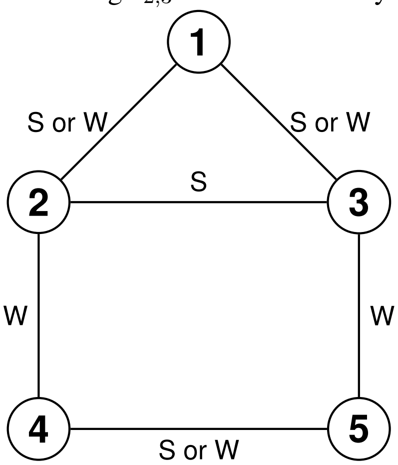
CSC 200—Social and Economic Networks

Sample Solution: Assignment 1 (Oct. 24, 2014)

1. Since $e_{2,3}$ and $e_{2,5}$ and $e_{3,4}$ do not exist we can conclude via the strong triadic closure property that $e_{2,4}$ and $e_{3,5}$ must be weak connections. Next since the only potential triangles $e_{4,5}$ can be a part of are $(2, 4, 5)$ and $(3, 4, 5)$, we see that $e_{4,5}$ can be a strong or weak connection. Applying similar logic to edges $e_{1,2}$ and $e_{1,3}$ and noting $e_{2,3}$ exists we see they can be of either strength also. So our network now looks like:



2. The missing edges with their probabilities of forming are:

- $e_{1,4}$, $\frac{1}{2}^{3-(w_{1,2}+w_{2,4})} = 0.28$
- $e_{1,5}$, $\frac{1}{2}^{3-(w_{1,3}+w_{3,5})} = 0.29$
- $e_{2,5}$, $\frac{1}{2}^{3-(w_{2,3}+w_{3,5})}$ + $\frac{1}{2}^{3-(w_{2,4}+w_{4,5})}$ - $\frac{1}{2}^{3-(w_{2,3}+w_{3,5})}\frac{1}{2}^{3-(w_{2,4}+w_{4,5})} = 0.46$
- $e_{3,4}$, $\frac{1}{2}^{3-(w_{2,3}+w_{2,4})}$ + $\frac{1}{2}^{3-(w_{3,5}+w_{4,5})}$ - $\frac{1}{2}^{3-(w_{2,3}+w_{2,4})}\frac{1}{2}^{3-(w_{3,5}+w_{4,5})} = 0.43$

3. We consider homophily test of the book, Chapter 4. Let p, q be the probability of a node being black or white respectively. For both networks, $p = \frac{5}{9}$ and $q = \frac{4}{9}$. The fraction of cross-color edges for Network A is $\frac{4}{9}$ and is $\frac{1}{10}$ for Network B. As both networks have the same frequency of each type of node and $\frac{1}{10}$ is smaller than $\frac{4}{9}$, Network B exhibits a higher degree of homophily.

4a-i. K should be connected to G. G has distance of 1 or 2 to to other people in the graph. The other nodes have at least one distance greater than 2 to another node. The closures for the initial connection of G are as follows:

- Week 1 or 2:
  - 4 triadic closures: K to E, F, H, I through G.
- Week 1:
  - 1 membership closure: K to C_1 through G.
- Week 2:

- 2 focal closures: K to D, J through C_1

4a-ii. K should be connected to C_1. The minimum number of weeks is 2. C_1 is the only club which has distances of 1 or 2 to other people in the graph. The closures for the initial connection of C_1 are as follows:

- Week 1 or 2:
  - 2 focal closures: K to D, J through C_1.
- Week 1:
  - 1 focal closure: K to G through C_1.
- Week 2:
  - 4 triadic closures: K to E, F, H, I through G.

4b. Yes. This is because the graph is connected and all three closures are possible. So, K can starts with one initial connection to any club or person. Then, it expands connections to all others over time by possible triadic, focal, membership closures.

4c.

- Week 1 or 2: $P(K\ to\ D) = 1 - (1 - p_f)^2$, $P(K\ to\ J) = 1 - (1 - p_f)^2$
- Week 1: $P(K\ to\ G) = p_f$
- Week 2: $P(K\ to\ E, F, H, I) = p_t^4$

All together: $P(K\ to\ D, E, F, G, H, I, J) = (1 - (1 - p_f)^2)^2 p_f p_t^4$

4d.

- Week 1 or 2: $P(K\ to\ E) = 1 - (1 - p_t)^2$, $P(K\ to\ F) = 1 - (1 - p_t)^2$, $P(K\ to\ H) = 1 - (1 - p_t)^2$, $P(K\ to\ I) = 1 - (1 - p_t)^2$
- Week 1: $P(K\ to\ C_1) = p_m$
- Week 2: $P(K\ to\ D, J) = p_f^2$

All together: $P(K\ to\ D, E, F, G, H, I, J) = (1 - (1 - p_t)^2)^4 p_m p_f^2$

4e.

$$
\begin{aligned}
&(1 - (1 - p_f)^2)^2 p_f p_t^4 > (1 - (1 - p_t)^2)^4 p_m p_f^2 \\
\implies &(p_f(2 - p_f))^2 p_f p_t^4 > (p_t(2 - p_t))^4 p_m p_f^2 \\
\implies &(2 - p_f)^2 p_f > (2 - p_t)^4 p_m
\end{aligned}
$$

5. *(20 pts)* Here is one sample answer. Of course, your actual results will vary, but should be in the ballpark.

	N=900		N=2500	
	%-Sim	Ticks	%-Sim	Ticks
t = 20%	Avg. 65.44 Min. 61.80 Max. 67.60	Avg. 6.2 Min. 4 Max 10	Avg. 54.90 Min. 53.70 Max. 56.70	Avg. 8.6 Min. 4 Max. 17
t = 30%	Avg. 78.14 Min. 76.90 Max. 79.20	Avg. 11.4 Min. 9 Max. 19	Avg. 75.42 Min. 74.20 Max. 77.10	Avg. 18.8 Min. 13 Max. 23
t = 55%	Avg. 97.66 Min. 97.00 Max. 97.80	Avg. 20.4 Min. 14 Max. 25	Avg. 96.10 Min. 95.50 Max. 96.70	Avg. 108.4 Min. 87 Max. 117

There are a variety of observations one could make and plausible explanations for them. Here are a few (you didn't need to get all of these, and you may have noticed others):

- The amount of homogeneity that results is much greater than that "desired" (i.e., the target similarity threshold) by the individuals (as discussed in class).
- With fewer individuals (900), things are slightly more homogeneous than with more (2500). The explanation is that with fewer individuals, people are located adjacent to more empty cells, which tends to increase the fraction of "like-colored" neighbours.
- It takes longer to converge with a higher similarity threshold. This can be explained by the fact that local random adjustments are much less likely to make someone satisfied with a higher threshold.
- It also takes longer to converge with more people (2500) than with less (900), for the same reason as above (random moves are less likely to make everyone happy).
- With a small number of people, there is more variance (or spread) in the final degree of homogeneity when the similarity threshold is smaller. (This may not be very noticeable with only five runs of each setting however.)