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

CompSci 396-0: Social Networking Analysis ***Win 2022***

Student Name: Jiaqi Guo NetID: JGR9647

● Responses to Question

■ Part I: Network Data Collection from Social Media

1. **(5 points) 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?**

   Gun control have long been a contentious issue in American politics. Gun rights supporters and gun control supporters (not opponents of gun rights) are **deeply divided** on the issue, often arguing over issues such as gun control, the cause and prevention of crime, and public safety. Generally speaking, these groups with different views correspond to different **social classes and regions** in the United States.

   My Topic: San Jose, California — This city voted Tuesday night to **require gun owners to carry liability insurance and pay an annual fee**, in what's believed to be the first measure of its kind in the United States.

   I am personally very interested in gun culture in the United States, and the proposal of this controversial bill is unprecedented in the history of the United States, so this issue is worth studying.

2. **(2 points) 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?**

   1. What role do nodes with high centrality play in the network, and what contribution do they make to the connection between different sub-groups?
   2. Is there an isolated sub-group? If not, how are all subgroups connected (randomly interconnected or centered on a particular subgroup)?
   3. How are connections made between different subgroups in the network (by a specific **bridge node** or by random connections between **individual nodes**)?

3. **(2 points) Is the graph directed or undirected?**

   It is directed

4. **(2 points) How many nodes and links does your network have?**

   My network has 290 nodes and 482 edges(links)

5. **(2 points) What is the number of possible links in your network?**

   Possible link = $n(n-1) = 83810$

6. **(2 points) What is the density of your network?**

   Density = 0.005751104

7. **(5 points) Briefly describe how your choice of dataset may influence your findings. What differences would you expect if you use different hashtags/urls?**

   Generally speaking, the network drawn based on highly controversial hashtag/ URL will have multiple nodes with high centrality, which usually represent the viewpoint of one party independently and play a central role in information transmission. Also, since 4 urls with similar content are chosen in my experiment, there are multiple sub-components inside the constructed network.

   Networks drawn around less controversial or niche topics tend to be very centralized and often single. And such networks are more likely to have isolated node groups

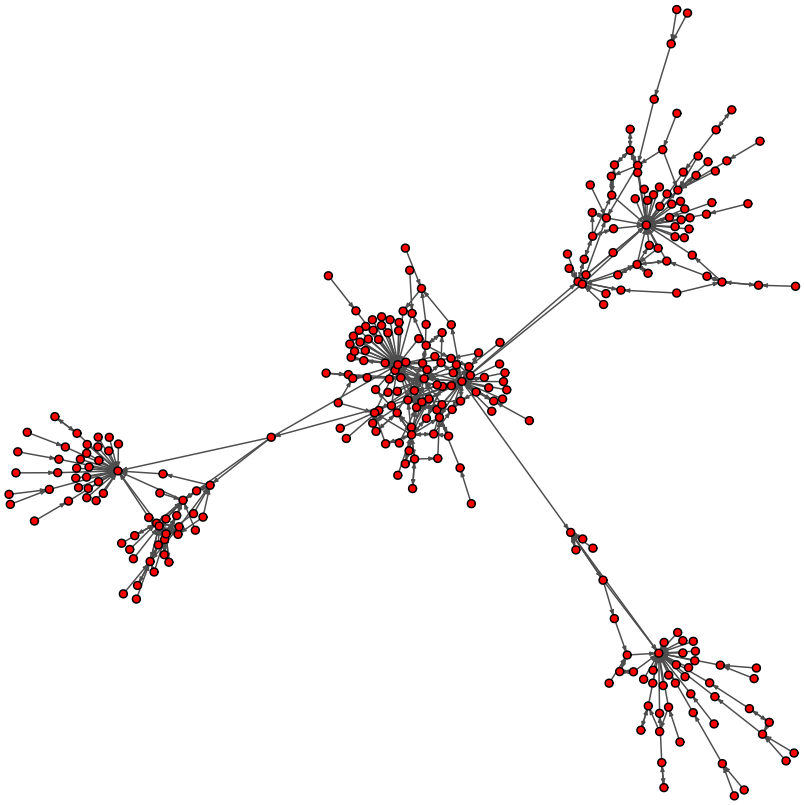
■ Part II: Network Visualization

1. **(5 points) 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?**

   Macro-Level: This "Graph", as shown in the figure below, *consists of four distinct sub-components*, in which the central component is connected to the other three in pairs by certain bridge nodes. Besides, there are no completely isolated nodes in this network which means there are no isolated components.

