Assignment 1

CSC 303: Social and Information Networks

Out:)
Due: February 14, 2020, 2:59 PM

Be sure to include your name and student number with your assignment. All assignments are to be submitted on Markus by the due date.

You will receive 20% of the points for any (sub)problem for which you write "I do not know how to answer this question." You will receive 10% if you leave a question blank. If instead you submit irrelevant or erroneous answers you will receive 0 points. You may receive partial credit for the work that is clearly "on the right track."

1. (10 points)

   Let $A(G)$ be the adjacency matrix of a simple graph G and let $B(G) = A(G) + I$ where I is the identity matrix and let $C = B(G)^2 = B(G) \times B(G)$.

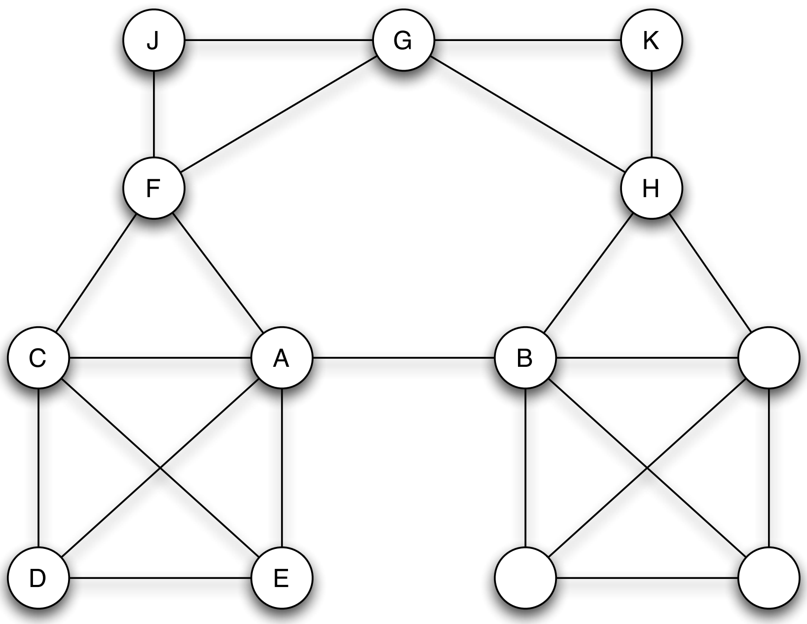
   (a) (5 points) What is the meaning of c_{ij}? (That is, the (i, j) entry in the matrix C.)

   (b) (5 points) Consider $B(G)^k$ and let d be the maximum diameter of the connected components of G. That is, if $G_1, \ldots, G_r$ are the connected components of G, then $d = \max_j diameter(G_j)$. What is the smallest value of k such that $B(G)^k$ can determine whether or not the graph G is connected.

2. (20 points)

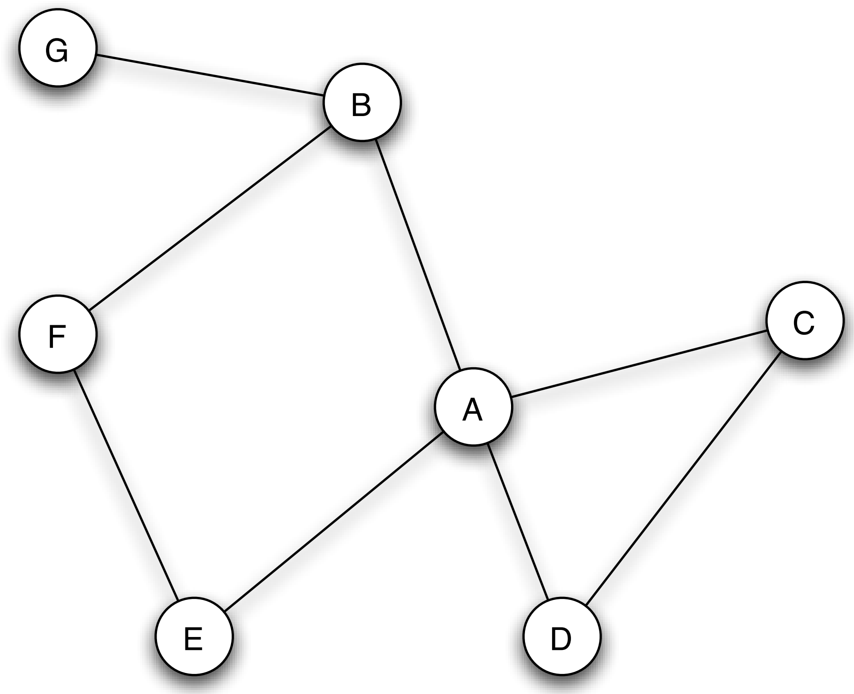
   Consider the following network (figure 3.4 in ET text):

   Suppose that triadic closure occurs at each open triangle independently with probability $\frac{1}{2}$ on a given day.



(a) (5 points) What is the fewest number of days in which nodes D and G will become friends?

