

CNT 5805 Project 2 Questions and Guidance

Please answer all of the questions that are bullets under each numbered paragraph below. Create corresponding numbered sections in your report. You must put your name and page numbers in your report.

1. The Zachary's karate club is a data set prepared by Professor WW Zachary which was published in a 1977 article. It is based on a true story in which the Professor experienced and is a classic network design problem that uses a process called community detection for analysis. You may download the file here [Gephi Github open source database](#). After downloading it, open it up with a text editor so you become familiar with the data. There are only 34 nodes and each node represents a member of the club.

Read the Professor's original article on the club (WW Zachary, An information flow model for conflict and fission in small groups, *Journal of Anthropological Research* 33(4), 452–473 (1977). It will explain the problem and give you some insight into community detection. You can find the article on-line either on the open Internet or in the UCF Library.

☐ Summarize the article (maximum of one page).

☐ Explain your purpose (e.g., inform, persuade, educate, entertain, predict, etc.) for analyzing this network? Try to use one of the terms for your purpose in the preceding sentence.

-The article goes into detail explaining the division of Zachary's Karate Club. The club was divided into two separate groups when a fission was created as a result of a disagreement between the two club leaders. When the disagreement happened, Mr. Hi left and along with his loyal students. The member of each of the clubs are represented as the nodes while the flow of communication is represented as the links between the students. The paper goes into more detail concerning the distribution of communication and distinguishes different types of nodes such as sinks and sources of information.

2. Now open the file using Gephi.

☐ What can you ascertain from the initial graph you see? Include a diagram of the initial graph in your report.

-Based on the initial upload, the karate community looks like a smaller rural community with only a handful of people with more than four or more links.

3. Now switch to the Data Laboratory view and interpret the spreadsheet you see.

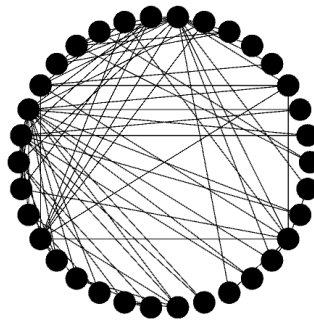
🔍 What have you learned from looking at this data?

-It seems clear looking at the Data Laboratory that the community is small and under forty nodes. Like I had hypothesized before, there are only a handful(3) of nodes that have more than four links.

4. Now run three to five Layout algorithms.

🔍 Show a small screenshot of each one and briefly explain what changed and why?

🔍 At this point, which Layout seems most useful and why?



🌟 **Figure 1. Circular Algorithm**

🌟 **Running the Circular Algorithm evenly spreads the nodes into a circle and highlights the distribution of the links.**

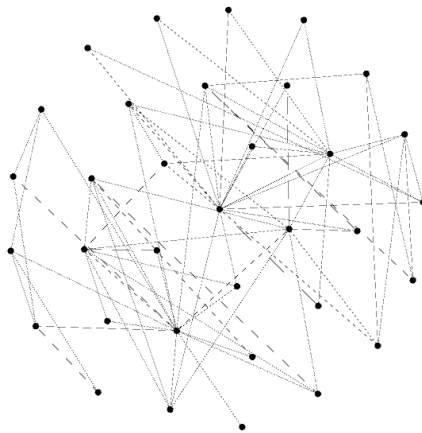


Figure 2: Fruchterman Reingold Algorithm

The Fruchterman Gold Algorithm evenly distributed the nodes into a circular grid with the most commonly linked nodes toward the center of the circular.

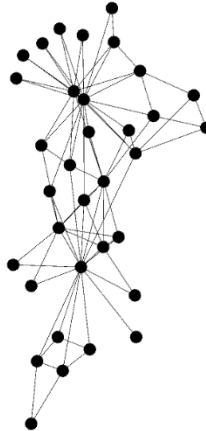


Figure 3: Yifan Hu Algorithm

With the most commonly linked nodes at the center, the Yifan Hu Algorithm strings out the nodes into the different clusters while highlighting the facets of information.

5. Now add some emphasis to your diagram by sizing, color coding, and naming the nodes.

☐ Show a screenshot of this action.

☐ What have you learned from your new diagram?

