

CSC303: A2

Due Mar 23 at 11:50PM, Toronto Time (EDT)

Include your name and student number with your assignment, **on the last page**. All assignments are to be submitted on Markus.

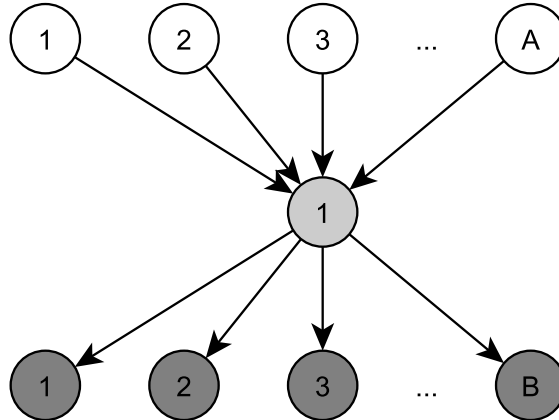
You will receive 20% of the points for any (sub)problem for which you write “I do not know how to answer this question.” If instead you submit irrelevant, erroneous, or blank answers then you will receive 0 points. You may receive partial credit for the work that is clearly “on the right track.”

A \LaTeX starter file for this assignment is available on Quercus, under the Files tab.

Note: Yed graph editor is a free, relatively simply, multiplatform, graph editor <https://www.yworks.com/products/yed>

Question 1: (15 points) Consider the following graph with $n = A + B + 1$ nodes, where $A, B \geq 1$. Note that the nodes are arranged in a simplified bow-tie structure. Provide a closed-form solution for the final authority values of the node(s) specified below, when running scaled PageRank with scale $s < 1$.

Assume that we initialize the graph such that the sum of all initial authority values is 1.
 HINT: You may want to check your solutions once you're done.



- (a) [5 points] In the general case, what are the equilibrium authority values of the nodes in the topmost row (i.e., the white nodes)? The only variables that can be contained in your solutions are A , B , s , and/or n . Your solutions do not need to be simplified. Briefly explain. Finally, evaluate your solution for $s = 0.8$, $A = 5$, $B = 4$, and state the result to 4 decimal places.
- (b) [5 points] In the general case, what is the equilibrium authority value of the node in the centre row (i.e., the light grey node)? The only variables that can be contained in your solution is A , B , s , and/or n . Your solutions do not need to be simplified. Briefly explain. Finally, evaluate your solution for $s = 0.8$, $A = 5$, $B = 4$, and state the result to 4 decimal places.
- (c) [5 points] In the general case, what are the equilibrium authority values of the nodes in the bottom row (i.e., the dark grey nodes)? The only variables that can be contained in your solutions are A , B , s , and/or n . Your solutions do not need to be simplified. Briefly explain. Finally, evaluate your solution for $s = 0.8$, $A = 5$, $B = 4$, and state the result to 4 decimal places.

Question 2: (10 points) You are given 2 social-affiliation networks:

- (a) [5 points] In the table below, we describe the first social-affiliation network. We list the social distances within the graph, the number of unordered pairs of nodes at each social distance, and the number of edges whose endpoints are at each social distance: based on what we know about decentralized search using social distance, does this graph's distribution of edges appear likely to support efficient decentralized search? Briefly explain.

Social Distance, s	$ \{\{v_1, v_2\} v_1, v_2 \in V, \text{social_distance}(v_1, v_2) = s\} $	$ \{(v_1, v_2) \in E \text{social_distance}(v_1, v_2) = s\} $
2	100	50
3	90	30
4	200	50
5	500	100
6	1200	200

- (b) [5 points] In the table below, we describe the second social-affiliation network. We list the social distances within the graph, the number of unordered pairs of nodes at each social distance, and the number of edges whose endpoints are at each social distance: based on what we know about decentralized search using social distance, does this graph's distribution of edges appear likely to support efficient decentralized search? Briefly explain.

Social Distance, s	$ \{\{v_1, v_2\} v_1, v_2 \in V, \text{social_distance}(v_1, v_2) = s\} $	$ \{(v_1, v_2) \in E \text{social_distance}(v_1, v_2) = s\} $
2	100	60
3	90	40
4	200	30
5	500	24
6	1200	20

Question 3: (17.5 points) Consider the threshold model of spread covered in lecture; recall that all nodes are initially using product B , and may switch to product A .

Assume we are running the model on some arbitrary undirected simple graph, $G = (V, E)$, and that our set of initial adopters is $I \subset V$.

Further assume that each node, v , has a rewards $a(v)$ and $b(v)$, which are the reward per agreeing neighbour when node v is using product A, or B respectively.

- (a) [5 points] In class we claimed that if a complete cascade doesn't occur, then there must exist a blocking cluster in $V - I$. Prove this, and describe how to determine what nodes are in such a cluster in the graph G , when at least one such blocking cluster exists.
- (b) [5 points] In class we claimed that if there is a blocking cluster in $V - I$, then a complete cascade cannot occur. Prove it.
- (c) [2.5 points] Provide a example of a product adoption process, where it would **not** be appropriate to apply the threshold model of spread. Briefly explain why not.
- (d) [5 points] Provide the best possible example you can think of, of a spread process in a social network where it would be appropriate to apply the threshold model of spread. Briefly explain why it is appropriate, and also state any limitations in your example.
You are not allowed to use the examples of messaging applications, competing TV shows, or voting.