(b) (15 points) What is the probability that D and K will become friends on the second day?



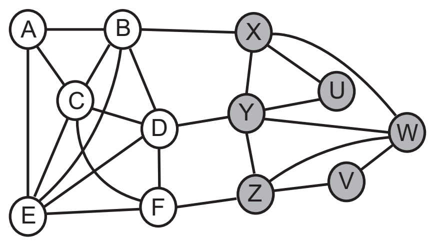
3. (15 points) This question concerns the strong triadic closure property. Consider the graph above.

   - [5 points] Suppose edge (A, B) is a strong edge. Label the remaining edges so as to maximize the number of strong edges (equivalently minimizing the number of weak edges) while satisfying the strong traidic closure property.
   - [5 points] Briefly describe how you went about labelling the graph once the edge (A, B) was labelled as being strong.
   - [5 points] Now suppose edge (A, B) is a weak edge. Label the remaining edges so as to maximize the number of strong edges while satisfying the strong traidic closure property.

4. (20 points) Consider figure 3.15 and the execution of the Givan-Newman algorithm in the text as depicted in figure 3.17. Now suppose we add one additional edge (3,10) to this network.

   - (5 points) Indicate all the bridges and local bridges.
   - (15 points) What is the betweeness of edge (5,7) and edge (3,10)?
     Note: Given the amount of symmetry in this network, you do not have to consider every pair of nodes (u, v) to determine the betweeness of edges (5,7) and (3,10).

5. (25 points) The teenagers in a small Northern Ontario community each engage in exactly one of two sports in the winter: volleyball or hockey. Each node in the following graph corresponds to a person in the community, with shaded nodes denoting those who play hockey and unshaded nodes denoting those who play volleyball. Edges in the graph represent close friendships. (There are 24 edges in the graph in total.)

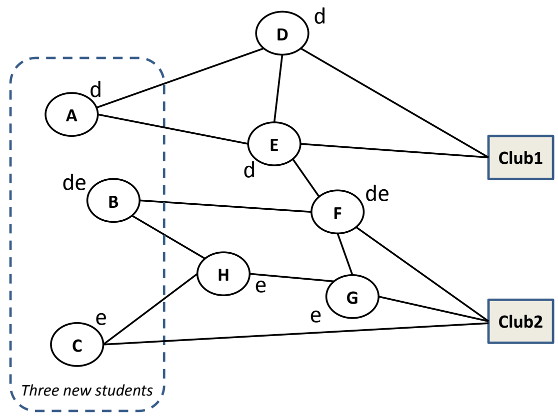


The network contains no nodes representing sporting activities, but we can easily imagine a social-affiliation network by connecting each person to a node representing their preferred activity (e.g., the hockey club or volleyball club). We can then sensibly refer to either *focal closure* or *membership closure* relative to these activities.

- (5 points) Does this graph provide evidence for homophily with respect to a person's preferred sport? That is, is there evidence that friendships and engagement in the same sport are correlated? Give a detailed *quantitative* justification for your conclusion.
- (5 points) Depending on your conclusion (either evidence of homophily or lack of evidence), provide one possible explanation for how the network of relationships and activities may have formed the way it did to support your quantitative conclusion in part (a).
- (5 points) Describe one edge that, if added to the derived social-affilation network, is explained by *focal closure* but not by triadic closure.
- (5 points) Suppose a teenager decides to engage in an additional new sport depending on his/her friends. Suppose that the probability of engaging (in one time step) in an additional new sport occurs independently with probability $p_m = \frac{1}{4}$ for each friend who is engaged in the other sport. What teenager(s) is (are) the most likely to engage is a new sport (in one time step)? Explain briefly.
- (5 points) Suppose that two hockey players, W and U, adopt a new warm-up routine τ. The warm-up routine will spread to a friend who has at least two friends using the new routing. Will the warm-up routing τ spread to all the hockey players? Explain briefly why this will or will not happen. Will the routine τ spread to all volleyball players? Explain briefly.

6. (40 points)

Consider the following social-affiliation network consisting of some college friends and two pubs (called clubs in the figure) in their college town that some students like to frequent on Friday evenings. Three new students, Alice (A), Bob (B) and Claire (C), have just transfered to the college, but have a friend or two in the group already, as indicated by the existing edges. We're interested in how new links might form.



Beyond the social-affilitation network, we are going to assume each person in the network can have one or both of two career interests. They can be be interested in engineeering (e) or in dentistry (d). Here is the list of career interests (the d and e annotations on the network indicate these as well):

	A	B	C	D	E	F	G	H
engineering	N	Y	Y	N	N	Y	Y	Y
dentistry	Y	Y	N	Y	Y	Y	N	N

For now we will consider these career interests to be immutable.

