

We have a system of two equations (3) and (4), which simplify to

$$x_1 = \frac{N-1}{2N-1} x_s$$

and

$$x_1 = \frac{N^2 - 1}{2N - 1} x_s,$$

respectively. Upon substitution, we find this has the trivial solution $\boldsymbol{x} = 0$.

Is this what we expect intuitively? This a short coming of the Page-rank centrality, that occurs in some cases when we have $k_j^{out} = 0$, called 'dead-ends', resulting in the centrality returned as a trivial solution. In lectures, to prove the columns of G sum to one, and that $\lambda = 1$ is an eigenvalue, we assumed $k_j^{out} > 0$. In this example, there will exist a column where each element is m/N, and therefore the column elements cannot sum to one, and our proof fails.

Note: Due to an ambiguity in the question, students have been awarded two marks for attempting the question, and the remaining three marks for taking the correct steps to find a solution.

3. (5 marks) Define an isolated pair of nodes as two nodes (i, j), with a link from i to j, each with degree 1. What is the expected number of isolated pairs of nodes in graphs generated by the G_{Np} model for a given N and p?

Solution: Let $X_{ij} = 1$ if nodes (i, j) form an isolated pair, and $X_{ij} = 0$ otherwise. The probability of (i, j) forming an isolated pair depends on $1 + 2(N - 2)$ Bernoulli trials, of which 1 is successful. Therefore, $P(X_i = 1) = p(1 - p)^{2(N-2)}$. Let X be the total number of isolated pairs of nodes in a G_{Np} graph such that

$$X = \sum_{distinct(i,j)} X_{ij}.$$

Using linearity of expectation gives

$$\langle X \rangle = \sum_{distinct(i,j)} \langle X_{ij} \rangle.$$

Since $\langle X_{ij} \rangle = P(X_{ij} = 1)$, and there are $\binom{N}{2}$ distinct node pairs, it follows that

$$\langle X \rangle = \binom{N}{2} p(1 - p)^{2(N-2)}.$$