**4602 Extra Credit Project**

Introduction:

A recent study found that the amount of people our brain allows us to know has expanded due to our social networking. The amount of network that is done via social media has increased rapidly, allowing us to stay in touch with people near and far. We’re now able to “@” people or tag them in certain posts which allows our algorithms to determine just how close we are to certain people.

Donald Trump is notorious for his tweets that come out almost daily. Trump typically enjoys @-ing his favorite users, which include users like Fox News and Fox and Friends as well as tweeting about people, such as Barack Obama and Ivanka Trump. What does his network of companies and friends look like? Is it as large as it should be for the amount of influence he has over people as president?

Research Question:

How accurately are people able to understand nodes and edges (the connections between Twitter users) within a busy network?

Hypothesis:

Using Gephi as a tool, people will be able to understand and comprehend certain layouts, like the ForceAtlas2 graph layout the easier than others. This is because this format pushes the most connected nodes to the periphery of the graph which aligns the nodes connected to them and around the most connected nodes. This is also plausible as many crime-related TV shows and movies tend to use this similar layout in their opening scenes or main title sequences, causing people to associate networks with this type of layout.

Null Hypothesis:

There will be no change in accuracy of reading the data between Gephi layouts.

Independent Variable(s):

The independent variable will be the data representation. By using different layouts, the user is forced to comprehend what is being shown, such as determining the most and least important nodes.

Dependent Variable(s):

The dependent variable will be the accuracy of determining who Trump tweets at the most.

Controlled Variable(s):

The controlled variables include the Trump tweet data itself as well as the colors chosen. The node size will stay consistent as well as the label size, font, edge with, network diameter and color. The degree ranking will stay at a consistent minimum of 10 and maximum of 100. The order of representation (layout) will remain consistent as will the participants.

Description of the Stimuli:

The participant will see 3 different layouts: ForceAtlas2, Fruchterman-Reingold, and OpenOrd [Screenshots will be included in Part 3]. These three were chosen as they emphasize different aspects of the data. The ForceAtlas2 and the Fructerman-Reingold layouts both emphasize complementarities, while the OpenOrd layout emphasizes divisions. I have chosen to compare two similar layouts with the same emphasis to observe if there will be a large gap in understanding the two most similar layouts. The ForceAtlas2 layout is made to spatialize small-world or scale-free networks. It’s focused on quality to allow a rigorous interpretation of the graph with the fewest biases possible. Meanwhile, the Fruchterman-Reingold layout stimulates the graph as a system of mass particles. The nodes are the mass particles and the edges are springs between the particles. This layout attempts to minimize the energy of this physical system. Finally, the OpenOrd layout expects undirected weighted graphs and aims to better distinguish clusters, (Gephi, 2013).

Experimental Procedure:

1. Pre-procedure, the three visualizations will be set up with all of the controls set in place.
2. The participant will be asked to sit at the laptop screen and ensure they are comfortable.
3. The researcher will stand near the laptop to monitor the experiment and take notes on the participant’s comments. The timer will start once the first layout begins to run.
4. During all trials, the participants will be allowed to zoom in and out of the visualization as well as scroll around the visualization as it doesn’t always fit the entire screen. They will not be able to touch any other buttons on the sides of each visualization.
5. The first layout that will be shown is the ForceAtlas2 layout. The participant will be asked to determine the most common users that Trump mentions in his tweets. Both the comments and time will be collected by the researcher on paper.
6. The participant will be asked and directed to switch to the second layout, which will be the Fruchterman-Reingold. The same question will be asked as written in step 4.
7. The third and final layout that will be presented is the OpenOrd layout. This decision will be the last in order to determine if the two similarly emphasized layouts have an impact on the participant. The data and time will be collected by the researcher.
8. At this point, the participant will have time to ask any questions or concerns about this mini research. They will be thanked for their time and given compensation with chocolate.

