Midterm exam

Fall 2022

Analysis of Network Data Tuesday 10/25

Please do not turn the page until instructed to do so.

This exam has 6 questions.

You can pick 4 of the 6 questions to submit for a total of 100 points.

Every question you pick will be equally weighted.

If you pick more than 4 questions, 4 of them will be selected at random for grading.

There are 6 pages in the exam.

The exam is **a take-home exam**.

It is due by class time (2:00 pm Central time) on Tuesday 11/01.

You are **not** allowed to collaborate or communicate with others about the questions in this exam.

You are **perfectly allowed** to discuss and collaborate with me!

Answer every question to the best of your knowledge.

Make sure to show your work for **partial credit**.

Some questions (like, for example, Question 2) will ask you to use Python, Gurobi, networkx or the software of your choice. You do **not** have to solve this one if you prefer to avoid coding during the exam.

Good luck!

Question 1: Smaller proofs and algorithms

(a) A matching is a set of edges such that no two edges include the same node. An independent set is a set of nodes such that no two nodes are adjacent (connected by an edge). If M is a matching and $\mathcal{I}$ is an independent set, show that $|M| + |\mathcal{I}| \leq |V|$.

(b) A cycle and a cut-set (set of edges that separate the network in two sets S and $\overline{S}$) intersect each other in an even number of edges. True or False? Prove or provide a counter example to the statement.

(c) An edge is called upward critical if increasing its capacity (and only that capacity) by 1 unit will cause the maximum flow to also increase by 1 unit. Is it true that every network has at least one upward critical edge or are there networks that no edges are upward critical?

(d) An edge is called downward critical if decreasing its capacity (and only that capacity) by 1 unit will cause the maximum flow to also decrease by 1 unit. Is it true that every network has at least one downward critical edge or are there networks that no edges are downward critical?

Question 2: Emergency crew routing

Last winter, a number of cities in the north states of the US faced threatening ice, snow, and cold weather conditions that caused provision of electricity to become intermittent. As a measure to prevent that and improve societal welfare, state and city authorities decided to locate repair crews and equipment in certain strategic areas in order to quickly assess, respond to, and repair such damages.

We represent this problem below on a graph. Figure 1 represents all candidate locations where crews can be stationed (in red), all arcs are given travel times, and all power system components (in blue) that can fail. If a power component is located in the middle of a street, then only half the traversal time needs to be "paid" before the crew reaches the power station. Assume that no streets are ever closed down, and hence the repair crews can always travel to address the issues arising in any power system component.

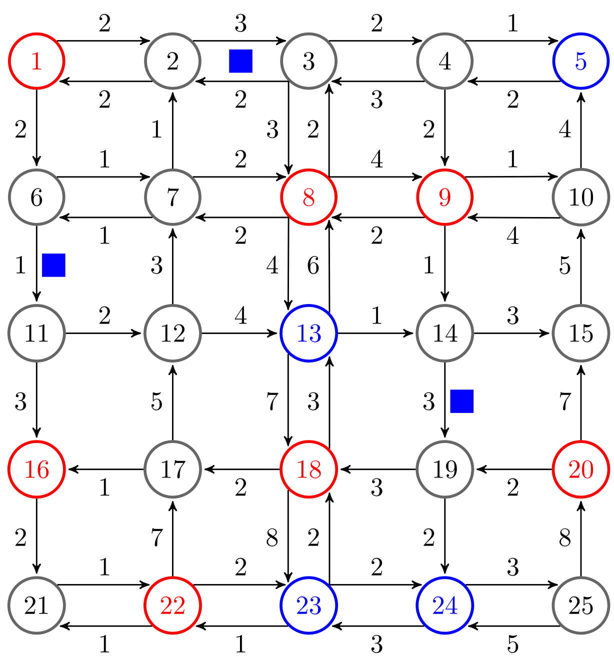

Figure 1: The network for Question 2. Note that the nodes in red are location candidates where your emergency crew can be stationed, while the nodes in blue are ones where damaged power station components can be found. If a blue node exists in the middle of a street, then you only need to "pay" half the arc time to get there from a starting node.

(a) Assume that your 2 crews are located in nodes 1 and 20. Formulate an optimization problem to evaluate the quality of the site selections, by deciding on which power systems get repaired by each crew, and the route that a crew would select to get to the damage site the fastest. Define your parameters, your variables, your constraints, and your objective function.

(b) Optimizing the same decisions as before, consider the case where the location crews are not known in advance, but instead are another decision variable. Formulate the problem under the assumption that the **2** crews are to be located on any node. Once more carefully define all your parameters, variables, constraints, and your objective function.

(c) Formulate and solve problem (b) in Gurobi.

Question 3: A crew scheduling problem

Consider the following airline crew scheduling problem. A crew can serve at most one flight at a time. When a crew serves a flight, it cannot serve another flight unless they arrive at their destination. As soon as the crew arrives in their destination airport, they can be scheduled to any flight starting from that destination or they can fly using an available flight to another airport from where they can get their next assignment.

We further assume that every flight has the following details:

$$(\text{Origin airport}, \text{Destination airport}, \text{Start time}, \text{Arrival time})\,.$$

To help you produce instances we provide you with a code called “flights.ipynb”: a screenshot with some indicative flights is provided in Figure 2. You may assume that times are all given in a universal clock setting; that is a number between 0 and 24.

Figure 2: A toy example for your crew scheduling problem.

