# Assignment 2

CSC 303: Social and Information Networks

First 5 questions posted February 18,2020
Due: March 16, 2020, 2:59 PM

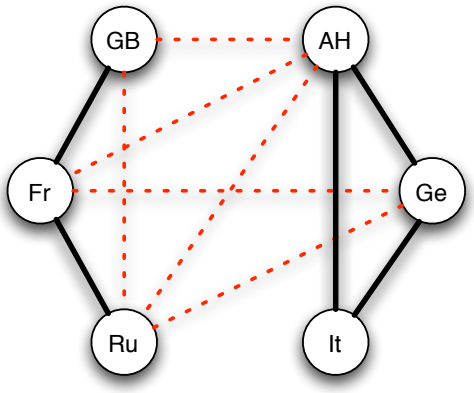
**Be sure to include your name and student number with your assignment. All assignments are to be submitted on Markus by the due date.**

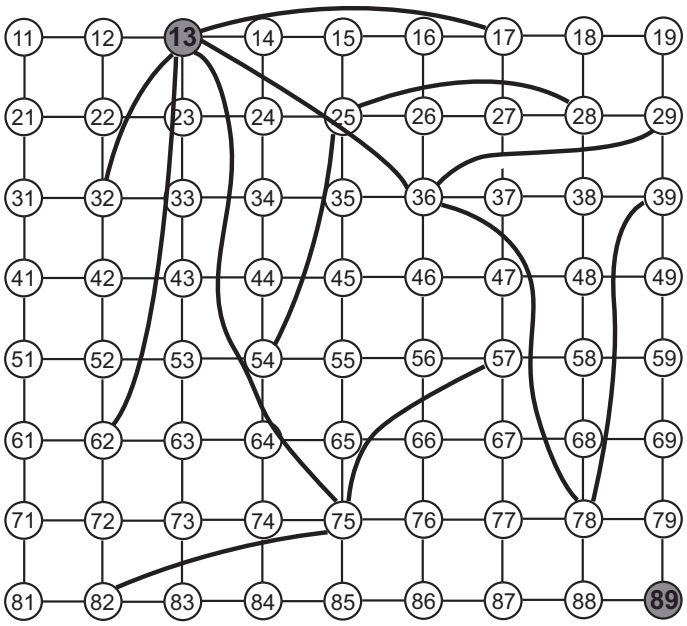
**You will receive 20% of the points for any (sub)problem for which you write "I do not know how to answer this question." You will receive 10% if you leave a question blank. If instead you submit irrelevant or erroneous answers you will receive 0 points. You may receive partial credit for the work that is clearly "on the right track."**

1. (10 points)

   Consider the above signed network representing countries (in the early 1900's ) where a solid edge denotes friendship (an edge labeled $+$) and a dashed edge represents enemies (an edge labeled $-$). No edge represents current neutrality. But sometimes, every country has to choose sides.

   Specify the minimum number of edge signs (if any) that need to be changed so that the network can be completed to be a strongly balanced network. Provide an explanation as to how the initial or modified network can be completed to be strongly balanced.






2. (10 points) We imagine a community where people are arranged on a grid, and interact almost exclusively with their local (vertical and horizontal) neighbors. However, a few people "travel" and develop (symmetric) friendships with others that are more distant on the grid. The graph above represents people as nodes, where nodes 13, 25, 36, 75, and 78 are the "travellers." Friendships are represented as edges on the grid. Homophilous friendships between local neighbors are shown using thin edges, while the friendships (or weak ties) involving connections between the travellers and non-local contacts are shown using bold edges. For example, node 25 (a traveller) has developed friendships with non-local nodes 28 and 54.

   We're interested in the process of decentralized search on this graph involving node 13 trying to communicate a message to node 89 (the two shaded nodes). 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.

   - 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?
   - 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?

