CE C88/CP C88 - Data Science for Smart Cities

Assignment 1

Spring 2021

Due date: February 10, 2021, 11:59 PM

Submission instructions:

- Please make sure to submit both a summary of your answers to all questions in a **PDF** and a separate **IPYNB** (Jupyter notebook) file containing your code on bCourses.
- Name your file using these format: *LASTNAME_FIRSTNAME_HW1.pdf* and *LASTNAME_FIRSTNAME_HW1.ipynb*
- Don't forget to write your name on the top of your PDF file submission.

Note: See Reading_SocialNetwork.ipynb and MyFirstNetworkExercise.ipynb for guidance in doing this homework.

Generate CE C88/CP C88 Network

Nodes:

Fill in the Google Doc (https://docs.google.com/spreadsheets/d/1XOcuDb5g9tFxRFaKaoG3ynvPLiSDBGDydojIC49bAXY/edit?usp=sharing) by adding your information including your major (*e.g.,* Data Science BA) and no. of terms in attendance (including this semester), and let's wait until almost everyone filled in their information until ***February 2nd, 2021, 11:59 PM***.
On ***Wednesday, February 3rd*****,** at **09:00 AM**, you should be able to download ***ce88_cp88_nodes.csv*** which is the finalized version of the Google Doc you helped us populate with your information. This shall now define your nodes for the network.

Edges:

Fill in this Google Doc (https://docs.google.com/spreadsheets/d/1rqeT006mtN0BHL2L_X4DMFSUtCePWoqO7h2nbSSrZdA/edit?usp=sharing) by adding your node_id based on the Google Doc above as the source node on four (4) consecutive rows (*i.e.*, your node_id should appear four (4) times on the 'source' column) and selecting a unique 'target' number chosen in random between 1 to 57 (excluding your own node_id) for each row. Please do this until ***February 2nd, 2021, 11:59 PM***. Download ***ce88_cp88_edges.csv*** from bCourses which contains the 'source' and 'target' columns on the same day that you'll download the nodes file which is on ***Wednesday, February 3rd*****,** at **09:00 AM.** This file shall be used to define the edges of the network.

Create the class network using the nodes and edges described above to generate an undirected graph (Graph) and answer the following questions:

Problem 1: ***Check if the CE C88/CP C88 Network has Small World property (7 pts)***

1) Fill the table below with the properties of the network such as number of nodes, number of links, average clustering coefficient, average degree, average shortest path based on the network generated on the CE C88/CP C88 Network from Problem 1. (2 pts.)

2) Using ***watts_strogatz_graph()*** from Network with ***random seed = 123,*** generate a Small World Network Model that matches its number of nodes and average degree of the CE_CP_C88_Network and add the information of the table. Also, select ***p*** based on the calculated value from $C(p) = C(0)(1-p)^3$ where ***C(p)*** is the *average clustering coefficient of the empirical network* and ***C(0)*** is the *average clustering coefficient at p = 0*. (2 pts.)

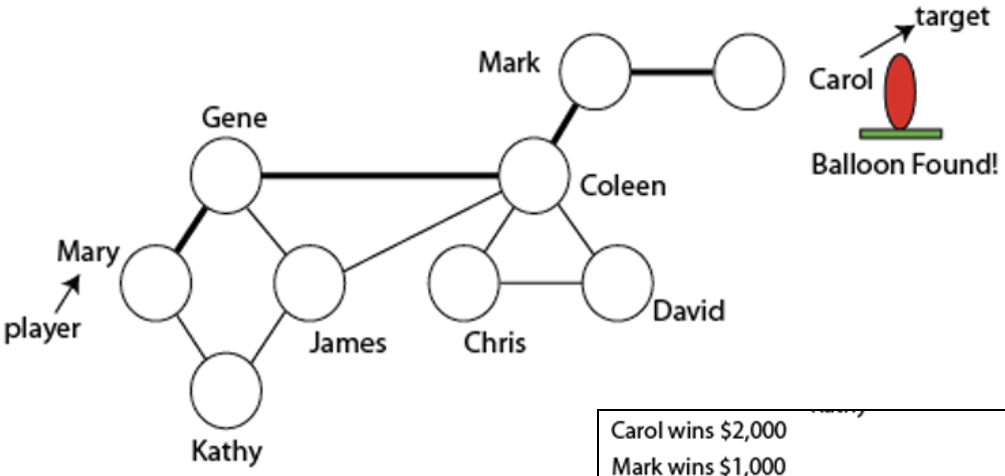
Network	# NODES	# LINKS	<C>	<K>	<L>
CE_CP_C88_ Network					
Small World Network Model *(p = _____)*					

