Shimon Dasgupta

DS 210

12/15/22

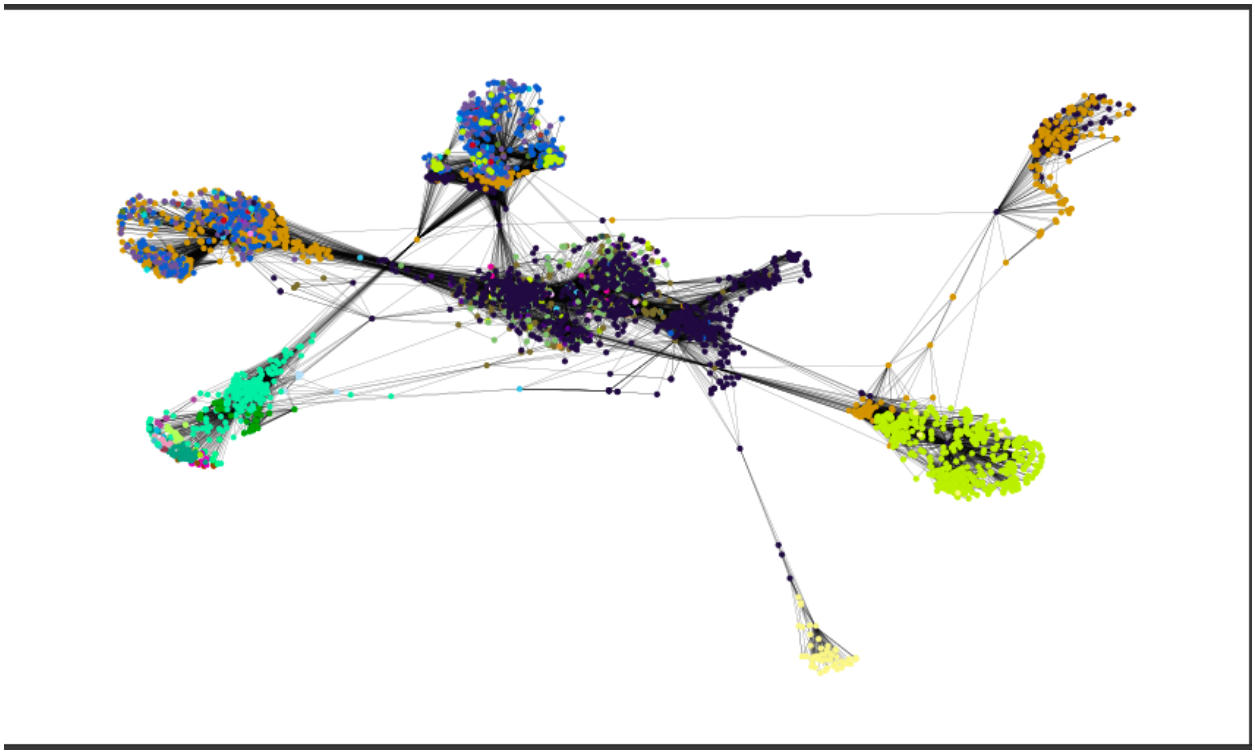
Leonidas Kontothanassis

Analyzing Facebook Ego Network Graph

All of us have used facebook before, whether you are active on it everyday or just have it to remind you of your family's birthdays, facebook has been at the forefront of social media. Because of this, Facebook is responsible for the personal data of millions of people. Their friends, families, interests, career, one look at a person's facebook page and you can gather a lot of personal information. Although facebook has been under recent fire for their data privacy policies, I want to take a look at another issue with facebook, and social media as a whole: the spread of misinformation.

The spread of misinformation or “Fake News” as it has been referred to recently, has been an underlying driving factor in the political and social divide our country is facing. Ever since the 2016 election, the separation from left wing and right wing has never been greater. What makes things even worse, is this divide is taking place in the middle of a data revolution. With companies and advertisers gaining access to more and more personal information about customers, it becomes easy to specifically target certain audiences with content that they would like. While this is usually a good thing for both customers and a company, the sale of a harmless product isn’t always what your data is being used for. Social media apps like facebook also advertise communities and content itself based on your political beliefs and overall views. This makes Facebook and Twitter tools of mob mentality and confirmation bias. Being surrounded and constantly fed “facts” that support your common beliefs are what create a divide and prevent us from coming together and challenging each other's ideas in an intelligent and respectful manner.

To dig deeper into the connectivity and spread of information in social media, I will be analyzing a graph of 10 users and their connections provided by facebook. I will be looking for specific statistics that can help gauge how easy it is for certain groups to spread information and how information can cluster amongst friend groups and communities.



