

NETWORK SCIENCE

# Assignment 8

STUDENT | Erfan Raoofian

ID | 47315221

INSTRUCTOR | Prof. Ifeoma Adaji



IRVING K. BARBER SCHOOL OF SCIENCE  
DEPARTMENT OF COMPUTER SCIENCE  
OKANAGAN CAMPUS

## 1 Question 1

Right after the vaccination against SARS-CoV-2 started in the world, in about a week, some cases of acute thrombosis emerged in the news. The fact that this virus had a huge impact on everyone's life and also the severity of the reported side effect of the vaccines led to a huge amount of discussions between people on social media. The main aim of this study is the analysis of Twitter discussions around this topic. They used the keyword 'blood clots' to gather related twitter data from 14 March to 14 April 2021. Furthermore, this paper tries to find the most important users, key concepts, different groups, and most popular tweets in this regard. The authors were able to create a social network graph using Twitter data to find out the positive and negative perceptions about SARS-CoV-2 vaccination, and more specifically, the possibility of forming blood clots after the vaccination. This network graph was produced by using the Twitter account as nodes of the graph, and the relationship between Twitter accounts as edges. The relationship between Twitter accounts consists of tweets, retweets, and mentions. A classification algorithm named Clauset–Newman–Moore was used to identifying different groups within the network and the key communities in the discussion. The study only focuses on the English language as it was the most commonly used language in this topic. The paper reports the most retweeted messages and the most influential users in the network and ranks the users according to their influence based on the betweenness centrality measure in the network. They also report the key hashtags, and the keyword pairs used in the tweets. At last, they claim that the network was a broadcast network as it had a high retweet ratio. In conclusion, they say that Twitter was used to highlight the low potential of developing a blood clot from vaccines and was an encouraging source for vaccinations among the public.

## 2 Question 2

They used the keyword 'blood clots' to gather related twitter data from 14 March to 14 April 2021. A dataset consisting of 266,677 tweets was retrieved at first, and a random sample of 5% of the dataset ( $n = 13,334$ ) was chosen to be entered into the NodeXL toolbox for further analysis on the dataset. The authors gathered data from Twitter using Twitter-search API and Twitter Archiving Google Sheets (TAGS). The network graph was produced by using Twitter accounts as the nodes of the network and the relationships between Twitter accounts as the edges. The relationships between Twitter account consist of tweets, retweets, and mentions to connect the nodes in the network. The authors do not mention if the generated network is directed or not, and they also don't use anything indicating that the network is directed. In short, Twitter was used to highlight the very low potential of forming a blood clot from vaccines compared to some other activities and they found out that the users on Twitter were encouraging vaccinations among the public.

## 3 Question 3

As I mentioned in the previous questions, they used the Clauset–Newman– Moore algorithm for community detection, and the betweenness centrality measure to find the most influential users. The betweenness centrality measure can correctly rank the most influential nodes as it measures the number of times each node has appeared in the shortest path between the nodes. It shows the importance of the node as it is used to propagate the information in the network effectively and optimally. They used the Clauset–Newman– Moore algorithm as it can show which groups of

nodes have the most interaction with each other and are separate from the rest of the network.

## 4 Question 4

There is a figure in the article that shows a high number of different communities. Among all the communities detected, they pick the 4 largest communities to study. They call these four largest communities Group 1, Group 2, Group 3 and Group 4. Among these, Group 1 is the largest group and Group 4 is the smallest. The shapes in Groups 1, 2 and 4 show a number of important users whose tweets were amplified by other users in the group and also across the network. They name Group 3, an ‘isolates’ group in which there are no retweets or mentions. The users in this group use the term ‘blood clot’ but did not mention or reply to any other user in the network. According to the authors, Group 4 is a self-contained broadcast group. They are labeled ‘echo chamber’ by the authors because they had little interaction with other users in the network.

## 5 Question 5

Some of the limitations that come to my mind are as follows:

### 5.1 Groupings:

We know that the authors used 5 percent of the tweets they found in their search and this portion was chosen randomly. On the other hand, they group some of the accounts in group 3 as an ‘isolates’ community that did not mention or reply to any other user in the network. I do not think that this interpretation of this group is fair because they may have mentioned or replied to other users outside of that 5 percent portion that the authors are studying. Other than group 3, the other groups identified are reasonable and have no problems.

### 5.2 Other metrics:

I think it would be interesting if the authors also used other metrics such as the eigenvector centrality measure for their analysis. Eigenvector centrality measure can measure the importance of the nodes relative to the quality of their connections. Because the authors of this article are trying to study the positive effect of social media on vaccination and the spread of true information, the quality of the connections can be important for their analysis. They could check the relationship between the eigenvector centrality score of each node and the correctness of their tweets according to scientific sources.