CSC303: A2

Due Thu Mar 21 at 11:50PM, Toronto Time

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."

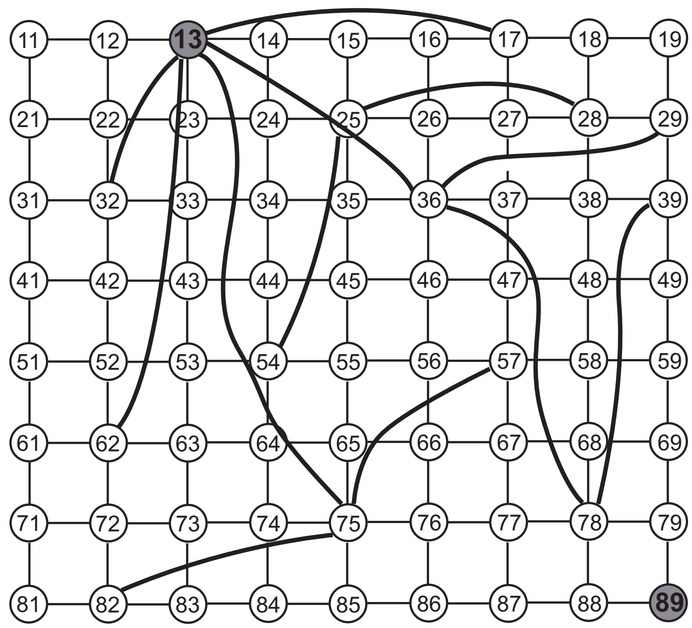
A LaTeX starter file for this assignment is available on Quercus, under the Files tab.

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/`

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: (15 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 to a target node t, then 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 (i.e., city block distance, or Manhattan 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, then n can send its message any one of these friends.

(a) [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). What is the length of the path?

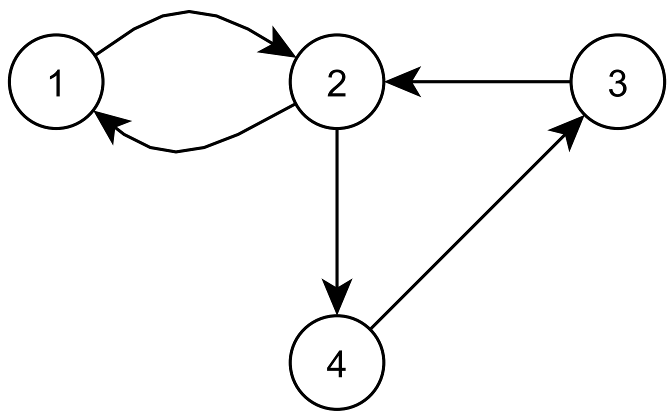
(b) [5 points] What is a shortest path that the message from 13 to 89 could take (not using decentralized search)? (Note: There may be several different shortest paths; just list one). How long is it?

(c) For each of the following scenarios, is the decentralized or shortest path more plausible? Why?

(i) [2.5 points] Nodes represent stops, warehouses, and transportation hubs in a freight system. Local edges are truck routes, and long-distance edges are train routes. All edges take the same amount of time to traverse. A freight company needs to deliver several tons of grain from node 13 to node 89.

(ii) [2.5 points] Nodes represent students in a class of high school graduates who graduated in 2010. All edges represent connections on social media, and roughly correspond to friendships that existed during highschool. Person 13 is trying to send person 89 an invitation to a high school reunion. Furthermore, person 89's social media account is under an alias, and thus they cannot be found directly through the social media's search function.

Question 2: (15.5 points) We are running a search engine, and we wish to rank the following 4 webpages that are related to the user's query. The hyperlinks between the webpages are represented as directed edges.



(a) [5 points] Use linear algebra to directly solve for the final unscaled PageRank values in the following graph. Explain the steps you took to solve. (HINT: you **shouldn't** be repeatedly applying the update rule; instead consider the properties noted in our proof of PageRank's convergence)

(b) [5 points] Run the Hubs-and-Authority algorithm on the graph for 3 steps (show the results for each step). Initialize all hub values with 1. Leave the normalization of the hub and authority values for the end.

(c) [0.5 points] Given your answer to b), what's the final ranking of the search results?

