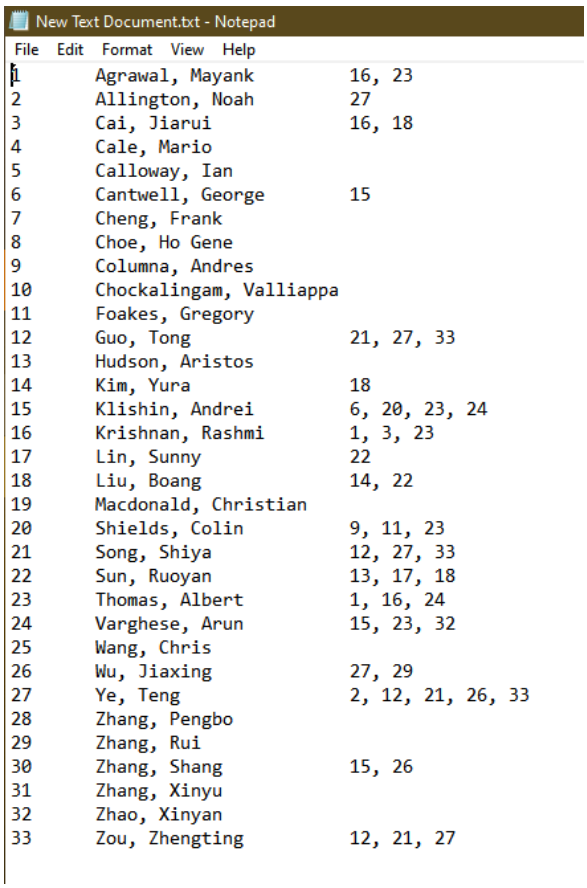
Homework 2

complex networks

Here are the results of the in-class social network questionnaire(I'll send you text file.):



Note that the network is directed, e.g., node 3 knows node 18 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*.

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* (www.gephi.org, available for PC, Mac, or Linux). You second task is to work out a way to turn the raw network data into a symmetrized undirected network and load it into Gephi. You can enter networks into Gephi by hand, but the program can also read network data in many file formats, so you may find it easier to convert the data file into a form Gephi understands and then read it in directly. (When I did it I used Emacs macros to turn the raw data into a simple adjacency list form, then wrote a short 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 visualization 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 vertices and edges 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) 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), and (v), your plots from parts (iii) and(v), and your results and explanation from (vi).