168 - Homework 3. Due on Gradescope by February 2nd

M. Roper

January 26th, 2024

Exercise 1. Make a post on Piazza describing a possible topic for your final project. Describe what kind of project you envisage working on (remember there are three broad categories – using a tool from Math 168 on a real world data set, developing a mathematical model of a real world phenomenon, using networks, or learning a new tool that was not used in class), including what the network would be, and what kinds of scientific question that network analysis might help you to answer.

Exercise 2. Comment on at least one other person's Piazza post from Exercise 1. For your first post, reply to a post that has no other replies. (If you write more than one reply, then you may add subsequent replies to any posts on Piazza). For example, your comment could augment the original post by finding a useful scientific paper, relevant to what they are studying, a mathematical tool that might be relevant, or an aspect of their proposed model that you find interesting (or, perhaps, that needs a bit of clarification or elaboration than was given in the original posting).

Exercise 3. A food chain is a directed graph with n nodes. The directed edges go from 1 to 2, 2 to 3, 4 to 5 and so on.

(a) What are the degree centralities of nodes 1, 2, $\ldots n$? (Remember that for a directed graph, degree centrality is usually identified with k_i^{in}.

(b) What are the eigenvalue centralities of nodes 1, 2, $\ldots n$? Explain your answer.

(c) Now calculate the Katz centralities of each node. (Hint: it is easier to start with the definition of Katz centrality, than the matrix inversion formula that we derived in class. Your answer will depend on the parameter α). .

(d) (* Optional question) In class we determined that the Katz centrality is defined for $\alpha < 1/\lambda$ where λ is the largest real eigenvalue of A. What is the condition for α in the Katz centrality calculation you did in part (c) ? Explain why α can be arbitrarily large for this network.

Exercise 4. Newman, Problem 7.2.

Exercise 5. Calculate the betweenness centralities of the two types of nodes (the central node and the $n-1$ peripheral nodes) in a star graph.

Exercise 6. Newman, Problem 7.6.

Exercise 7. Recall from class that the global clustering coefficient, C of a network is given by:

$$C = \frac{6 \times \text{no. triangles}}{\text{no. 2-paths}} \ ,$$

(a) Assuming that the graph is simple and undirected, explain carefully why:

$$6 \times \text{no. triangles} \ = \sum_i (A^3)_{ii} \ ,$$

(b) Again assuming a simple, undirected graph, derive a formula for the no. 2-paths. (Hint: The answer is not just $\sum_{i,j}(A^2)_{ij}$).

(c) Using your answers from (a) and (b), calculate the global clustering coefficient of the the graph with adjacency matrix:

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

It is fine to use e.g. Matlab to calculate the various powers of A needed for your solution. Hint: To calculate the clustering coefficient, we need A to not have any self-loops. How do we remove the self-loops in our adjacency matrix?

Exercise 8. (Simulations of random networks) Use simulations to calculate how the global clustering coefficient, C, of Erdös-Rényi graphs depends upon the parameters n and p.

Make a graph that contains 3 lines, one for each of the values: $p = 0.05$, $p = 0.1$, $p = 0.2$, and with $n = 5,\ 10,\ 15, \ldots 40$. to show how C depends on n. n will be shown along the x-axis and the global clustering coefficient, C, will be the y-axis of your graph.