Question 4: (15 points) This question revolves around a strange and terrible tale that strikes fear into the hearts of computer scientists and engineers everywhere – the dread tale of the hardware virus! Or more accurately, the hardware prion.

Once upon a time, in a faraway land, unsuspecting engineers daily used DVI-to-VGA adapters to connect their laptops (which only had DVI connections) to various VGA projectors. Little did they know of the horrors that would be unleashed upon them! One day, a single pin was bent in an adapter. This bent pin, when forced into a laptop’s DVI port, caused the corresponding hole of the port to break. When a new adapter was inserted into the laptop’s DVI port, the broken port would bend the pin of the adapter in the same way. Thus, the infection would spread from laptop, to adapter, to laptop, and so on.

In this question we will model these events with the SIR model. In the following network, nodes are pieces of hardware (either laptops or DVI adapters), and edges indicate where one piece of hardware has been connected to another (i.e. an adapter has been inserted into a laptop).

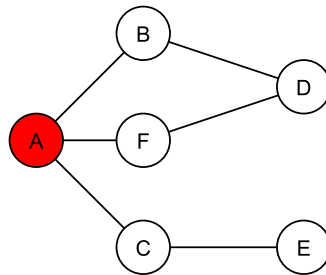
We will assume that each laptop is used with every adapter that it is has an edge to, every day. We will be optimistic, and always assume that laptops and adapters always break at the end of the day (i.e., a newly broken piece of hardware, can’t break anything else until the next day).

On day 0, the unimaginable happens. A careless user strikes the DVI adapter *A* against a table edge in such a way that the pins of the adapter are bent. The adapter is now contagious in the manner described above. We will now model the spread of the hardware prion throughout the company’s hardware.

In this simulation we will be optimistic: assume that any infection occurs at the end of the day (e.g., if laptop *B* is infected on day 1, then the earliest it can potentially infect *D* on day 2).

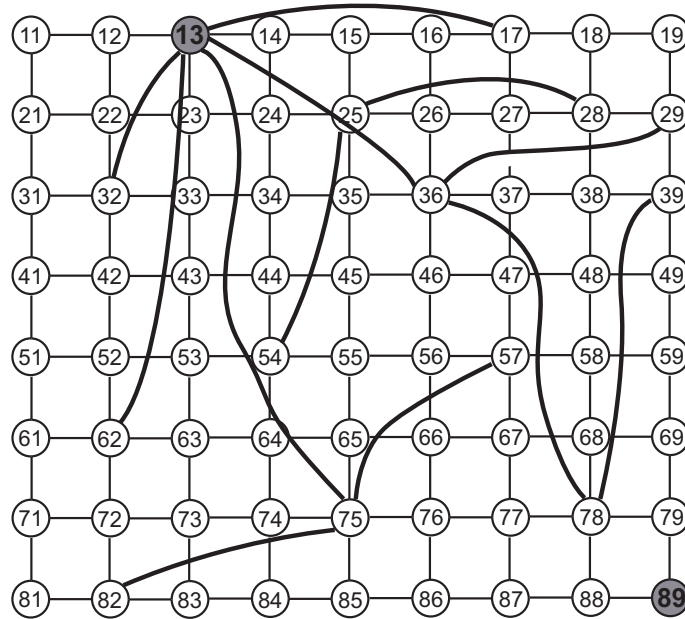
The “duration” of the infection is 2 days, after which the node cannot be reinfected. In other words, it takes 2 days for someone to diagnose the problem and to send the hardware for replacement or repair. Ergo, if a piece of hardware is infected on day 5, then it can potentially infect other pieces of hardware on day 6, or day 7. After this, the node is permanently removed from the network (i.e. it is no longer in the network from day 8 onward).

Additionally, assume that on each day that an uninfected node is exposed to an infected node, there is an independent 60% chance that the infection will spread.



- (a) [7.5 points] For C,D and E, what is the probability that this piece of hardware will be broken after 10 days? Briefly justify.
- (b) [2.5 points] What is an example of an edge could be added to this graph to increase the probability that the node *E* would be damaged by the hardware prion? Briefly explain.
- (c) [5 points] Suppose that we removed the assumption that hardware always breaks at the end of the day. How might we modify our model? Briefly explain and state any new assumptions you make. There are many possible correct answers.

Question 5: (10 points) Consider the following communication network



Recall from class, the process of decentralized search. In decentralized search, if a node n is asked to forward a message so that it will reach a target node t quickly, it must forward the message to one of its friends f (who will then continue the process). Node n will forward the message to the friend f that is closest to target node t , where closeness is measured by grid distance (or city block distance). The grid distance is simply the length of (smallest) path between f and t using only local edges (thin edges in the picture). If there are several friends f that are equally close to the target, n can send its message any one of these friends.

- [2.5 points] 13 is trying to get a message to node 89 using the decentralized search process. What path will the message take? (Note: There may be more than one acceptable answer but you only need to provide one path.). How many hops (links) will the message need to traverse?
- [2.5 points] What is a shortest path that the message from 13 to 89 could take (not using decentralized search)? (Note: There may be several different shortest paths; just list one). How long is it?
- [5 points] Compare and contrast centralized & decentralized search; briefly explain. Be sure to include at least two advantages for both. The advantages can be situational (i.e., the advantage is only relevant for certain types of problem or situation). There are many possible solutions to this question.

END OF ASSIGNMENT 2

If you are typesetting the assignment using the provided L^AT_EX, 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.