Consider the following models of triadic, membership and focal closure, where we look at the following probabilities for a new link to form during a one-week period:

- Triadic closure will occur between two people X and Y in any given week in a way that depends on both the number of friends they have in common, and the number of career interests (d, e, or both) they have in common. Let X and Y be two unconnected people with f common friends during a specific week and i common interests. The probability of a friendship forming due to triadic closure during the next week is:

$$Pr(X \text{ and } Y \text{ become friends }) = (1 - 0.5^f)\frac{i+1}{3}.$$

  For instance, with two common friends ($f = 2$) and one common interest ($i = 1$), the weekly probability of a friendship forming due to triadic closure is $(1 - 0.5^2)\frac{1+1}{3} = (1 - 0.25)\frac{2}{3} = 0.5$, that is, there is a 50% chance of a friendship forming.

- Focal closure will occur between two people X and Y in any given week in a way that depends on the number of career interests (d, e, or both) they have in common. Let X and Y be two unconnected people with a common focal point (pub frequency) during a specific week, and i common interests. The probability of a friendship forming due to focal closure during the next week is:

$$Pr(X \text{ and } Y \text{ become friends }) = (1 - 0.8^i).$$

For instance, with two common careeer interests ($i = 2$), the weekly probability of a friendship forming due to focal closure is $(1 - 0.8^2) = (1 - 0.64) = 0.36$, that is, there is a 36% chance of a friendship forming.

- Membership closure will occur between a person X and a pub P in any given week in a way that depends on the number of X's friends that frequent P. Let X and C be unconnected (that is, X doesn't freqeunt P), and suppose X has f friends that do frequent P. The probability of membership closure during the next week is:

$$Pr(X \text{ frequent } P) = (1 - 0.7^f).$$

However, every person can frequent *at most one pub*. So if X already frequent a different pub P', this new closure with P will only occur if X has strictly more friends that hang out at P than at P'. In this case, X will move to P and no longer hang out at P' (i.e., that link will dissolve).

- Notice that two unconnected people with both common friends and a common focal point may become friends due to triadic closure or due to focal closure. In any such situation, we treat each closure as acting independently, so the probability of closure in such a case is:

$$p_t + (1 - p_t)p_f = 1 - (1 - p_t)(1 - p_f),$$

where p_t is the probability of triadic closure for the pair (as defined above) and p_f is the probability of focal closure (as defined above).

The following questions refer to this graph.

- (20 points) Suppose the network shown illustrates the situation on the week that the new students arrive. Identify each new link that can form due to triadic, focal, or membership closure involving the new students (nodes A, B and C). State the probability that each of these links will form, using the model above, after their first week in town. Very briefly explain your calculation by explaining which forms of closure (or closures) you are using to derive your probability (and the number of assumed common friends, career interests or any other relevant information).
- (10 points) Which new student is most likely to start (for the first time) hanging out at one of the two pubs imediately after the first week? That is, we ignore the fact that C already hangs out at Club 2. From this small sample of students, can you characterize each of the two pubs in any way?
- (5 points) Which edge in the graph is most *embedded*? (See the definition in Section 3.5 of the text.)
- (5 points) Is there some evidence of homophily among the two career interests (engineering and dentistry) in the initial network? Give a quantitive justification for your answer. For definiteness, if two indiviuals share any interest, consider this a homlogeneous link.

7. (20 points)

   The following question requires you to use the NetLogo software package. You may either install it on your own computer or run it on a CDF machine with the command `netlogo`. Please ask TAs this week if you are having trouble with Netlogo.

   Start Netlogo and load the Segregation model from the SampleModels/SocialScience Library. This implements a version of the Schelling model discussed in class. We would like you to run *five* simulations of the Segregation model setting the parameters as follows: consider two different numbers of agents, 900 and 2500; and consider four settings of the threshold variable (or "% similar-wanted" as it is called in the software), 20%, 30%, 50%, and 70%. Notice that you have eight combinations of settings, and must run five simulations for each. (You can set the speed faster to ensure each simulation proceeds quickly, or slower if you want to watch the patterns emerge).

   For each simulation, record the final "% Similar" once the simulation converges (when all agents are happy) and the number of rounds of movement, or "Ticks" required. For each of the eight combinations of settings, report: (i) the average (over the five simulations) of "% Similar" value and the "Ticks" value at convergence in the table provided; (ii) the minimum value observed over the five simulations; and (iii) the maximum value. *Please hand in the table on the final page of the assignment with these values to make marking easier.*

   On the basis of your observations, draw some qualitative conclusions about the impact of the number of agents and the similarilty threshold on the final degree of population homogeniety and the time taken for the Schelling model to converge. Provide possible explanations for these observed patterns.

   NOTE: It is likely that for the setting $N = 2500$ and $t = 70\%$, the simulation will not terminate. For any setting that does not terminate, indicate for how long it ran, and what conclusions, if any, can be observed from the plots provided by netlogo.

	$N = 900$		$N = 2500$	
	%-Sim	Ticks	%-Sim	Ticks
$t = 20\%$	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.
$t = 30\%$	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.
$t = 50\%$	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.
$t = 70\%$	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.