Planned Analysis:

The dependent variables will be analyzed based on their accuracy and time it takes to determine their answers. While accuracy is important, I am more interested in how quickly a user is able to understand the different layouts of the nodes and edges, which will represent how important a layout is to a user comprehending the network of Twitter users.

**Part 2: Building the Apparatus**

Screenshots of the three different layouts produced by users 1, 3 and 5 using Gephi:

Image 1: ForceAtlas2 Layout Demonstration by 19-Year-Old Female Subject

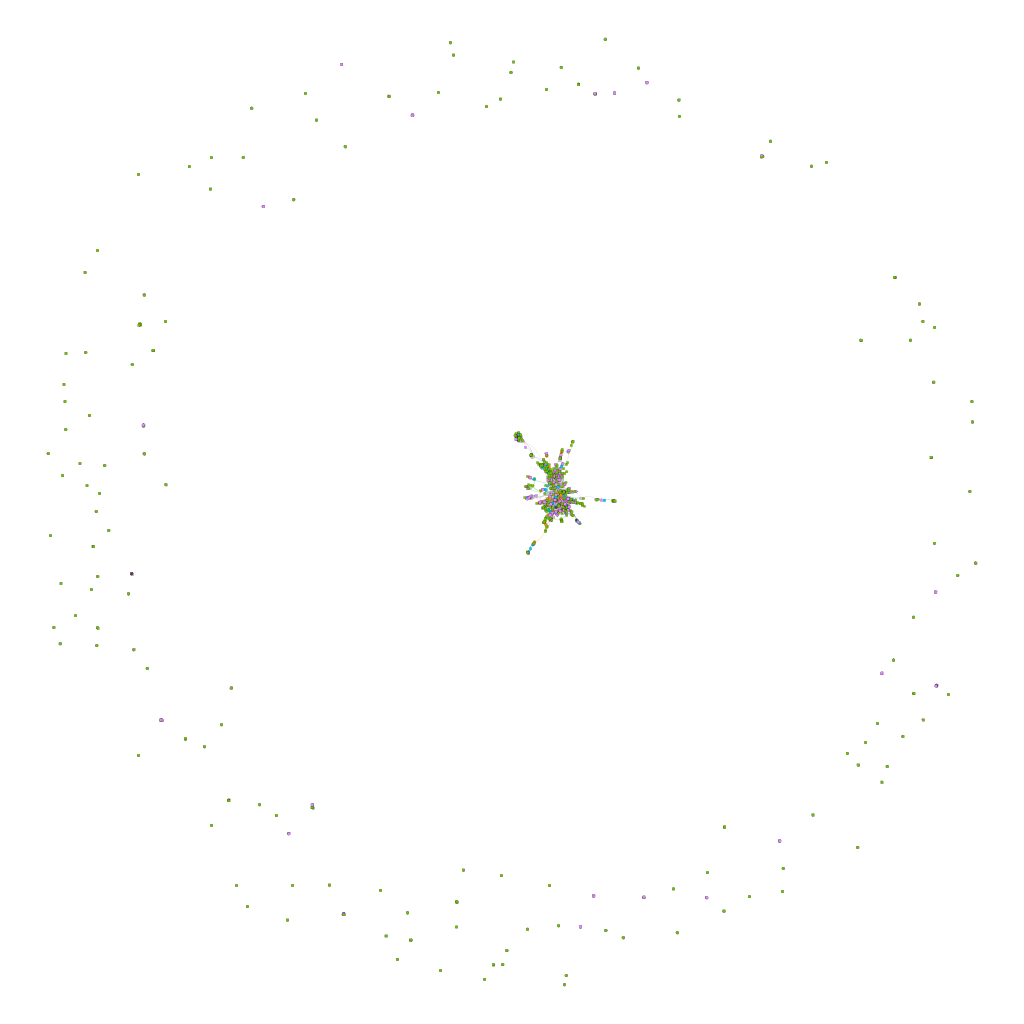


Image 2: Fruchterman-Reingold Layout demonstration by 20-Year-Old Female Subject

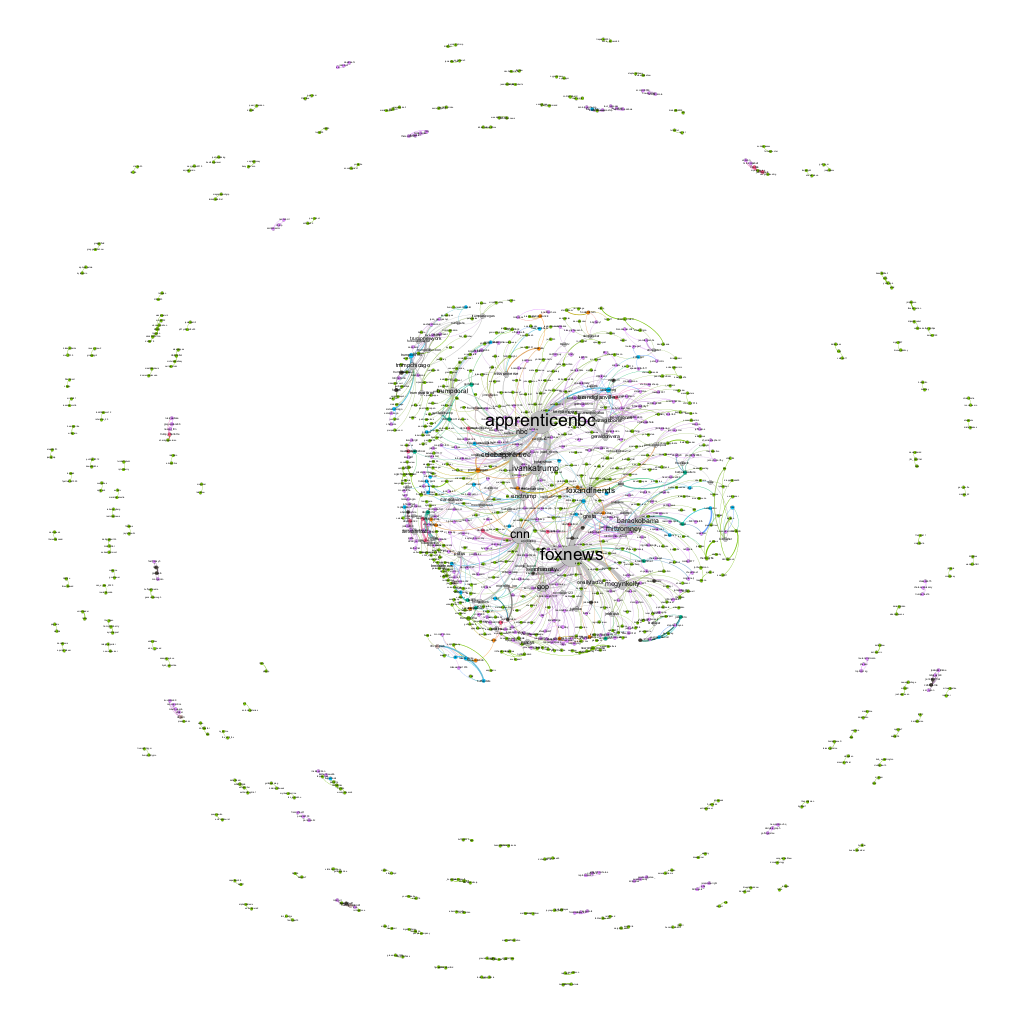
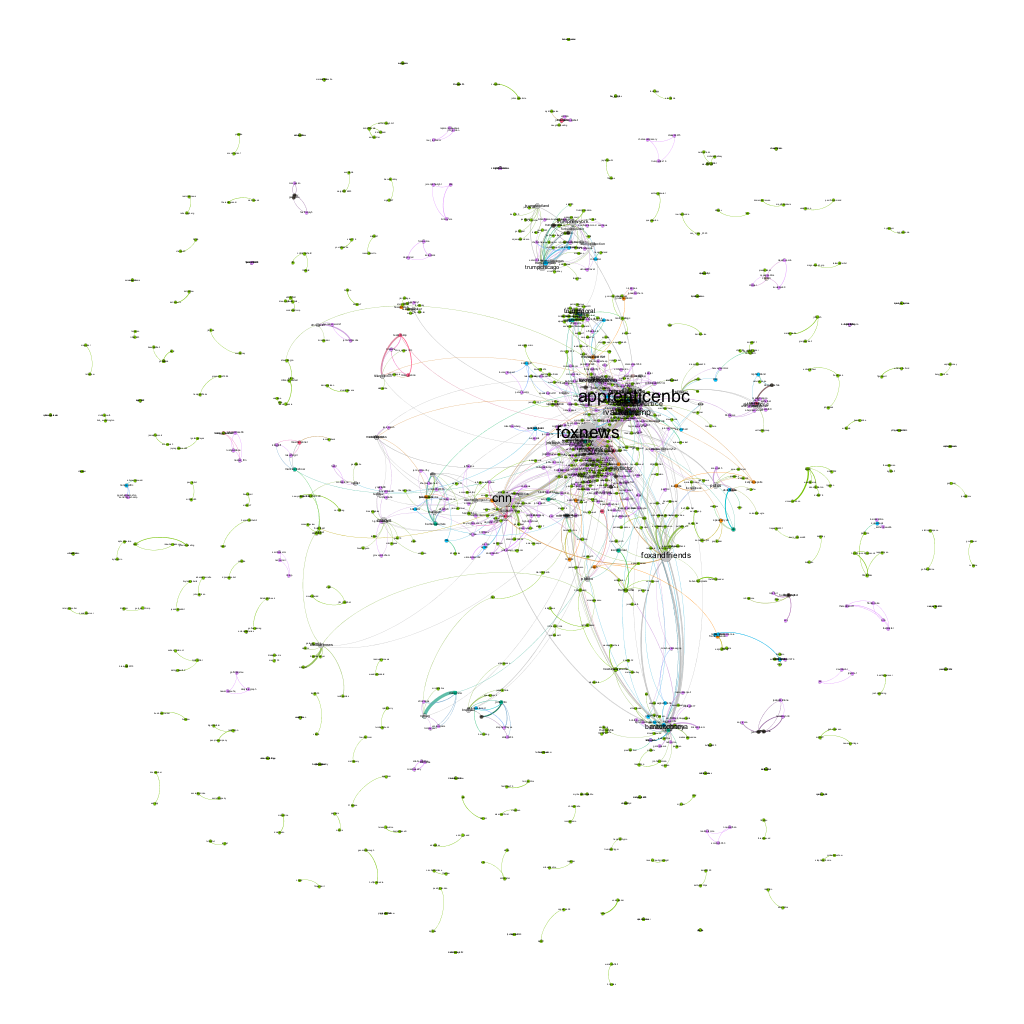