(d) [5 points] Can you modify your method from a) to possibly directly solve for the final authority values of the Hubs-and-Authorities algorithm? If so, then what changes would you have to make? [Note: you don't actually have to solve, just outline how you'd have to adjust your method]

Question 3: (7.5 points) Recall the threshold model of spread from class, where all nodes begin by using some product B, except for a small set of initial adopters that use product A. Now, suppose we change the model as follows:

If a node v is using product A, then they obtain an intrinsic award of $r_A(v)$, as well as a reward $a(v)$ for every neighbour also using A, and a smaller reward $m_a(v) < a(v)$ for every neighbour using B

If a node v is using product B, then they obtain an intrinsic award of $r_B(v) < r_A(v)$, as well as a reward $b(v)$ for every neighbour also using B, and a smaller reward $m_b(v) < b(v)$ for every neighbour using A

(a) [5 points] Under this modified model, can we say that there will not be a complete cascade iff there is a blocking cluster of non-initial adopter nodes, for some corresponding definition of a blocking cluster? Explain. If yes, specify the corresponding definition of blocking cluster.

(b) [2.5 points] Let's say that on a given a day, at UofT there is a CSC303 study session in your Bahen tutorial room, taking place at the same time as a meeting of the chess club at Hart House. We are interested in the behaviour of CSC303 students, and have already collected a complete social network of the friendships between CSC303 students. Is the extension of the threshold cascade model defined above, run on this social network, a good and practical way to model this situation and student decisions? Briefly explain.

Question 4: (7.5 points) In class we saw that the distribution of in-links on the internet follows a power law. Suppose we're studying a subset of the internet where every page has at least one in-link. In this subset we find that, empirically, the number of in-links follows a power law with an exponent of 2.1 (i.e., the probability that a webpage chosen uniformly at random from our dataset has k in-links, approximately follows a power law with exponent 2.1).

(a) [5 points] Suppose that it's the case that, in the subset being studied, there are 1000 webpages with exactly 3 in-links. Estimate how many webpages have at most 3 in-links. Briefly explain.

(b) [2.5 points] Now suppose that you are designing a search engine. Give an example of a design decision that could exacerbate the inequality in the number of in-links a webpage has. Briefly explain.

Question 5: (13 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 perhaps 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, you will build a model for this scenario.

Assume that every employee has their own dedicated laptop.

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). e.g., if a laptop or adaptor breaks on day 1, then the earliest it can potentially infect another piece of hardware is on day 2

Assume that whenever employees visit a room, they have an independent 60% chance of trying to use the projector (some days have several presentations, other days are purely discussions and have no presentations). If they need to present and the room has several adaptors, then they select an adaptor within the room uniformly at random. If some hardware breaks when they try to present, then they give up on using the projector and spend the meeting trying to debug the problem, while any others may attempt to present.

It takes 2 days for someone to diagnose the problem and to send the hardware for replacement or repair. On the third day they get a temporary replacement that's always kept in good condition, and on the start of the 4th day they get the repaired hardware back. 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, it's being repaired on day 8, and it's back in the hands of users on day 9.

Below, employees are listed, alongside their schedule for days 1 through 10:

- Anita attends every day in room 100, and on days 2-4 in room 200
- Briar attends in room 200 from days 1-4, in room 300 from 4-8, and from 7-10 in room 100
- Cecilia attends in room 200 on days 3-6, and in room 400 on day 8
- Darius attends in room 100 on days 2-3, and in room 300 on days 4-6

Note, all the date ranges above are inclusive (i.e., they are presenting on the first & last dates in the range).

Most rooms have 1 adaptor, except for room 200 which has 2 adaptors

On day 0, the unimaginable happens. As the day is wrapping up, a careless user strikes a DVI adapter 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.

(a) [10 points] Draw a contact network for this scenario, and describe a reasonable set of rules that could be used to model spread in your network.

(b) [0.5 points] Is it possible for the adaptor in room 400 to break, and thereby cause all other hardware to break at least once? Briefly explain.

(c) [2.5 points] If we could reorder and/or add to the meetings that occur, and if we're sufficiently (un)lucky, then would it be possible for a node to reinfect itself through a length 5 cycle? Explain.

END OF ASSIGNMENT 2

If you are typesetting 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.