CSC303: A1

Due Thu Feb 15 at 11:59PM

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."

Some graph editors that may be of help: Yed (a free, simple, multiplatform, graph editor `https://www.yworks.com/products/yed`), past students have also used `https://csacademy.com/app/graph_editor/`

A LaTeXfile of this assignment is available on Quercus.

Before you're finished with this assignment, please check the submission checklist:

1. Your name and student number are written down **on the last page**
2. Any use of generative AI has been documented, and declared (see course syllabus for the details of what is required – in particular, you should transcribe and include all your interactions with the tool)
3. Your solutions are contained within a *legible* PDF of *reasonable filesize* (I believe the MarkUs filesize limit is about 8MB this year).
4. The correct .pdf file has been uploaded to MarkUs.
5. Your submission has the exact filename requested in MarkUs, including the ".pdf" file extension.
6. Your submission can be viewed from *inside* of MarkUs. Note that if MarkUs can't preview your submission, you may be missing the ".pdf" file extension.

Question 1: (20 points) Let $G = (V, E, w)$ be a signed, connected, undirected, graph. Assume that G can be completed to a strongly structural balanced graph. Let $L(G)$ be the signed Laplacian matrix of G.

(a) [5 points] Prove that if G has only positive edges, then the vector of all ones ($\mathbf{1} \in \mathbb{R}^{|V|}$) is a zero eigenvector of $L(G)$.

(b) [5 points] Let G' be a signed undirected graph, with only positive edges, and K connected components. Using the results from a), give a lower bound for dimensionality of the null space of $L(G')$. Explain. (HINT: The linear algebra review topics may be helpful)

(c) [10 points] Recall Harary's balance theorem. It turns out that there is a close relationship between the communities, and a zero eigenvector of $L(G)$. Suggest one such relationship, and prove it holds. (NOTE: G is a signed, connected, undirected graph that may contain negative edges).

Question 2:(20 points)

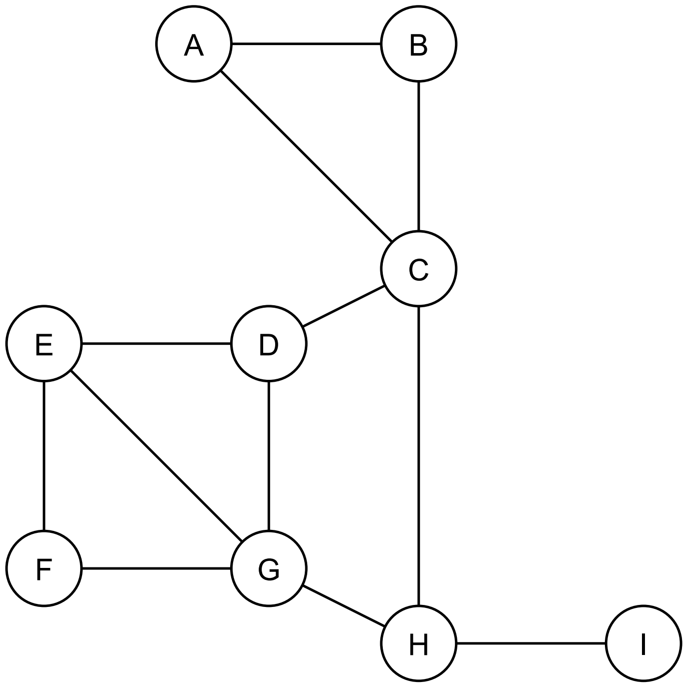
(a) [5 points] Suppose you are studying a transportation network with symmetrical routes, represented by the undirected graph $G = (V, E)$. You find an edge $e = (u, v) \in E$ with high dispersion. What are some possible conclusions that we can reach? Explain.

(b) [5 points] Suppose you are studying a connected collaboration network of actors. Let B be the highest Bacon number in this network. Consider the shortest paths between any two actors; what is the length of the longest such shortest path? Explain, answering as best you can.

(c) [5 points] How many edges are there in a **rooted** tree with N vertices? Provide a proof.

