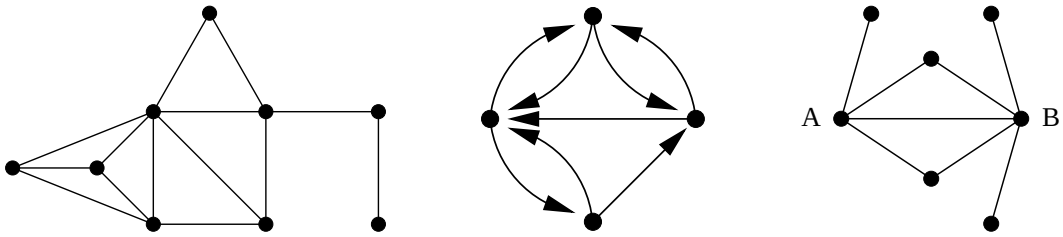
Complex Systems 535/Physics 508: Homework 3

1. Consider these three networks:



   (i) Find a 3-core in the first network.

   (ii) What is the reciprocity of the second network?

   (iii) What is the cosine similarity of nodes A and B in the third network?

2. A 2016 article on the website FiveThirtyEight.com described the results of a study by Eitan Hersh and Yair Ghitza of political alignment among heterosexual couples in the United States. Hersh and Ghitza estimated the fractions of partners with each combination of major-party alignment (Democrat, Independent, or Republican) to be as follows:

		Women			
		Democrat	Independent	Republican	Total
Men	Democrat	0.25	0.04	0.03	0.32
	Independent	0.06	0.15	0.05	0.26
	Republican	0.06	0.05	0.30	0.41
	Total	0.37	0.24	0.38	

   Assuming these results to be representative of the network of relationships, calculate for Democrats, Independents, and Republicans the numbers e_r and a_r that appear in Eq. (7.61) of the course pack. Hence calculate the modularity of the network with respect to political persuasion. What do you conclude about homophily in this network?

3. What (roughly) is the time complexity of:

   (i) Vacuuming a carpet if the size of the input to the operation is the number n of square feet of carpet?

   (ii) Finding a word in a (paper) dictionary if the size of the input is the number n of words in the dictionary?

   Explain your reasoning in each case.

Continued over the page →

4. To complete this homework problem you'll need to use network analysis software, so your first task is to download and install the network analysis software *Gephi*, which you can get from `www.gephi.org` (for Windows, Mac, or Linux).

   Next, download the results of the in-class social network questionnaire from here:

   `http://www.umich.edu/~mejn/courses/2017/cscs535/classnet.txt`

   Note that the network is directed: person A may know person B but not *vice versa*. Let us define an undirected *symmetrized network*, in which there is an undirected edge between nodes i and j if there is a directed edge either from i to j *or* from j to i.

   In order to analyze this network, you will need to enter it into Gephi. You can enter networks into Gephi by hand, but the program can also read network data in many file formats, and you will probably find it easier to convert the data into a form Gephi understands and then read it in directly. (I used some Emacs macros to turn the raw data into a simple adjacency list form, then I wrote a short Python program to convert that to the GML file format and then loaded the result into Gephi. But everyone has their own way of doing these things.)

   (i) When you load up the network the program will produce a picture of it, but the default picture is not very clear. Work out how to do a clear visualization. (Hint: Try out the different vizualization algorithms and see what works best for you.) Work out how to zoom in and out of the picture to examine different parts of the network. Work out how to turn on the node labels, so you can see which nodes correspond to which individuals. You may also need to change the size of the labels. (Hint: Look at the buttons and sliders along the bottom and sides of the network.)

   (ii) What are the numbers n and m of nodes and edges in the (symmetrized) network? Hence what is the average degree?

   (iii) Calculate the degrees of all the nodes and make a histogram of the degree distribution.

   (iv) What fraction of the network is occupied by the largest component? What is the diameter of the largest component?

   (v) Calculate the eigenvector centrality of every node. Work out how to color the nodes on the screen according to eigenvector centrality so you can see which have the highest score. (Hint: the "Nodes" panel on the left is a good place to start.) Save a copy of the resulting visualization. Which three people in class have the highest eigenvector centralities?

   (vi) Find the "communities" in the network using the modularity maximization method and make a picture of the network with node colors representing the communities. Roughly speaking, what is the computer doing when it performs this calculation? (We will look at community detection in networks in more detail later in the course.)

   (vii) Do one other cool thing with this network using Gephi—calculate some interesting thing, or produce some interesting visualization. Explain what you did.

   For full credit turn in your answers to the questions in parts (ii), (iv), (v), and (vi), your plots from parts (iii), (v), and (vi), and your results and explanation from (vii).