Social Network Lab 1a: Descriptive Network Analysis – Collecting and Visualizing Data

Student: Yijian Li Net ID: YLS9426

Part I Data collection questions:

1. **Provide a high-level overview of the hashtags/urls you included in the data collection. Why did you choose this collection of hashtags/urls? Was there a specific, overarching question - intellectual or extracurricular curiosity - that motivated this collection of hashtags/urls?**

Answer:

Guns are restricted in most countries around the world, but in the United States, gun culture has a long history, and it has always represented the attitudes, feelings, values and behaviors of a society or certain social groups. Therefore, the gun control has also been a huge problem in United States now, and everyone has a very unique perspective on guns.

I love gun models since I was in the primary schools, and I always want to experience the feeling of shooting on the range myself.

As there is a recent news that San Jose passes **first U.S. law requiring gun owners to get liability insurance and pay annual fee**, so I am pretty curious in analyzing how people will react to this law and what do they think about the influence of this new law.

As this law might lead other cities in US to start plan to set same kind of restriction to gun users, I think this is an important topic for us to dig into.

I choose to analyze 4 posts related to this topic to see the connections in this specific hashtag.

2. **What are the insights you hope to glean by looking at the network of hashtags/urls - in terms of individual node metrics, sub-grouping of nodes, overall global network properties?**

Answer:

- What the graph structure would be with this hashtag and how these posts connect to each other?
- Will the node metrics different in these 4 posts?
- Will there be several sub-groupings of nodes or just a huge group of nodes collecting all the comments?
- How many strong trust ties in the global network?

3. **Is the graph directed or undirected?**

Answer:

Yes, it is a directed graph.

4. **How many nodes and links does your network have?**

Answer:

It has 280 nodes and 571 links/edges in this network.

5. What is the number of possible links in your network?

Answer:

It has possible links: n x (n - 1) = 78120

6. What is the density of your network?

Answer:

It has a density of 0.007309268

7. Briefly describe how your choice of dataset may influence your findings. What differences would you expect if you use different hashtags/urls?

Answer:

There is a lot of posts with this topic since it is really hot in reddit, and I choose 4 of them with appropriate size of comments to analyze. Since there are top 2 posts that have 2000+ comments will make other link lose their importance in graph chart, so I did not choose to use them. In addition, I choose these posts from different communities in reddit which will have more rational results and I could know the data I collected is from different part of people instead of just some people in the same community.

With different urls, I do think there will be a huge difference in the results. As I mentioned above, the reaction in different communities in reddit varies a lot and also represents disparate social groups, especially the graph structure and sub-groups they will form. Also, if we use different topics/hashtags, the number of nodes, links will all change and make a new network.

PART II: Network Visualization

1. Create a visualization of the whole network and include it in your report (the first visualization). In a paragraph, describe the macro-level structure of your graph based on the visualization. Is it a giant, connected component, are there distinct sub-components, or are there isolated components? Can you recognize common features of the subcomponents? Does this visualization give you any insight into the interaction patterns of your topic? If yes, what? If not, why?

Answer:

- **Macro-level Structure:**

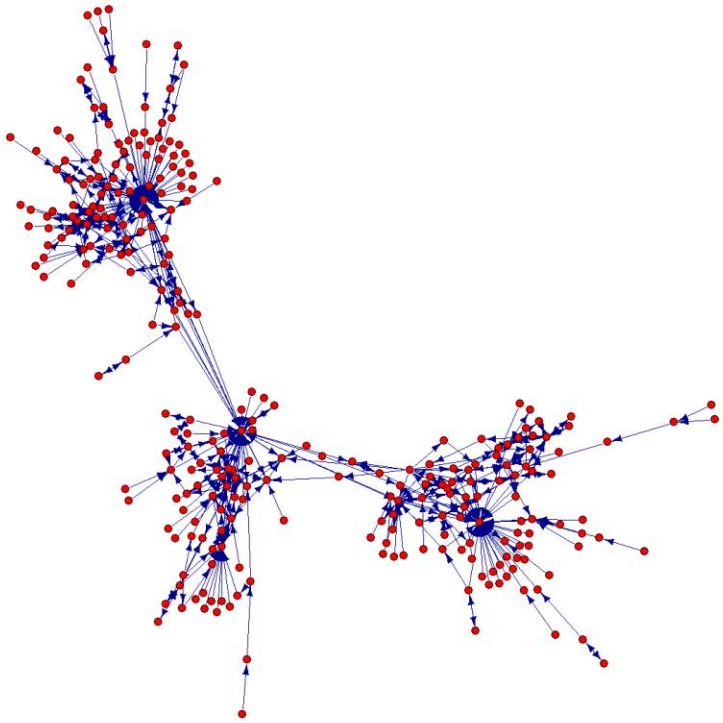
In graph 1 below, we can see that there are three distinct sub-components, and they all have links with each other. They different in size of nodes and links with distinct shapes. In addition, we cannot find any isolated components in the whole graph, so I assume there is no isolated component in graph 1.