3) In the same figure, plot the degree of histograms of both the CE C88/CP C88 empirical network and the Small World Network graph. (2 pts.)
4) Does the CE-CP-C88 empirical network closely resemble the Small World Network based on both the histogram of the degree and table generated? (1 pt.)

Problem 2 (8 pts)

This idea is based on a DARPA network challenge in December 2009. As a brief summary of the challenge, participants were tasked to find 10 red weather balloons distributed all throughout the US. This challenge was designed as a crowdsourcing kind of activity where people used social networks and developed social media tools on solving this problem. You can also watch this YouTube video to learn more about the challenge: https://youtu.be/gELZYrjIKEI.

A team from MIT won the challenge. They created a system where you and the people who helped you (*i.e.,* being able to recruit people to be a part of a chain who finds a balloon) get prize money for finding the balloons (Chu, 2011). Below is a sample structure of how the money is given on a chain that successfully finds a balloon:



Description of the Game: One `player` signs up for the game and informs the contacts in his/her social network. If someone in the network (`target`) finds the balloon, he/she receives a prize of $2000, additionally, everyone in a chain connecting the target with the player gets half of the next person's reward. In the example above, Carol (`target`) found the balloon so she gets $2000, Mark invited her in the search so he gets $1000, but Mark was recruited to join the search by Coleen so she gets $500, Gene asked Coleen to be a part of it as well so she gets $250, and finally, Mary (`player`) wins $125 and the total cost of the chain is $3875.

We are going to adapt this challenge and do somewhat the same balloon search in class using the network that you'll build in Part I; and, in Part II, you should determine how much it will cost to find the balloons in the network created in Part I.

Reference: Chu, J. (2011, October 28). *Searching for balloons in a social network: The key to mobilizing large numbers of people is incentives, study finds.* MIT News Office. https://news.mit.edu/2011/red-balloons-study-102811

Part I. The Network (4 points)

1) Report the different "Majors" and "no. of students" in each Major? (1 pt.)
2) Report the different "Terms" and the no. of students based on the number of their terms in attendance? (1 pt.)
3) How many nodes and links are there in the network? (0.5 pt.)
4) What is the average degree of the network? (0.5 pt.)
5) Determine the maximum in-degree of the network. Who is/are the students with the largest in-degree in this network? (1 pt.)

Part II. The Game (4 points)

The game described above is hypothetically done in class. But instead of 10 balloons, there are only (three) 3 balloons. All the students are supposed to be players but only the chain started by students 9, 42, and 8 successfully found one balloon each. While the lucky students (*disclaimer: these numbers are randomly picked*) who found the balloon are students no. 3, 21, and 50, respectively.

1) If it takes the shortest path to reach the target from a player, fill in the table below by writing the sequence of students who will receive money from finding each of the balloons from the `player` to the `target` (write all possible paths) and the cost of chain for finding each balloon if the same rule applies in getting money based on the description above. (3 pts./0.5 pts. per cell)

Balloon	Player	Target	Chain created (from player to target)	Cost of chain
1	9	3		
2	42	21		
3	8	50		

2) How much is the total cost of the chains for finding all the three balloons above? (0.5 pt.)
3) If there is a fourth balloon search which was started by Student No. 18 and ends with Student No. 11 finding a balloon, how many shortest paths can there be? How much is the cost of a chain created? (0.5 pt.)