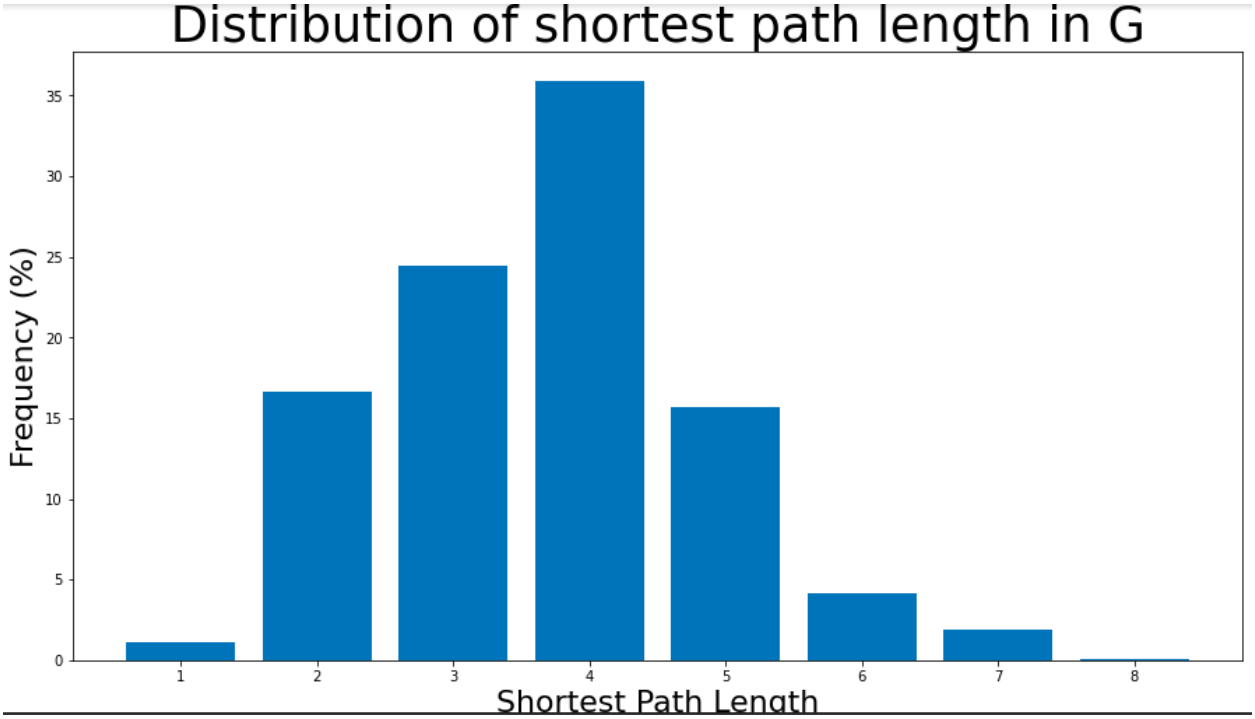
I began by reading the edges of all the nodes into Rust to get an understanding of the scope of the project. Using a longest path algorithm, I was able to find the longest possible path from through all the edges which was 18446744073709551615. This was after I had to half the sample from 88k to around 50k due to memory allocation issues. If i went with an Adjacency list instead of Matrix It may have worked. Obviously 18446744073709551615 is quite a large number but with all the possible connections and combinations it makes sense. But when you analyze more factors such as ego_features and circles you begin to see how connected everyone really is. Because my experience in rust visualization is very limited I constructed some graphs in Python's Networkx's library for graph analysis. With this library I was able to test how often friends of a node are friends of each other as well as to what extent the six - degrees of separation held true. All of this was coupled by multiple measures of centrality that could tell us which users had the most access to the network as well as which users served as bridges to other users. Users that are more central to the network can directly feed information to the most number of users. This means that they can spread misinformation instantaneously compared to someone on the outer edge of our network who has to connect with multiple

nodes.

```
126 | fn read_graph(num_nodes: usize, edges: Vec<(usize, usize)>) -> Vec<Vec<usize>> {
    |    ^^^^^^^^^^
    |
    = note: `#[warn(dead_code)]` on by default

warning: `hw10` (bin "hw10") generated 2 warnings
    Finished dev [unoptimized + debuginfo] target(s) in 0.03s
     Running `target\debug\hw10.exe`
The longest path in the graph is 18446744073709551615
```

Regardless of individual users' connectivity. Overall I found the graph to be quite clustered with a clustering coefficient of 0.6 and 95% of the network following the 6 degrees of separation rule. The maximum number of nodes any person had to travel to reach another was 8 and that accounted for less than 1% of the network.



Overall, this project demonstrated how connected people really are to each other in social media. The time it would take for any post or message to span the entire network wouldn't be long at all. 60% of every node's friends are friends with each other which created over 16000 triangles. This means that regardless of position in the network, your messages and posts have the ability to easily reach and influence everyone. With this study of only 10 users, I can't imagine how connected and powerful the entire facebook network would be. Unfortunately I didn't have enough time to test run all my tests on other features of the social network but I hope you can see in my code I was close but wasn't able to debug in time. In the future I hope

to finish this project in its entirety and get a better understanding of just how connected out world truly is.

Work Cited/References

Allcott , Hunt. “Trends in the Diffusion of Misinformation on Social Media - Sage Journals.” Trends in the Diffusion of Misinformation on Social Media, https://journals.sagepub.com/doi/full/10.1177/2053168019848554.

“Rust API Guidelines.” Documentation - Rust API Guidelines, https://rust-lang.github.io/api-guidelines/documentation.html.

“The Rust Programming Language.” The Rust Programming Language - The Rust Programming Language, https://doc.rust-lang.org/book/.

Kirenz, Jan. “Social Network Analysis with Python.” Jan Kirenz, 13 May 2020, https://www.kirenz.com/post/2019-08-13-network_analysis/.