3. (15 points)

   A music retailer has begun and is promoting an online site for downloading popular international songs not well known in Canada. The retailer is charging only $.25 cents per download. The retailer will not let you listen to the music but does provide lots of information on internation sales.

   After the first month, the retailer has sold 70,000 downloads amongst the first 1000 top downloads. The retailer has also observed that the number of sales is satisfying a power with exponent $-1$.

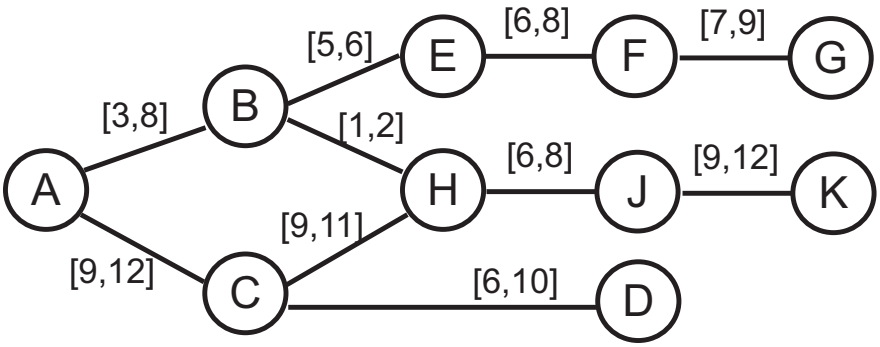
   - (5 points)Estimate the number of sales of the most popular download.
   - (5 points) aEstimate the total number of sales of the top 100 downloads.
   - (5 points) Provide an explanation as to why the number of downloads is following this power law.

4. (20 points) Question 3 of Chapter 14 in the EK text. Parts (a) and (b) are worth 5 points each and part (c) is worth 10 points.

5. (10 points) Question 6 of Chapter 14 in the EK text.

6. (10 points) Consider a moderate size sparse (in terms of the degree) social network with say around 10,000 people and say around 10 friends per person. We have some evidence that these are politically independent voters are trying to decide about their vote for in a local election. We believe voters are positively influenced when a candidate comes personally to answer questions. And we know that even if someone wasn't visited, they are impressed if a friend was visited. The candidate only has time to vist around 100 people. As the campaign manager, you only have the social network (as a graph) and no other information about people in the network.

   State two methods for choosing which people should the candidate visit. What are the relative benefits for each method? Explain breifly.




7. (20 points) Consider the contact network above in which the earliest and latest period of contact (in days) between two people is shown on each edge. Suppose we are contending with the spread of a very contagious flu such that a person who has this flu will transmit it with probability $\frac{1}{2}$ on any given day of contact during a time $[t, t+1)$ . For example, Since $B$ and $E$ have contact at time $t = 6$, the probability is $\frac{1}{2}$ that $B$ would transmit the flu to $E$ at time $t = 6$ if $B$ were still contagious and then $E$ becomes contagious on day 7. Assume person $A$ has this flu on Day 0 and is contagious starting on Day 1.

   - (20 points) Suppose anyone who gets the flu will remain contagious for 15 days (which for this network is the same as staying contagious indefinitely). For each person (i.e., node) $X$ in the network, determine the probability that $X$ will get the flu. Indicate how the flu is possibily being spread to each person.
   - (0 points) **NOTE:** I am removing the second part of this question. The first part sufficiently illustrates the contact network concept. You might want to think about this part of the question as preparation for the final.

     Suppose that prior to $A$ getting the flu on Day 1, two vaccines have been developed. The vaccine $V_1$ limits the period of contagion to 3 days but does not change the probability of spreading during the contagious period. For example, $A$ will be contagious on days 1,2,3. Vaccine $V_2$ does not change the period of contagion but lowers the probability of spread in a given day of contact to $\frac{1}{4}$. If we measure the success of a vaccine by the expected number of people who do not get the flu, determine which vaccine is more effective. That is, we apply exactly one of the vaccines to every node (person) in the network and then determine which vaccine is more effective.