Homework Assignment 1

CSC 303: Social and Information Networks

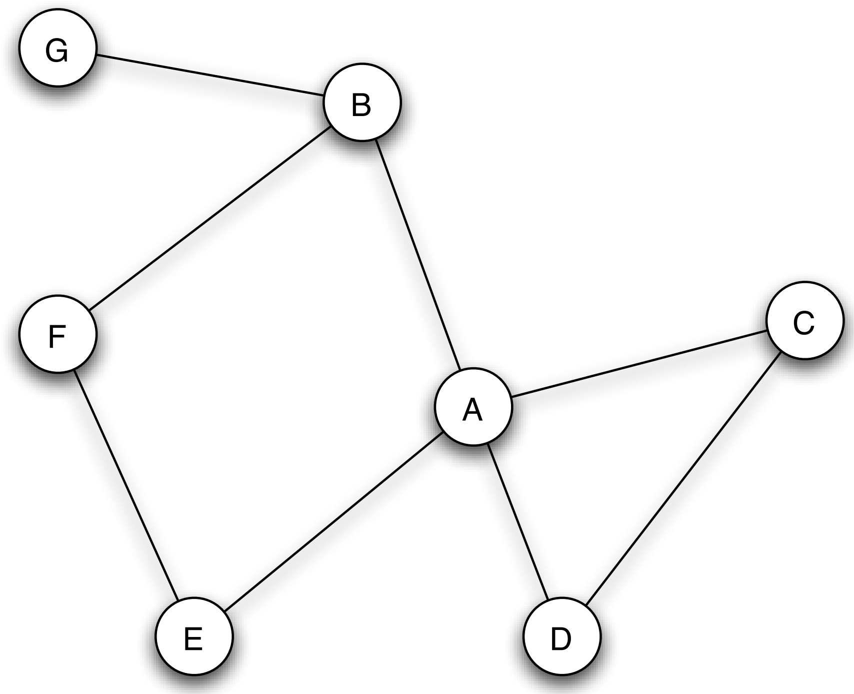
Out: January 17 (first two questions) and January 28 (remaining quuestions)
Due: February 15, 2019

Be sure to include your name and student number with your assignment. If your handwriting is *possibly* illegible, be sure to hand in your assignment in some typed form. All assignments are to be submitted on Markys 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. The following is just a simple exercise regading basic definitions of graophs and directed graphs.

   - [5 points] Consider the directed network in Fig 13.8 of the EK text. What is the largest (in terms of the number of nodes) strongly connected component LSCC in this directed graph? (Thst is, indicate the nodes in the LSCC.)
   - [5 points] Now consider the LSCC as an undirected graph. That is, make all of the edges undirected in the LSCC. Call this graph UCC. Which edges have have the largest *span*.
   - [5 points] Does the undirected graph UCC have a bridge? Explain without checking all edges.



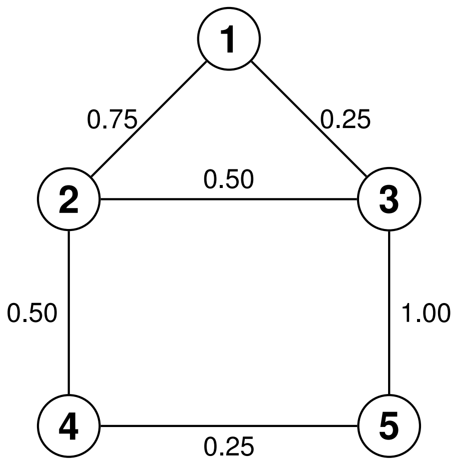
2. This question concerns the strong triadic closure property. Consider the graph above.

   - [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.