(d) [5 points] Consider an edge (u, v). Given that the nodes u and v have N_u and N_v neighbours respectively, what are upper & lower bounds on the embeddedness and dispersion of the edge? Briefly justify.

Question 3:(20 points) Consider the undirected graph $G = (V, E)$, drawn below.



(a) [5 points] Draw the corresponding graph G_T, such that solving MinSTC on G is equivalent to solving the min vertex cover problem on G_T.

(b) [5 points] Propose a labelling of each edges as strong or weak, that maximizes strong edges while maintaining the STC property.

(c) [2.5 points] What is the clustering coefficient of the node C?

(d) [2.5 points] Is there any way to increase the clustering coefficient by adding or removing nodes, while maintaining the set of edges unchanged? How about decreasing the clustering coefficient? Briefly explain. [Clarification: If we add a node, we are not allowed to add any edges. If we delete a node, then then *only* the edges that are no longer well defined should be removed.]

(e) [5 points] List all local bridges, and their spans.

Question 4:(10 points) In class we saw an algorithm that could determine whether a connected graph G was completable to a strongly balanced graph (you can assume that G has at least one node). You classmate claims that they can determine whether or not G is completable to a strongly balanced graph using the following modification of BFS (see Algorithm 1). This modification of BFS either returns "NO SOLUTION", or returns "SOLUTION EXISTS" after colouring every node as either black, or white.

In the pseudocode, let $\mathcal{N}_+ : V \to \mathcal{P}(V)$ be the ***positive*** neighbourhood function for the graph G (i.e., $\mathcal{N}_+(v)$ is the set of nodes connected by a single ***positive*** edge to v). Let $\mathcal{N}_- : V \to \mathcal{P}(V)$ be defined equivalently, but for negative edges.

Prove the algorithm works, or explain why it doesn't work. You are allowed to use the correctness of any algorithm, characterizations, or theorems from class, without further proof.

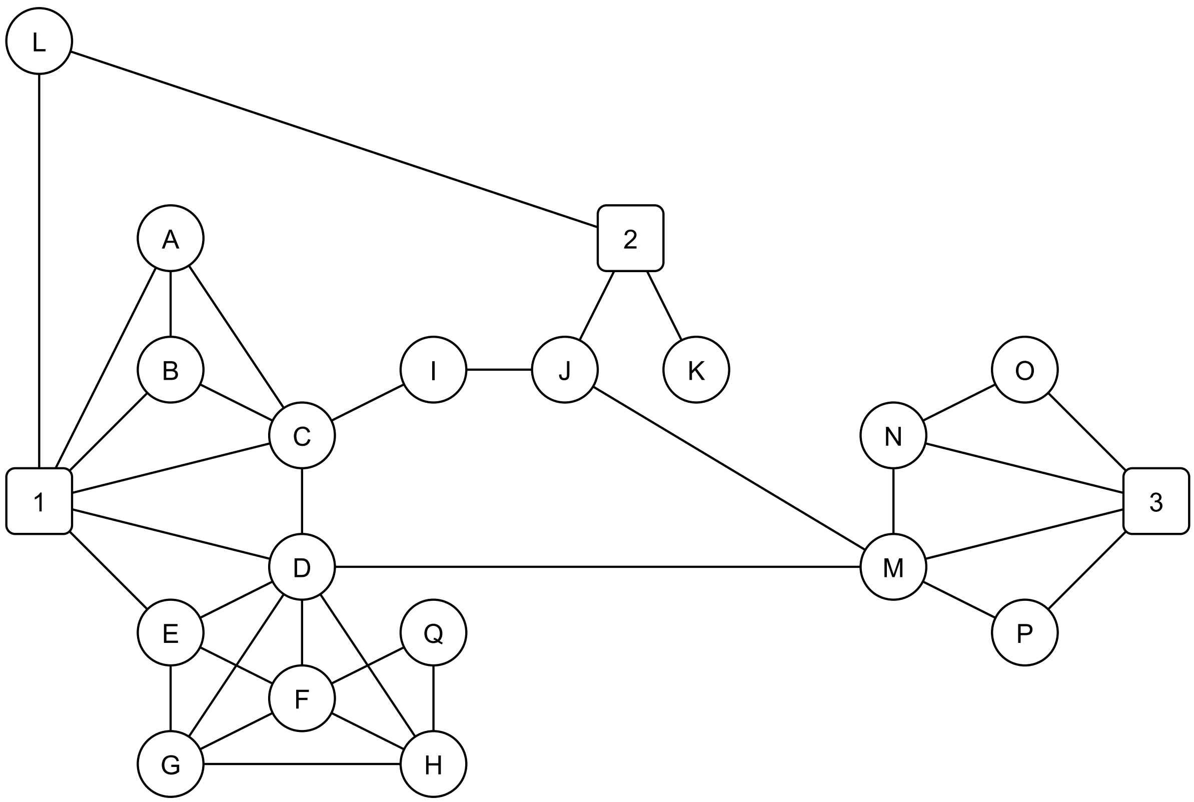
Algorithm 1 Completable

```
for v ∈ V do
    v.colour = None
end for
v_1 = choose_random_node(V)
v_1.colour = black
queue = [v_1]
while queue != ∅ do
    v = queue.pop()
    for p ∈ N_+(v) do
        if p.colour == None then
            p.colour = v.colour
            queue.push(p)
        else
            if p.colour != v.colour then
                return NO SOLUTION
            end if
        end if
    end for
    for n ∈ N_−(v) do
        if n.colour == None then
            n.colour = opposite(v.colour)
            queue.push(n)
        else
            if n.colour == v.colour then
                return NO SOLUTION
            end if
        end if
    end for
end while
return SOLUTION EXISTS
```

