CSC303: A1

Due Feb 18 at 11:50PM, EST

Be sure to include your name and student number with your assignment, *on the last page. The last page should contain only your name and student number (i.e., the last page should have no solutions)*. This is done as part of an effort to combat unconscious bias in grading. All assignments are to be submitted on Markus.

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

Note: Yed graph editor is a free, relatively simply, multiplatform, graph editor. It may be helpful when completing this assignment.

Question 1: (25 Points) Let $G = (V, E)$ be a simple, undirected, unweighted, graph with n nodes, $v_1, \dots v_n$.

For an unordered pair of nodes, $(v_k, v_l) \in E$, let us define the $n \times n$ matrix $\tilde{L}(G, (v_k, v_l))$ as:

$$\tilde{L}(G, (v_k, v_l))_{ij} = \begin{cases} 1 & i = j = k \\ 1 & i = j = l \\ -1 & i = k, j = l \\ -1 & i = l, j = k \\ 0 & \text{else} \end{cases}$$

From these smaller matrices, let us define $L(G)$ as:

$$L(G) = \sum_{e \in E} \tilde{L}(G, e)$$

(a) [5 points] What is the meaning of values on the diagonal of $L(G)$? Explain your answer.

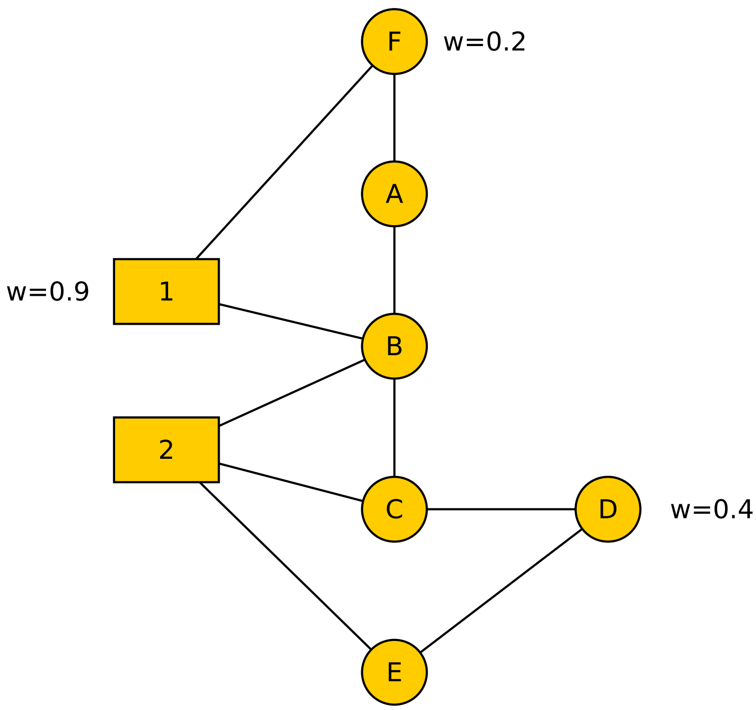
(b) [10 points] Show that if G is a connected graph, then the nullspace of $L(G)$ is of dimension one. Provide a corresponding zero-eigenvector. (HINT: there are many ways to solve this problem, but it may be helpful to consider $\mathbf{x}^T L(G)\mathbf{x}$).

(c) [10 points] Show that for an arbitrary G, then the nullspace of $L(G)$ is of dimension equal to the number of connected components in G. You can assume the results from (b). (HINT: You can assume, without loss of generality, any particular ordering of the nodes of G. It may also be helpful to partition $L(G)$).

Question 2: (25 Points) Considering the following social-affiliation graph. This graph represents the friendships and affiliations of a small group, on day 0. Assume that each day, t, triadic, focal, or membership closures can occur – assume that the graph does not change until day $t+1$ at which point all relationships are updated at once.

Note that a subset of the nodes of the graph are weighted. Assume that for any specific triadic, focal, or membership closure where the the shared node is labelled, that closure occurs with probability equal to the shared node's weight. When there is no such label, triadic closures occur with probability 0.5, membership closures with probability 0.4, and focal closures with probability 0.3.

If a missing edge can be created by multiple closures, then assume that each potential closure works independently.

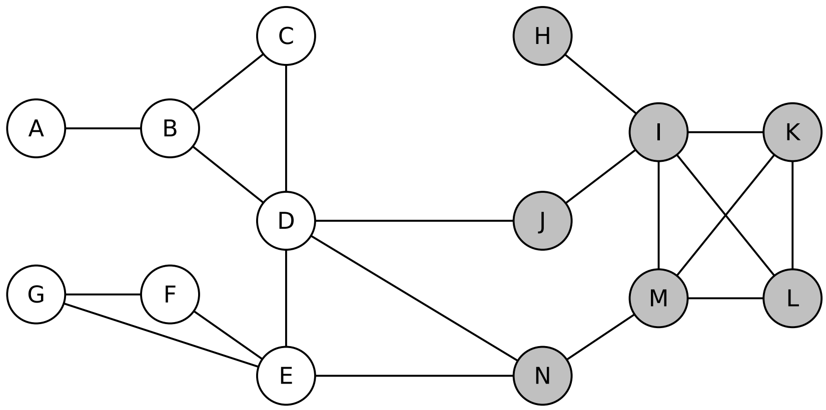


(a) [15 points] For each edge that could be created during the day, list the type(s) of closure(s) that would produce this edge (i.e. triadic, focal, or membership closure) and the probability that it occurs.