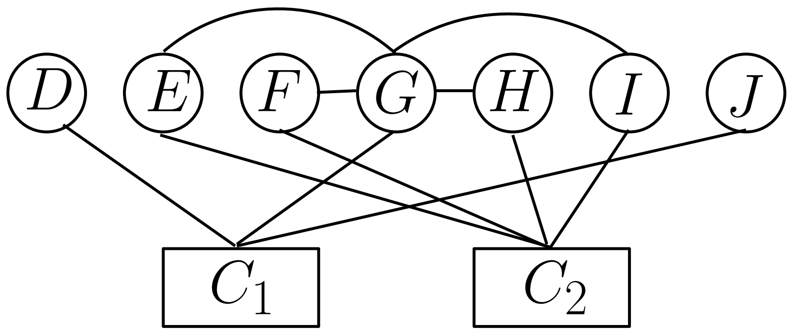
3. [15 points]

Consider the following probabilistic model of triadic closure. A connection $e_{a,b}$ has a weight $w_{a,b} \in [0, 1]$. Given an open triangle ($e_{a,b}$,$e_{a,c}$) the probability that a new connection $e_{b,c}$ forms is Pr[$e_{b,c}$ forms] = $\frac{1}{2}^{2-(w_{a,b}+w_{a,c})}$.

For each missing connection in the following network state the probability that it becomes an edge due to the triadic closure process described above. Assume that if a missing edge can complete two triangles each of the potential closures works independently.



4. Consider the following social-affiliation network of a small neighbourhood community, consisting of some people D–J and two clubs C_1 and C_2. A new person K has recently moved to the neighbourhood. We assume K can make an effort to initially get connected to only one of D–J, C_1, or C_2. Consider triadic, membership and focal closures happen over a one-week period with the probabilities of p_t, p_m, and p_f respectively. We further assume that these closures only happen if K is involved in them as the rest of the network is stable (i.e., no further closures happen amongst D–J,C_1,C_2).

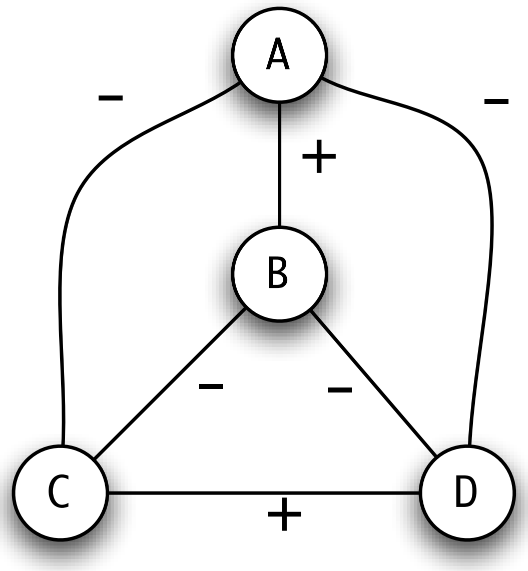


(a) [10 points] Ignoring the probabilitites, K would just like to have the opportunity of connecting to every person D–J as soon as possible (i.e. in the fewest number of weeks) by either making friends with one person or joining one club.

(i) To which person should K initially become friends (in week 0)? Specify an appropriate sequence of closures that makes such friendships possible and briefly explain (using properites of the network) why there is no better person with whom K should become a friend.

(ii) To which club should K initially join? Specify an appropriate sequence of closures that makes such friendships possible and briefly explain (using properites of the network) why there is no better club for K to join.

(b) [5 points] Is it possible that no matter to which person K initially becomes friends or to which club K initially joins, that there is some way to become friends with everyone? Briefly explain.



balanced

5. [10 [points] Consider the strongly balanced network above (from figure 5.2 of the text). Suppose now that every edge has only been correctly labelled with some high probability (say $p = .9$). What is the probability that the network is strongly balanced? You can answer this question in terms of an arbitrary p or for the specific $p = .9$ as suggested.

6. [10 points] You are given an incomplete graph whose edges have been labelled positive or negative. Describe an computationally efficient algorithm that can determine whether or not the missing edges can be labelled so that the resulting network is weakly balanced. What is the time complexity of your algoeithm. Briefly justify correctness and running time.

7. [15 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`.

   Start Netlogo and load the Segregation model from the Models 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% and 55%. Notice that you have six 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 six 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.

	$N = 900$		$N = 2500$	
	%-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.
$t = 55\%$	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.