COMP 200 & COMP 130
Assignment 6: Shortest Paths, Probabilistic Reasoning

Be sure to read the course policies, as posted on the course web site:

http://www.clear.rice.edu/comp200/policies.shtml

Work in assigned pairs on this assignment. Each pair should put all assignment answers in one Python file called `netid1_netid2_6.py`, substituting in the students' NetIDs. Put the answers to non-programming problems in comments. When submitting, only one of the member should turn in the pair's answers.

Total points: 100 for COMP 200, 150 for COMP 130.

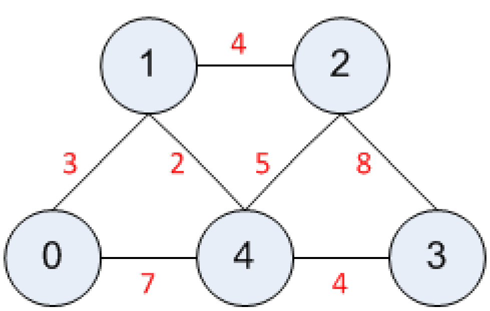
Wherever reasonable, use functions that have been defined previously on this assignment, on a previous assignment, or in class.

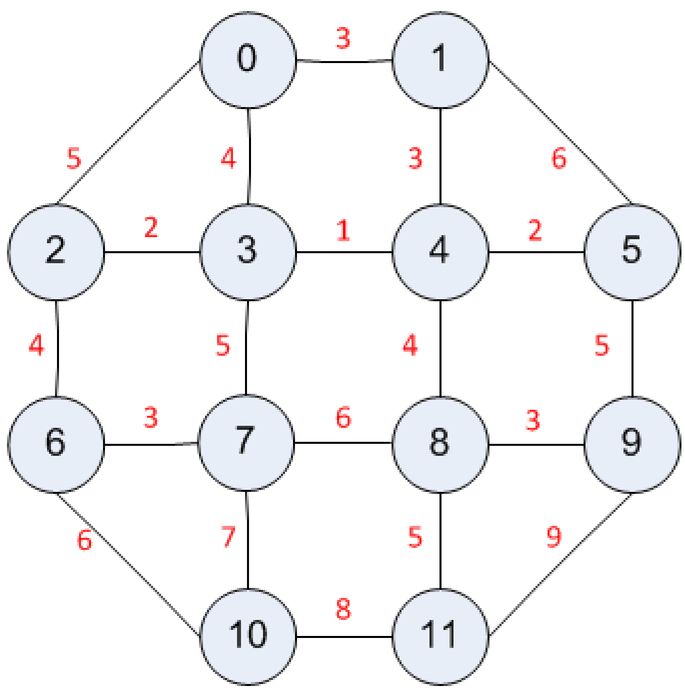
Dijkstra's Shortest Path Algorithm – COMP 200 & COMP 130 (80 points total)

For these problems, do not use any of the `networkx` functions for computing shortest paths or path lengths. Implement Dijkstra's algorithm yourself, and use your implementation.

1. (15 points total)

   Consider the two undirected graphs g_1 (left) and g_2 (right), respectively, shown here:





You should understand Dijkstra's algorithm without first having implemented its details in the next problem. In the following two parts, you will step through Dijkstra's algorithm in your head, and write some of the values that are calculated by the algorithm.

(a) (9 points)

For graph g_1, will we search the entire graph starting from node 0. *At the end of each loop iteration*, indicate which nodes have been visited, what is the lowest total distance currently known for each node, and what is the best path known so far to each node.

(b) (6 points)

For graph g_2, will we search the entire graph starting from node 0. *At the end of the algorithm*, indicate the lowest total distance currently known for each node, and what is the best path known so far to each node.

2. (40 points total)

Complete the in-class implementation of Dijkstra's algorithm. Your code should use the `networkx` library. It should use one of the priority queue implementations given in class.

(a) (20 points)

Complete the outline given in class that just returns the shortest distance from the given source node to the given target node. It returns `None` if there is no such path.

As described in class, the priority queue stores the lowest total distance for each node encountered so far.

(b) (20 points)

Provide a second version that also returns the corresponding path.

As described in class, the priority queue also stores the "parent" node for each node encountered so far. I.e., if in the best path known so far to t, we used the edge from s to t, then s is t's parent. Refer to the previous homework's sample solution for details.

3. (10 points)

Give an example that shows that Dijkstra's algorithm doesn't always work if the graph has any negative-length edges. Briefly explain the problem. Your wording should be in terms of the high-level algorithm, not the specifics of your Python code.