- **Common Features of Subcomponents:**

One of the features is that in all three subcomponents, they all have a central node which have most of the links with their nodes. Also, the central nodes all have connections with other centers, in an indirect way.

- **Insights:**

We can see the connections from outside nodes normally point to the centrals nodes through other nodes that are closer to central nodes, which means these 2 kinds of node have the same attitude or they both related to the content of central node with their own comments. Moreover, the central nodes receive/link with so many nodes means either their comments represent most of the attitudes in this post or their comments are extremely controversial. That means they have more value to study.



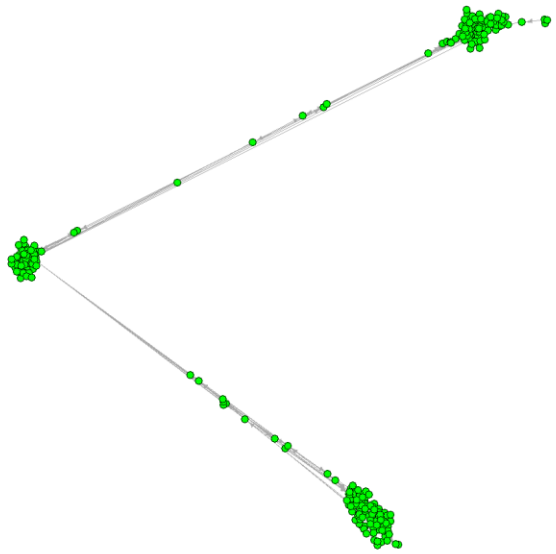
Graph 1

2. **Create a second visualization, now using only the single largest component of the network (i.e., "giantGraph" if you work with the provided R code) and include it in your report. Are there any differences between the first visualization and second one? If so, why? If not, why not?**

Answer:

From processing of graph 2, we get 280 nodes and 571 edges, which is the same as graph 1.

Since we do not have any isolated subcomponents, the nodes in graph could find all the nodes through their connections/links, thus we have the same result.



Graph 2 (Giant)

3. **Create a third visualization using a different 'igraph' layout option and include it in your report. Experiment with visualization options to make your layout better or add additional information to the plot. Explain your choice of layout options. In a few sentences, describe what types of observations are easier to make using one plot or the other.**

Answer:

To optimize the graph we draw, I made several changes below:

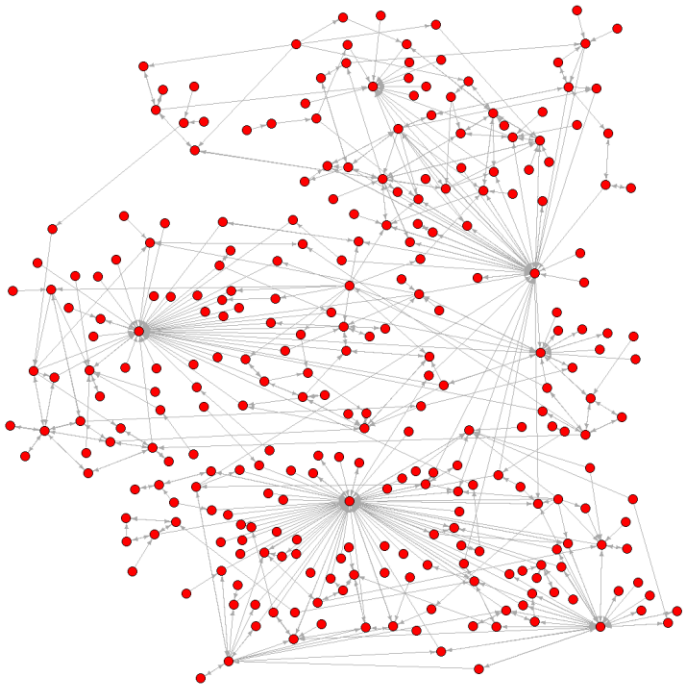
- Set arrow **size** to 0.3 to make arrow smaller in graph
- Set arrow **color** to 'darkgray' which make arrow less prominent
- Use **Davidson and Harel algorithm**, we could have shorter edges, vertices away from each other and minimize edge crossings.
- Set node **color** to 'brown1' that make it stand out in the graph. (Previous: red)
- Set node **size** to 3, which is the appropriate size to analyze graph. (Previous: 2)

So, we have graph 3 below, which is much better than previous.
In the graph3, firstly we notice the whole setting or layout is valid and clear which means we could easily see connections between every node.

Also, we can find with those nodes have high in-degree values, they are kind of isolated and leave with a web in the center. Compared with graph in former questions, we find instead of 3 different clusters, this new graph looks better as it becomes a compact network.

Moreover, the color comparison in the graph 3 make node more distinct and also the arrow becomes less staring.

In addition, there is a small improvement here, we see the edge length in average has become less and we can have a better observation of this graph.
In sum, applying **Davidson and Harel algorithm** could really improve the experience in observation and identifying those key contents in the graph with a little changes in color and size of our components in network.



Graph 3