Question 5: (30 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.

Assume that triadic closures occur with probability 0.6, membership closures with probability 0.3, and focal closures with probability 0.25.

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



(a) [10 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. You only need to list edges that directly involve D (i.e., D is one of the endpoints of the created edge).

(b) [5 points] What is the probability of O joining focus 1 within 2 days (i.e., allowing for two steps where new closures can form)?

(c) [5 points] Give an example of a node in this network with bonding capital, and an example of a node in this network with bridging capital. Briefly justify.

(d) [5 points] Over a sufficiently long period of time, what will this model converge to? (you do not need to justify this answer). Suggest a way to modify the model so that the behaviour is less unrealistic.

(e) [5 points] Now, considering the subgraph that consists only of the members of the focii 1 and 3. Then at $t = 0$, is there evidence for homophily? Be as rigorous in your answer as you can.

Question 6: (15 points)

The following question requires you to use the NetLogo 6.3.0 software package. I strongly recommend running it on a teaching lab machine with the command `netlogo`. Please ask on Piazza, or during the week 4 tutorial, if you are having trouble with Netlogo.

Start Netlogo and load the Segregation model. Go to File, Models Library, and select SampleModels/SocialScience/Segregation. 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. If they have no none-empty neighbours, they are also satisfied. This has no significant impact on the observed trends.

You can remotely access the teach.cs machines by following these instructions: `https://www.teach.cs.toronto.edu/faq/#GS3`

Note that the simulation *technically* can be run in a web browser but this is very slow! I do not suggest it, as the runs will likely take an unreasonably long time to converge.

In the model, all individuals are blue or orange. Happy individuals are represented by squares, and unhappy individuals are represented by X's.

We would like you to run *three* simulations of the Segregation model setting the parameters as follows: consider two different densities, 70% and 95%; and consider three settings of the threshold variable (or "% similar-wanted" as it is called in the software), 25%, 50%, and 65%. 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 six combinations of settings, report:

(i) the average (over the three simulations) of "% similar" value and the "ticks" value at convergence in the table provided;

(ii) the minimum value observed over the three 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 within 5000 ticks (the tick counter is at the top of the display), indicate for how long it ran, and what conclusions, if any, can be observed from the plots provided by netlogo. When the desired % similarity is high, you will want to increase the simulation speed.

	density = 70%		density = 95%	
	%-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.
$t = 65\%$	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.