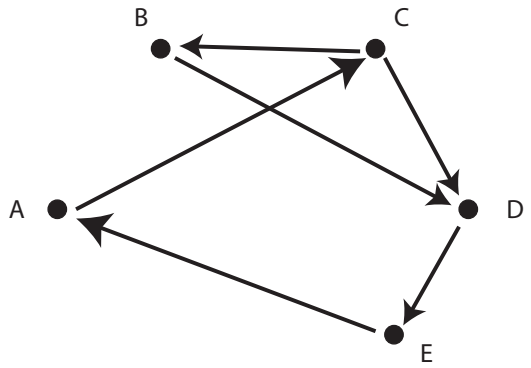
Assignment 3 (Graph Theory and Networks)
Discussed in class on March 7

---

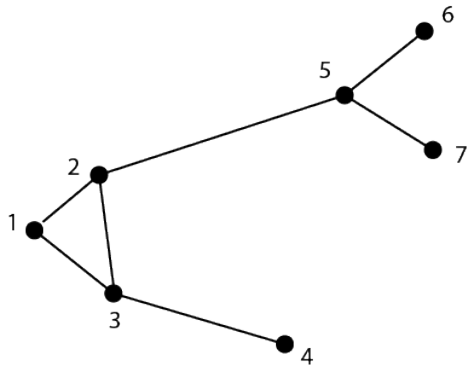
(1) (a) Calculate the closeness centrality for the directed network below. (Since the network is small, no need to normalize.)

(b) Now treat the network as an undirected network (remove the arrow heads). What is the betweenness centrality of the undirected network? (Again, no need to normalize.)



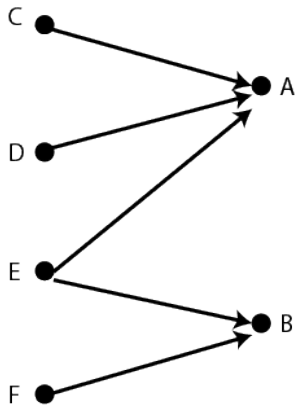
(2) Provide an example of a network such that the node with the highest degree is not the one with the largest closeness.

(3) Consider the undirected graph below. For each question, in case of a tie, answer with all the tied top nodes.



(a) Which node has the highest degree centrality?
(b) Which node has the highest betweenness centrality?
(c) Which node has the highest closeness centrality?

(4) Web designers strive to find ways in which their Web pages score highly on search engine rankings. This question explores some of the reasoning that is involved. The goal is to create Web pages to achieve large authority scores, given an existing hyperlink structure, which in our example is the directed graph below, where each node represents a Web site.



(a) Suppose you want to create a new Web page X and add it to the network so that it could achieve a normalized authority score that is as large as possible. One thing you might try is to create a second page Y as well, so that Y links to X and thus confers some authority to it. You may wonder, though, whether it helps X's authority to have Y link to other nodes as well. Specifically, suppose you add X and Y to the network. Two options are

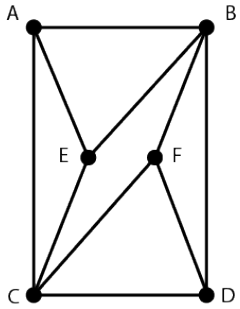
- *Option 1:* Add new nodes X and Y to the network, create a single link from Y to X, and create no links out of X.
- *Option 2:* Add new nodes X and Y to the network; create links from Y to each of A, B, and X; and create no links out of X.

For each of these two options, we'd like to know how X fares in terms of its authority score. So, for each option, show the (non-normalized) authority values that A, B, and X get when you run two iterations of the hub-authority calculation. For which of options 1 or 2 does page X get a higher normalized authority score? What is a plausible explanation for this result?

(b) Suppose you create three new Web sites, X, Y, and Z to add to the network and again try to strategically create links out of them so that X gets ranked as well as possible. Describe a strategy for adding three nodes X, Y, and Z to the network with links of your choice so that when you run two iterations of the hub-authority calculation and rank all pages by thier (non-normalized) authority score, node X shows up in second place.

[In this network, there is no way for X to do better than second place when only three nodes are added to the network as described.]

(5) Perform one iteration of the Kernighan-Lin partitioning algorithm on the graph below. Based on this, what is the best partition or partitions of the graph?



(6) Consider the seven data points $a \cdots g$ with coordinates $a = (0, 0)$, $b = (1, 2)$, $c = (3, 0)$, $d = (5, 2)$, $e = (7, 2)$, $f = (7, 0)$, and $g = (5, 0)$. Use two iterations of K-means clustering to group these into two clusters. For cluster A, use point b as the initial centroid, and for cluster B use point c as the initial centroid. Show your calculations.