(b) [5 points] What is the probability that the nodes E and F are friends on day 2? Justify your answer.

(c) [5 points] In class and tutorial, we discussed how studies suggest that influence is a factor in mutable factors such as obesity. In principle, we could apply social-affilation network such as the above to model obesity (obesity would be an affiliation), but this is problematic as the model does not allow for the dissolution of edges – i.e., obese nodes can never get back into shape. To resolve this, we propose adding a probability, p, with which membership in the obesity affiliation can decay. What new limitation does this change cause? How could we fix this?

Question 3:(30 points) Consider the following social graph in which white nodes correspond to chess players, and grey nodes correspond to checkers players.



(a) [5 points] Is there evidence of homophily among chess and checkers players? Give a detailed *quantitative* justification for your conclusion.

(b) [15 points] To try and determine communities in the network, solve the MinSTC problem on this graph,$G = (V, E)$, using the reduction to Vertex Cover, from the week 2 tutorial. Specifically, you should:

 (i) (10 marks) Draw the corresponding graph ,$G_T = (V_T, E_T)$. Remember that the nodes in V_T correspond to the edges in the original graph, E. Follow the naming convention where nodes in V_T are named after the endpoints of their corresponding edge, in alphabetic order (e.g., $(B, C) \in E$ is named $BC \in V_T$, $(D, B) \in E$ is named $BD \in V_T$).

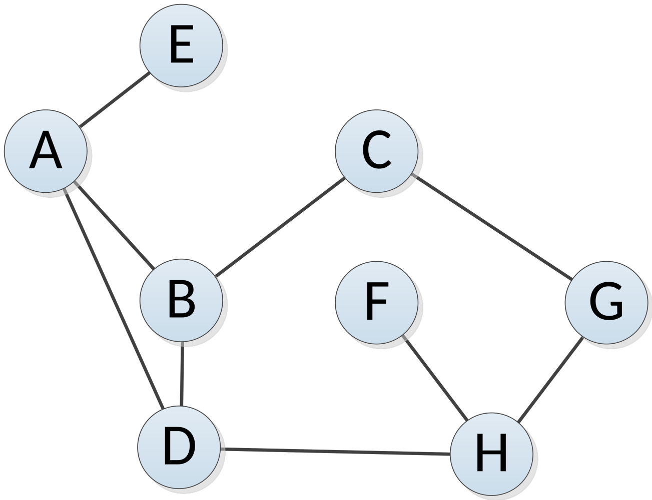
 (ii) (4 marks) Create vertex cover of G_T using maximum degree greedy algorithm (i.e., repeatedly choose the vertex that covers the most uncovered edges). Ties should be broken alphabetically. Make sure to state the order in which you add nodes to the cover, and the number of uncovered edges in the complementary graph covered by adding the node.

 (iii) (1 mark) Report the resulting communities that are connected by strong edges.

(c) [5 points] In general, for an arbitrary G and corresponding G_T, what can you conclude about the neighbourhood overlap of an edge $e \in E$ (see week 2 slides for definition) which corresponds to a degree-0 node in the G_T graph? Justify your answer.

(d) [5 points] In the graph, which node (if any) would you say is likely to have bridging capital? What about bonding capital? Why?

Question 4:(20 points) This question concerns the strong triadic closure property. Consider the graph below.



(a) [5 points] Suppose edge (D, H) 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 triadic closure property.

(b) [5 points] Briefly describe how you went about labeling the graph once the edge (D, H) was labelled as being strong.

(c) [5 points] Now suppose edge (D, H) is a weak edge. Label the remaining edges so as to maximize the number of strong edges while satisfying the strong triadic closure property.

(d) [5 points] Are any of the edges bridges, or local bridges? If so, list these edges and their spans.

Question 5:(10 points) You are given an incomplete graph, $G = (V, E, w)$, whose edges have been labelled positive or negative. Describe a 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 algorithm? Briefly justify correctness and running time.

Question 6: (15 points)

The following question requires you to use the NetLogo software package. I strongly recommend running 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. Note that there is a slight difference, instead of X agents desiring at least n of their neighbours to also be X, in this variant X agents desire at least $n\%$ of their non-empty neighbours to also be X. This has no significant impact on the observed trends.

We would like you to run *three* simulations of the Segregation model setting the parameters as follows: consider two different numbers of agents, 1200 and 2400; and consider four settings of the threshold variable (or "% similar-wanted" as it is called in the software), 25%, 50%, and 75%. Notice that you have six combinations of settings, and must run three 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: For any setting where the model does not converge, indicate for how long it ran (the tick counter is at the top of the display), and what conclusions, if any, can be observed from the plots provided by netlogo. When the desired % similarity is high, you may wish to increase the simulation speed.

	$N = 1200$		$N = 2400$	
	%-Sim	Ticks	%-Sim	Ticks
$t = 25\%$	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.
$t = 75\%$	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.	Avg. Min. Max.

END OF ASSIGNMENT 1

If you are typsetting the assignment using the provided LaTeX, then please write your name and student number below.

NAME: Your name should go here, on the last page.
STUDENT NUMBER: You student number should go here, on the last page.