   Common Features: Each of these components has a strong centrality because we can easily find at least one central node inside the node group. And these centers are usually closely connected or connected by an adjacent node.

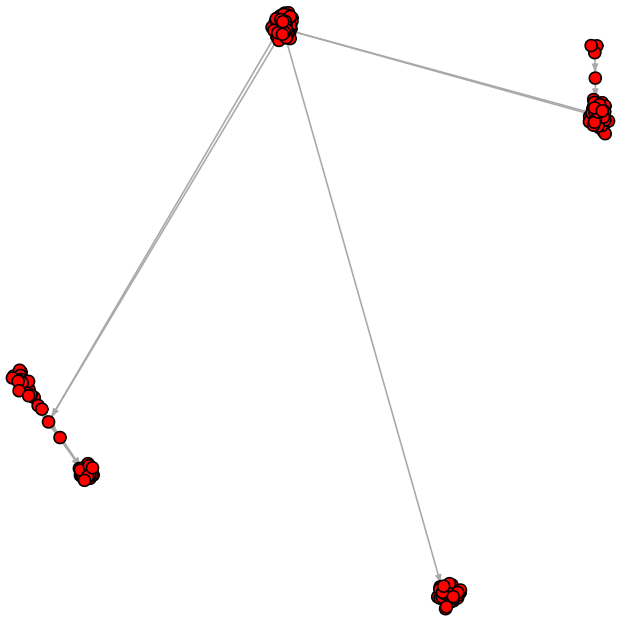
   Interaction Pattern: Yes, this topic originates from the central sub-component of the network, which contains several nodes with obvious high centrality. From these three central nodes, they are connected with the central nodes of the other three sub-components directly (or through a short path) and respectively for message interaction.



2. **(5 points) 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?**

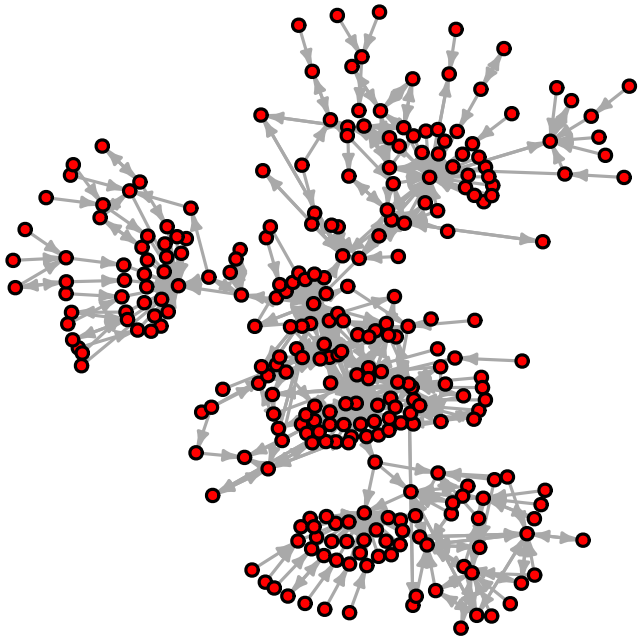
   No, this "GiantGraph" also contains 290 nodes and 482 edges (links). This is because there are no *isolated nodes* inside this network. Below is the graph plotted using (**layout-drl**)

   The differences come from the distance between each pair of nodes and the displaying angle of the network. This is probably because this sort of layout increases the impact of intercorrelation on edge length.



3. **(5 points) 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**

This Graph is created using the **"Spring(layout-kk)"** graph layout, the size of the nodes and arrow are properly adjusted.



Through this graphic layout, we can not only easily observe the macro structure of the network, but also obtain more micro information compared with **layout-drl.**