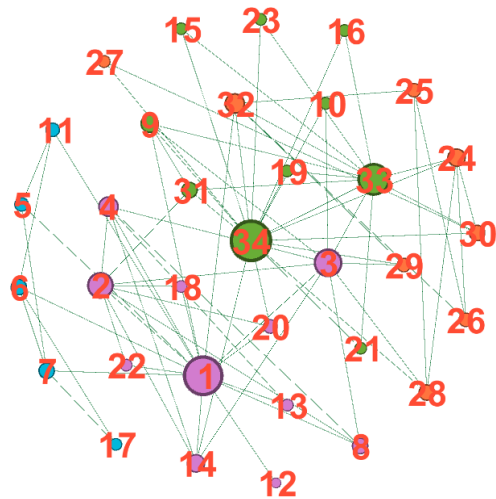


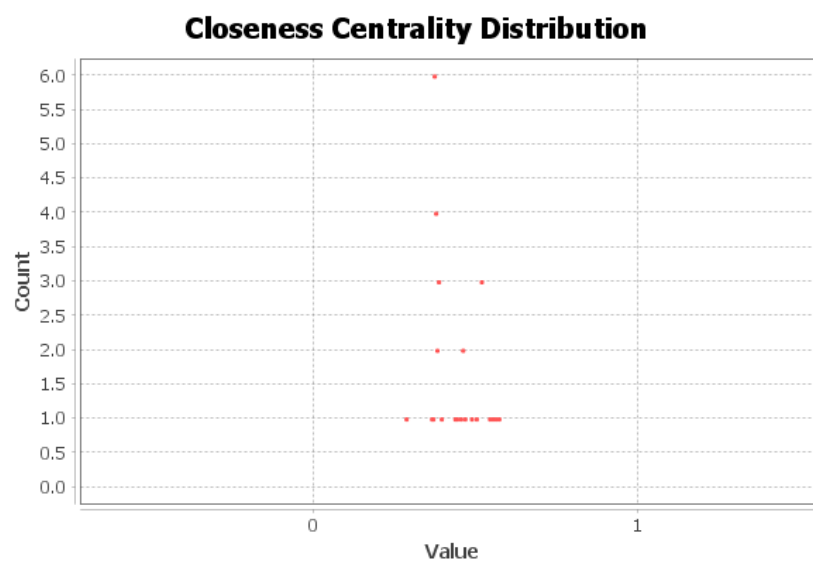
Figure 3. Color-coded, sized and weight added

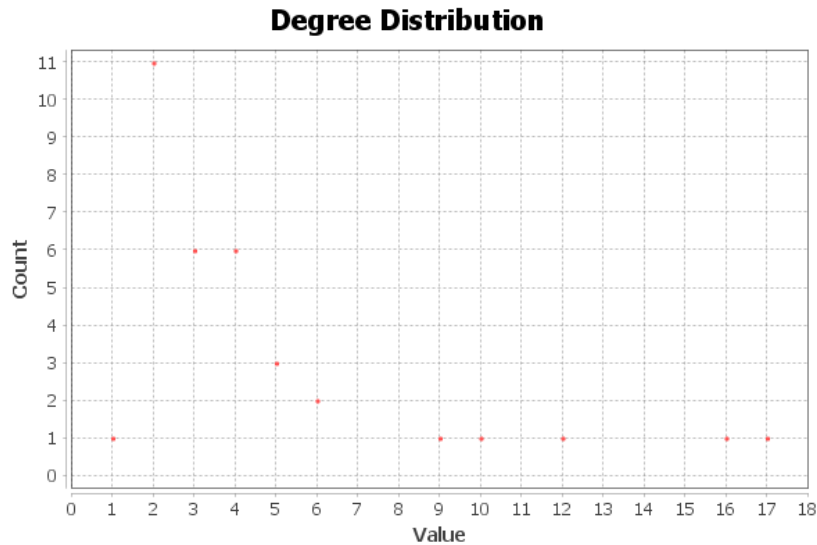
The new diagram highlights the degree of each node really well by showing the increased amount of links.

6. Now run some statistics about the network such as connectivity, centrality, clustering, etc.

You decide which ones are most relevant.

📝 Explain what statistical results you found. Please show graphs and/or numbers generated by Gephi in your explanation.





The Closeness Centrality Distribution shows how far one node is from the other nodes. In this case, it's easy to pick out who the members of the club that are furthest from the information distribution versus the members that are the closest.

The Degree Distribution was chosen to highlight the different faucets of information throughout each of the classes. By looking at the count of links for each degree, we can see who the person was who distributed the most amount of information. This person was likely the source of any information.

☐ Are there any communities or giant components, please explain?

There is a clear distinction between four groups in the classes with two of these groups being the main classes.

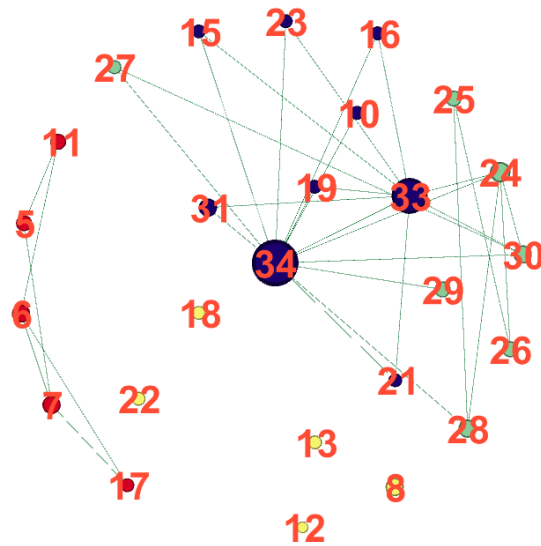
☐ Do you see any homophily or density in your network, please explain?

-Density is shown by comparing the number of edges to the maximum possible number of edges in the graph. By running the graph density statistic, the result turns out to be 0.139 which is relatively low.

7. Now run some filters on the network such as connectivity, centrality, clustering, etc. You decide which ones are most relevant.

☐ Explain what results you found. Please show graphs and/or numbers generated by Gephi in your analysis.

-Using the Eigenvector Centrality filter and playing with the range makes it easy to see the number of connections. The number of links decrease as the eccentricity increases.



8. Run Modularity Analysis to come up with your final graph. Suggest you review YouTube videos by Dragan Gasevic and/or jengolbeck for further guidance.

☐ What have you found?

-Once the Modularity Analysis runs, the communities within the network become apparent. In the case of the karate club, four communities were found with two distinct ones.

9. Final contemplation (three paragraphs maximum).

☐ Can you think of any other relevant network data that would have enhanced or made the study more interesting? Please explain.

☐ Summarize what you learned in doing this project.

-A larger study over a longer period could have given the project more depth or even a larger school. Overall, it was interesting to learn the ways that information can flow throughout a group and which nodes show influence in a community. From source to sink, the information can journey through many different paths.

10. In submitting your project assignment: You must upload your paper into WebCourses in PDF

format and your file name should follow this naming convention: last name-first nameCNT5805-research project2.pdf