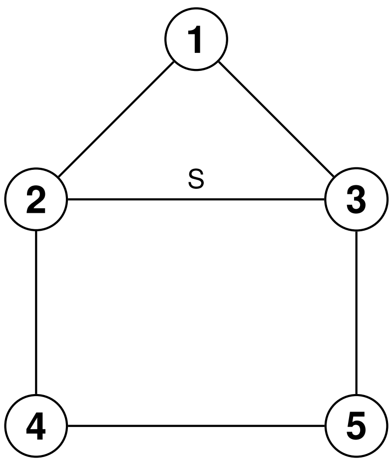
Homework Assignment 1

CSC 200Y: Social and Economic Networks

Out: October 1, 2015
Due: October 16, 2015

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.

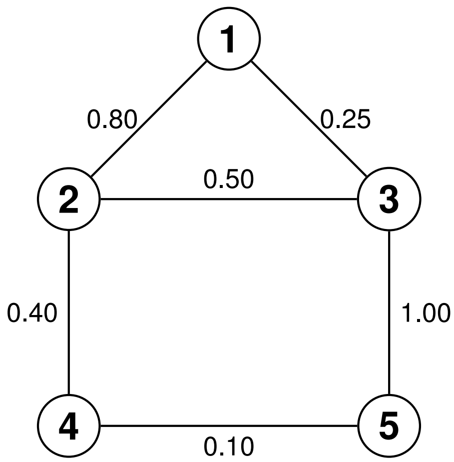
1. Consider the following network:



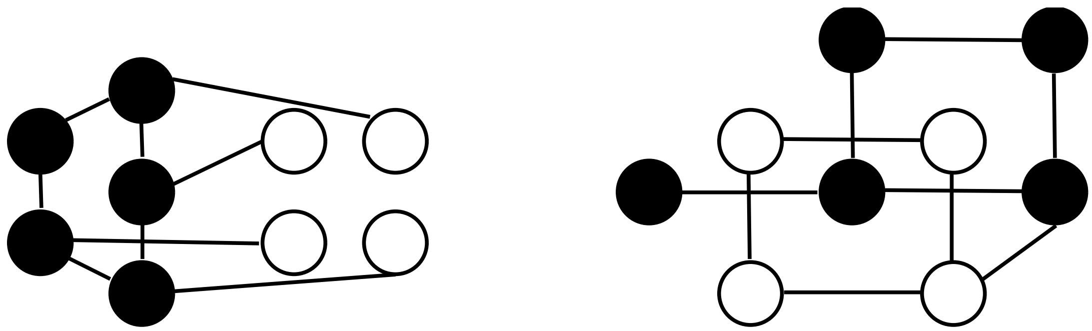
Assume that the nodes obey the strong triadic closure property and the only edge weight known is a strong connection between nodes 2 and 3. List all the valid configurations of the other edge weights. For example, one of the valid configurations is when all missing edge weights are weak (W). Note that since there are 5 free edges and 2 possible settings there could be 32 possible configurations. Using the fact that the edge between 2 and 3 is strong and the graph is not a clique will result in fewer valid configurations.
Definition: $G = (V, E)$ is a clique if $E = \{(u, v)|u, v \in V \text{ and } u \neq v\}$.

2. Consider the following model of triadic closures. 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] = $\frac{1}{2}^{3-(w_{a,b}+w_{a,c})}$.

   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.



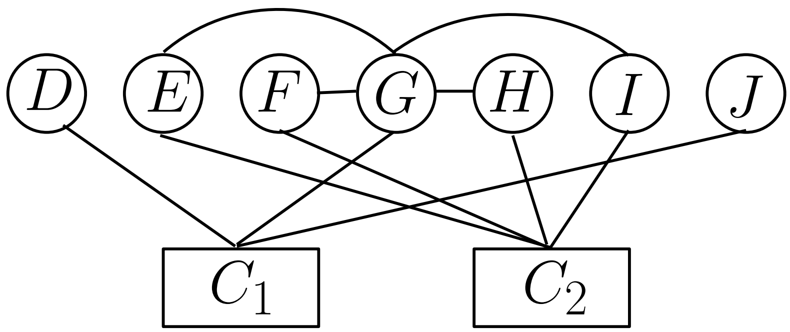
3. Which of the following networks exhibit the higher extent of homophily? Justify your answer with the relevant numerical computations.



Network A   Network B

‘

4. Consider the following social-affiliation network of a small neighbourhood community in Toronto, 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 (i.e. in week 0) 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 will ever happen when restricted to just D–J,C_1,C_2).



(a) 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) 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.

(c) Specify a mathematical formula based on p_t, p_m, and p_f which captures the probability of K being connected to every person after two weeks (i.e. at the end of week 2) if it is initially connected to C_1.

(d) Specify a mathematical formula based on p_t, p_m, and p_f which captures the probability of K being connected to every other person after two weeks if it is initially connected to G.

(e) Suppose K has only two possible choices of G and C_1 for making an initial connection. Also, K has the goal of connecting to all other people in two weeks. Under which condition does connecting to C_1 have a higher chance of success than iconnecting to G in order for K to meet his/her goal? [Hint: You should find a relationship between p_t, p_m, and p_f].

5. 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. A copy of this sheet can be found on the CSC200 web page (Assignments tab) if you misplace it.*

   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.