#	Origin	Dest	Start	End
1	ATL	DCA	11.14	12.14
2	ATL	DFW	13.18	15.18
3	LGA	MCO	10.57	13.57
4	MCO	PHX	9.4	13.4
5	LAX	DFW	13.83	16.83
6	PHX	LGA	1.93	5.93
7	LGA	LAS	12.13	16.13
8	MCO	DEN	0.76	4.76
9	SFO	DCA	5.57	10.57
10	SFO	JFK	16.92	21.92
11	LAS	LGA	12.72	16.72
12	CLT	ORD	8.51	10.51
13	DFW	ATL	1.91	3.91
14	LAS	CLT	4.98	8.98
15	SFO	PHX	8.33	9.33
16	SFO	ORD	17.81	21.81

(a) Recommend an approach to solving the problem as a network flow problem when the goal is to *minimize the number of crews you need to use throughout the day.*

(b) Use the instance generator provided to solve a problem with $n = 50$ flights in networkx and report how many crews you need to accommodate all flights.

(c) Assume that some flights are bigger than others and need two or more crews. Based on the network flow problem you formulated in part (a), recommend an extension that you would implement to accommodate for that.

Question 4: Eulerian cycles and extensions

In class, we saw the Eulerian path and cycle problem. As a reminder, a graph is Eulerian if it has an Eulerian cycle. In an Eulerian cycle, we may start a path from a node and end in the same node after traversing every arc exactly once. Answer the following questions.

(a) Consider a simple and undirected graph $G(V, E)$. Assume you "double" every edge in the network. With "doubling" we mean that for every edge (i, j) you add a second copy of it and you now have two of each. Call this new (multi)graph $\hat{G}(V, \hat{E})$. Is $\hat{G}$ Eulerian or not? Prove it or provide a counterexample.

(b) Let us consider the following problem. A person is visiting an area to inspect all the streets (that is, they need to traverse all edges). They will park at a node and then start inspecting street by street by traversing them one at a time. They can traverse a street multiple times if they need to, but they would prefer not to. Based on your answer in part (a), devise a plan for them to find a cycle that uses at most $2|E|$ streets from and to their parking spot after having visited every edge in the network at least once.

(c) For the same problem in part (b), can you devise a polynomial time algorithm that produces a cycle from and to the parking spot that traverses every edge at least once and that is of cost at most $|E| + |V| - 1$?

Question 5: Matchings and networks

We assume that different professors can teach multiple classes in a program. That said, everyone has their preferences, and they would much rather teach a specific subset of the program classes. Assume for each professor $i \in \mathcal{I}$ we have a list with their preferences of courses $j \in \mathcal{J}$ that they can teach. In network terms, their preferences can be shown as in Figure 3, for example. Answer the following questions.

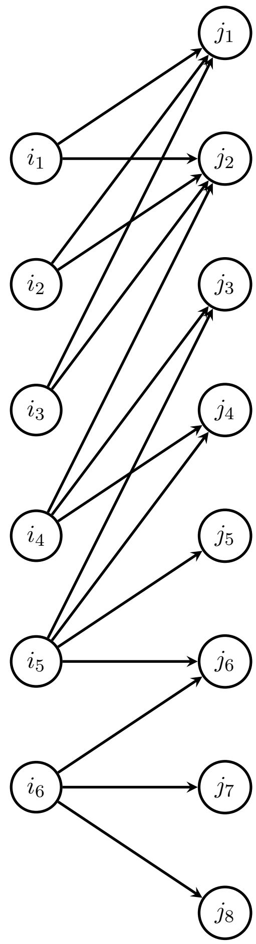

Figure 3: The preference network of Question 5.

(a) Recommend an approach to solving the problem as a network flow problem when the goal is to *maximize the number of professors who get a class they would prefer to teach.*

(b) Assume that all professors can teach all classes, but they provide a score from 0 to 10 on their level of preference (10 being the best assignment and 0 the worst). Based on the network flow problem you formulated in part (a), recommend an extension so that you may accommodate for preference scores.

(c) Show that a tree has at most one perfect matching. As a reminder, the definition of a perfect matching in a general graph is a set $M \subset E$ such that no two edges in M share a node.

Question 6: An evacuation planning problem

During an evacuation, people living in each of the locations in a network are to be routed towards one or many shelter (safe) locations. Assume you are provided with an undirected transportation network $G(V, E)$ (as an example, you can use the provided network emulating the main street arteries of West Virginia). Further assume that your shelter is randomly located somewhere in the state (i.e., simply select a random node s anywhere in the network to serve as your shelter). Answer the following questions.

(a) Consider the following problem: first, identify a minimum cost spanning tree in the whole network. This is ensured to lead everyone (eventually) to the shelter. Then, calculate the average time it takes every node to get to the shelter. Report this average time using networkx and the West Virginia network provided.

(b) The minimum spanning tree is a little too strict, in the sense that all nodes are equally important. Instead, assume that a set of nodes $S \subset V$ are the most important locations in the network (e.g., they have the biggest population; they are located in especially dangerous locations; they have more vulnerable populations; etc.). Solve a Steiner tree problem where you connect the set of nodes $S \cup s$ (i.e., the important locations and the shelter). Then, identify the shortest path from each of the non-important locations to any of the nodes in the Steiner tree. Again, report the average time it takes for each location to evacuation in the West Virginia network.

(c) Formulate the optimization problem of identifying an evacuation tree of minimum cost. This would be a tree, rooted at the shelter, where every node has **only one path to safety**. Formulate the problem and solve it using Gurobi and networkx. Report the average time it takes every node to get to safety for the West Virginia network provided.