**SNA – Spring 2023**

**Lab 1b: Descriptive Network Analysis – Local and Global Properties**

**Deadline: April 27th at 3:29 pm CT**

**Deliverables to submit on Canvas:**

1. Your report as a .pdf file
2. Your code as a .R file
3. Your data as a .Rdata file

Please upload each file separately – do not upload as a zip file!!

**Assignment Description:**

In Lab 1a, we created and visualized semantic networks from both collective and artificial intelligence text sources. In this lab, you will use descriptive network analyses to compare these two networks. We will continue using the statistical software package R. We will begin by looking at the local differences in the positions of nodes in the two graphs. This analysis will answer questions like, “are the most central terms in the collective intelligence network also the most central in the artificial intelligence network?” Then we will look at the global differences in the overall structure of the two networks. This analysis will answer questions like, “which of the two networks has a higher density?”

**General Instructions:**

1. For this lab, you will continue using the same dataset that you gathered in Lab 1a. You should be able to reload the exact same data using the RData file you generated and submitted in Lab 1a.
2. This lab has two parts indicated with Roman numerals (**III, IV**) in the outline below (continuing from parts I and II in Lab 1a).
3. Prepare a report that includes your responses to **all the questions** in parts **III and IV** outlined below. Label your responses with the instruction and prompt number (for example, “9”). Incorrectly labeled responses may receive a lower grade. For each response in your report, you should report your results and interpret them as specified in the prompt. Insert network images into your report in the appropriate places. In RStudio, you can click “Export / Copy to Clipboard” and paste directly into the Word document. You will be graded primarily on the completeness and accuracy of your responses, but the clarity of the prepared report will also affect your grade. While students may work together to perform the analysis, each student must execute his or her own code, and is responsible for writing the narrative in the report and submitting it.
4. Upload your report as a PDF, R code script and RData file to the Lab 1b assignment in Canvas. Please delete the instructions from your final hand-in.

**PART III: Local Network Properties (25 points)**

In this part, you will compute individual-level network measures and identify some key users in your network. Further, you will conclude this lab exercise with discussion in your main findings based on the visualizations, measures and your own analysis.

**Local Network Properties Instructions:**

For both the collective and artificial intelligence graphs:

1. Generate the giant component graph (only the single largest component) for your networks from Lab 1a. **You will use the giant component graph, rather than the full network, for all analysis in this lab**.
2. For each node in the networks, calculate different centrality measures: (a) in-degree, (b) out-degree, (c) betweenness, (d) in-closeness, (e) out-closeness, (f) eigenvector, (g) Burt’s network constraint, (h) hub score, and (i) authority score.

**Local network properties questions to answer in your assignment:**

1. **(5 points)** Briefly describe each centrality measure. How is each computed and what does its number mean in your network (e.g., a high centrality score means…)?

For both the collective and artificial intelligence graphs:

1. **(10 points: 5 points per network)** Provide a table ranking the top 5 nodes in your network on each centrality measure. Each centrality means (a) in-degree, (b) out-degree, (c) betweenness, (d) in-closeness, (e) out-closeness, (f) eigenvector, (g) Burt’s network constraint, (h) hub score, and (i) authority score.
2. **(6 points: 3 points per network)** How does the centrality of nodes vary with different types of centrality metrics? Why is this the case? Please offer some potential explanations using certain nodes as examples.
3. **(4 points)** Compare the local properties of the two networks. Think back to what you gleaned from your network visualizations in Lab 1a. Do your results in this part of the assignment align with your expectations or do they surprise you? Please explain.

**PART IV: Global Network Properties (40 points)**

In this part, you will compute global properties of your network, such as subgroups, which provide a plethora of information to social network researchers. A variety of algorithms have been developed to identify and measure subgroups. You will use some of igraph’s built-in tools to identify subgroups and central nodes for visual inspection.

**Global Network Properties Instructions:**

For both the collective and artificial intelligence graphs:

1. Calculate the coreness of each node in the giant component graph. Plot the graph and color nodes based on their coreness.
2. Run a community detection algorithm for the graph. If you want to use another algorithm, replace with your choice. Check how many communities are created. Calculate the modularity score using these communities. Plot the graph using the community detection results. Make sure that your nodes are colored based on the communities.
3. Create a plot for the in-degree distribution of the graph. Create a log-log plot based on the in-degree distribution. Calculate a power law fit to the in-degree distribution. Hint: use ‘power.law.fit()’ in ‘igraph’.
4. Create a plot for the out-degree distribution of the graph. Create a log-log plot based on the out-degree distribution. Calculate a power law fit to the out-degree distribution.
5. Compute the clustering coefficient and the average path length for the graph. Also, compute the clustering coefficient and average path length for 1,000 randomly reshuffled networks based on the graph. Plot the distribution of 1,000 simulated clustering coefficient values from the reshuffled networks and add the vertical line on the plot indicating the value of average path length from the graph. Create the same plot for average path length.
6. Run a one-tail t-test to examine whether the value of clustering coefficient from the graph is different from the simulated distribution. Run the same t-test for average path length.

**Global network properties questions to answer in your assignment:**

For both the collective and artificial intelligence graphs:

1. **(3 points)** Briefly describe (a) what k-core is, (b) what insight this k-core decomposition method provides, and (c) what is the highest/maximum level, k, of cores present in your networks (e.g., Do any 3-cores exist in your networks? Do any 4-cores? 5-cores? etc.)?
2. **(3 points)** Visualize your networks using k-core decomposition and include the visualizations in your report. In a few sentences, discuss your interpretation of the visualizations. Do the results of k-core decomposition make sense based on your expectations of the network?
3. **(3 points)** Pick one community detection algorithm to run on your networks. Which community detection algorithm did you choose and why?
4. **(3 points)** How many communities have been created in each network? For your networks, what might a community of nodes potentially have in common? Do these align with the topic modeling communities you visualized in Lab 1a?
5. **(3 points)** What is a modularity score? Interpret the modularity score of your results of community detection.
6. **(6 points)** Plot the communities and include the plot image in your report. What information does this layout convey? Are the communities well-separated, or is there a great deal of overlap? Describe the actors between any components and cliques (i.e., brokers). What are common features of these actors?
7. **(3 points)** Present and interpret the in- and out-degree distribution based on your networks as well as a log-log plot. Compute and interpret the estimate of the slope (the *alpha* value). Note that a *p* value (KS.p) less than 0.05 indicates the empirical data does *not* fit with the power-law distribution.
8. **(8 points)** To determine whether your observed networks demonstrate small world properties, (a) Plot the observed and simulated values for the global clustering coefficient based on the original network and 1,000 randomly shuffled networks. Conduct a t-test to check whether the observed value is statistically different from the simulated distribution. Then, (b) Plot the observed and simulated values for the average path length based on the original networks and 1,000 randomly shuffled networks. Conduct a t-test to check whether the observed value is statistically different from the simulated distribution. Finally, (c) based on these results would you conclude that the observed network demonstrates small world properties? If so, why? If not, why not? Include the plots and t-test results in your report.
9. **(8 points)** In two or three paragraphs, discuss your major findings based on all the analyses you’ve done between Lab 1a and Lab 1b. Your answer here will be evaluated based on depth and comprehensiveness. Thus, you’re encouraged to utilize extra information to answer this question. For instance, you can take a look at your original text data. Similarly, if you need more insights from your network, feel free to run other types of analyses based on your data collection.