Image 3: OpenOrd Layout demonstration by 22-Year-Old Male Subject



**Part 3: Conducting the Study**

Use your apparatus to collect data from at least 5 different people.

Table 1: Qualitative Data and Timing of Reading and Analyzing the Gephi Layout for 19-Year-Old Female Subject for Different Gephi Layouts Including ForceAtlas, Fruchterman-Reingold, and OpenOrd (± 1 second)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Layout** | **Time 1** | **Time 2** | **∆ inTime** | **Observations** |
| ForceAtlas2 | 0:00 | 1:30 | 1:30 | Foxandfriends, melaniatrump |
| Fruchterman-Reingold | 1:50 | 2:59 | 1:09 | Apprenticenbc, foxnews, cnn |
| OpenOrd | 3:15 | 4:45 | 1:30 | Apprenticenbc, foxnews |

Table 2: Qualitative Data and Timing of Reading and Analyzing the Gephi Layout for 20-Year-Old Female Subject for Different Gephi Layouts Including ForceAtlas, Fruchterman-Reingold, and OpenOrd (± 1 second)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Layout** | **Time 1** | **Time 2** | **∆ inTime** | **Observations** |
| ForceAtlas2 | 0:00 | 1:00 | 2:00 | foxandfriends |
| Fruchterman-Reingold | 2:00 | 2:45 | 1:45 | Foxnews, cnn, apprenticenbc |
| OpenOrd | 3:20 | 6:00 | 1:40 | Apprenticenbc |

Table 3: Qualitative Data and Timing of Reading and Analyzing the Gephi Layout for 20-Year-Old Female Subject for Different Gephi Layouts Including ForceAtlas, Fruchterman-Reingold, and OpenOrd (± 1 second)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Layout** | **Time 1** | **Time 2** | **∆ inTime** | **Observations** |
| ForceAtlas2 | 0:00 | 1:50 | 1:50 | Foxandfriends, melaniatrmp |
| Fruchterman-Reingold | 2:10 | 3:30 | 1:20 | Foxnews, apprenticenbc |
| OpenOrd | 3:50 | 6:15 | 3:05 | foxnews |

Table 4: Qualitative Data and Timing of Reading and Analyzing the Gephi Layout for 24-Year-Old Male Subject for Different Gephi Layouts Including ForceAtlas, Fruchterman-Reingold, and OpenOrd (± 1 second)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Layout** | **Time 1** | **Time 2** | **∆ inTime** | **Observations** |
| ForceAtlas2 | 0:00 | 1:40 | 1:40 | Foxandfriends |
| Fruchterman-Reingold | 2:00 | 3:15 | 1:15 | Foxnews, apprenticenbc, cnn |
| OpenOrd | 4:00 | 5:30 | 1:30 | Apprenticenbc, foxnews |

Table 5: Qualitative Data and Timing of Reading and Analyzing the Gephi Layout for 22-Year-Old Male Subject for Different Gephi Layouts Including ForceAtlas, Fruchterman-Reingold, and OpenOrd (± 1 second)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Layout** | **Time 1** | **Time 2** | **∆ inTime** | **Observations** |
| ForceAtlas2 | 0:00 | 1:50 | 1:50 | Foxandfriends |
| Fruchterman-Reingold | 2:00 | 3:35 | 1:35 | Apprenticenbc, foxnews, cnn |
| OpenOrd | 4:00 | 6:25 | 2:25 | Apprenticenbc, foxnews |

**Part 4: Inferential Analysis**

This experiment concluded that the hypothesis was incorrect, while the null hypothesis was rejected. It was discovered that the layout that had the most accuracy and quickest time was the Fruchterman-Reingold layout. The average time users spent determining who Trump mentioned the most was 1:25 seconds. Meanwhile, the averages of the ForceAtlas2 layout was 1:54 while the OpenOrd layout was 2:00.

Which stimulates the graph as a system of mass particles and emphasizes complimentataries was found to be much more efficient due to its more spread out layout. This also allowed the text to be read more easily and allowed the participants to gain an understanding of networks the quickest. All users mentioned that the color made the data more aesthetic and approachable however did not do nearly as much as the labels of each node. The graph below demonstrates the average time that it took each participant to determine the most mentioned users via each layout.

Limitations to this study include the minimal research, due to lack of participants as well as the slight learning curve that Gephi has to maneuver its many buttons. There is a slight skew in demographics as the majority of participants were female and 20 years old or younger while the males were both older.

# Works Cited

Gephi. (2013, June 11). *Gephi-Tutorial-Layouts.* Retrieved April 27, 2018, from gephi.org: https://gephi.org/tutorials/gephi-tutorial-layouts.pdf