4. (15 points)

The provided `network.txt` file is the measured network topology of an Internet Service Provider. The edge weights approximate the latency (speed) of the network links, meaning that network traffic would be preferentially routed over links with lower weight (time for sending data).

You can load this graph via the following code:

```
import networkx
graph = networkx.read_weighted_edgelist("network.txt",nodetype=int)
```

What is the average shortest path length? Do not include any path from a node to itself.

Probability – COMP 200 ONLY (20 points total)

5. (20 points – 4 points each)

On the show "Let's Make a Deal", host Monty Hall set up the following problem for his wildly dressed contestants. There were three doors. Behind one was a fabulous prize, but behind the other two was something undesirable. The contestant picks a door. The host would open a different door, showing you that it didn't have the prize, and offers to trade your unopened door for the remaining unopened door.

Briefly explain your answer to each of the following questions.

(a) What is the probability that the prize is behind the door you initially chose?

(b) What is the probability that the prize is behind either of the two doors you didn't choose?

(c) What is the probability that the prize is behind the door he opens?

(d) What is the probability that the prize is behind the door he doesn't open?

(e) In conclusion, should you make the swap? Does it matter?

COMP 130 ONLY (70 points total)

As the head of marketing for SocialWidgets-R-Us, you know that college students, particularly progressive Rice students, are a prime market for your products. Your latest product is "Beer Bike Buddy" whose purposely obscured purpose will surely be a hit at Rice. But how to get the buzz going so that "BBB" will fly off your virtual store shelves? A time-honored tradition is to hand out freebies, but if you hand out too many, you will cut into your profits. So, given a limited number of BBB's to distribute, how can you distribute them most effectively? Luckily, Zuckerberg, et al., at Facebook have no stated mantras about "do no evil" but rather to "help you connect and share with the people in your life" and what better relates to Beer Bike than that? And so, minus an arm and a leg and the shirt off your back, but now armed with Facebook friends data, surely your computer science skills can tease out the optimal people to whom you should give your limted supply of BBB swag to maximize your capitalistic success.

Some business facts:

Each Beer Bike Buddy costs you $5, but is sold for $20. (Gouging students is another time-honored tradition!) All other costs can be considered fixed costs and thus can be ignored in your calculations.

Some behaviorial observations:

Obviously, a student given a free BBB won't buy one.

The probability of a student buying a BBB is proportional to and hyper-linear (i.e., stronger than linear, e.g., quadratic) in the number of their friends who have a BBB.

The probability of a student buying a BBB is proportional to and hyper-linear in the number of their friends who have heard of a BBB.

The likelihood of a person hearing about a BBB exponentially decreases with their friend distance from a person who has a BBB. That is, the likelihood of someone telling their friend about BBB gets less and less, the farther away they are from someone who has one.

The effect on the probability of buying a BBB is additive with respect to having friends that have a BBB and friends who have heard of a BBB.

Obviously, the probability of buying a BBB cannot be more than 1, but it can be zero.

6. (25 points total)

   Consider a single round of purchasing where you give out your free samples, and word about it spreads across campus.

   (a) How many free samples should you give out to maximize your profit during the first round of selling?

   (b) How many BBBs did you sell and what was your net profit, i.e., total sales minus the cost of all BBB's sold and given away.

7. (25 points total)

   Now consider multiple rounds of selling, that is, after students have gone out and bought BBB's, word about them spreads again. The free samples are only given out at the beginning, to start the process.

   (a) Calculate and graph your sales volume (number of units sold) and profits over multiple rounds of selling. Show total and per round figures.

   (b) Explain any phenomenon you see happening. Does what you see affect how many free samples you think you should give out in order to maximize your overall profits? Support your decisions with concrete simulation results.

8. (20 points)

   What is the probability that a a student's friend owns a Beer Bike Buddy if a student buys one? Use Bayesian analysis to compute this probability. Note: You may need to determine certain probabilities by analyzing your data after running a round of product buying as above. You may not determine the requested probability directly from the data otehr than to verify your Bayesian analysis results. Hint: Can you relate this problem to examples given in class?

Feedback – COMP 200 & COMP 130 (0 points)

1. Roughly how many hours did you spend on the two sets of finger exercises?
2. On a scale of 1 (very easy) to 5 (very difficult), how difficult were the finger exercises?
3. Roughly how many hours did you spend on this homework?
4. On a scale of 1 (very easy) to 5 (very difficult), how difficult was this homework?
5. Which material did you find most challenging?
6. Did you feel that the class material adequately prepared you for the homework?