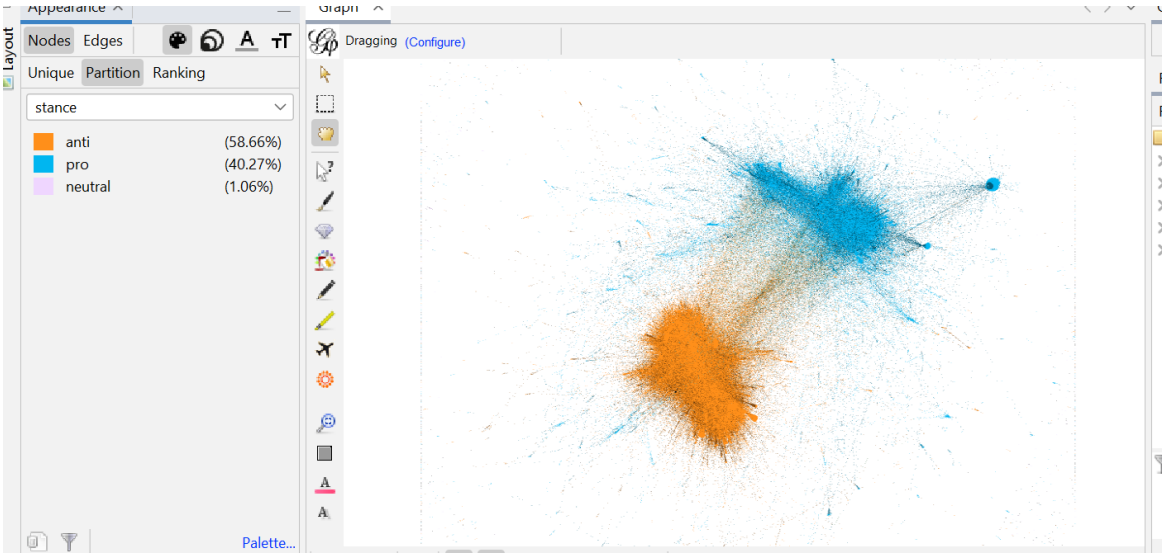
Sprint 8 Report

By
Conor Barry
Tom Higgins
Jane O'Brien

"I want to visualise the results of my dataset analysis"

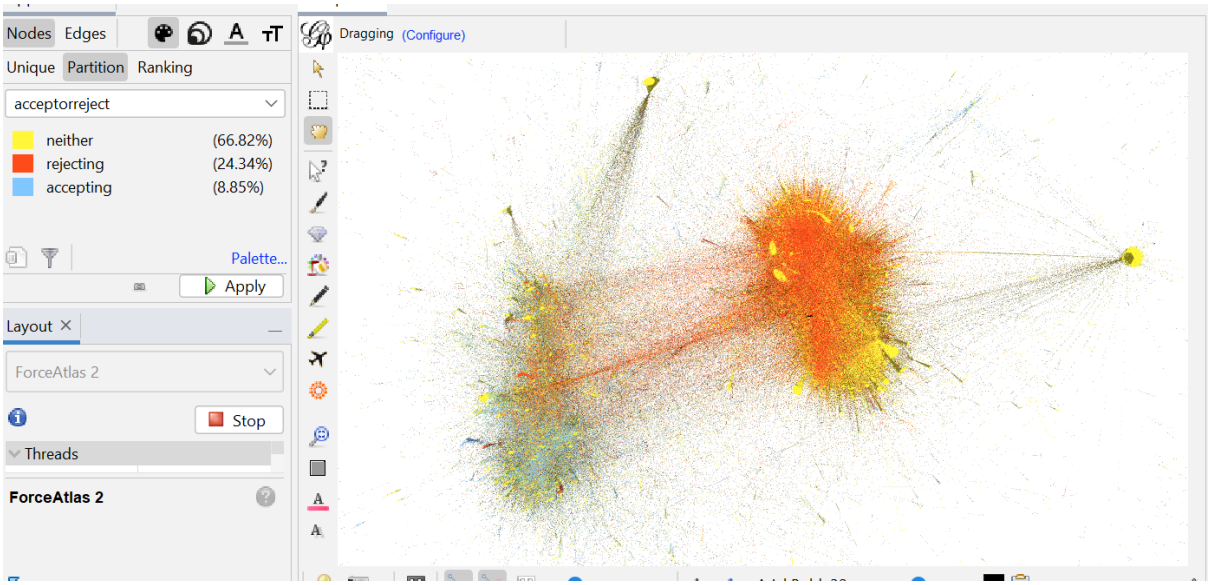
Retweet/Retweeted:

The following graphs contain TwitterUser nodes. There are edges between the users and the users they retweet in the retweet graph, and edges between users who get retweeted by other users in the retweeted graph.



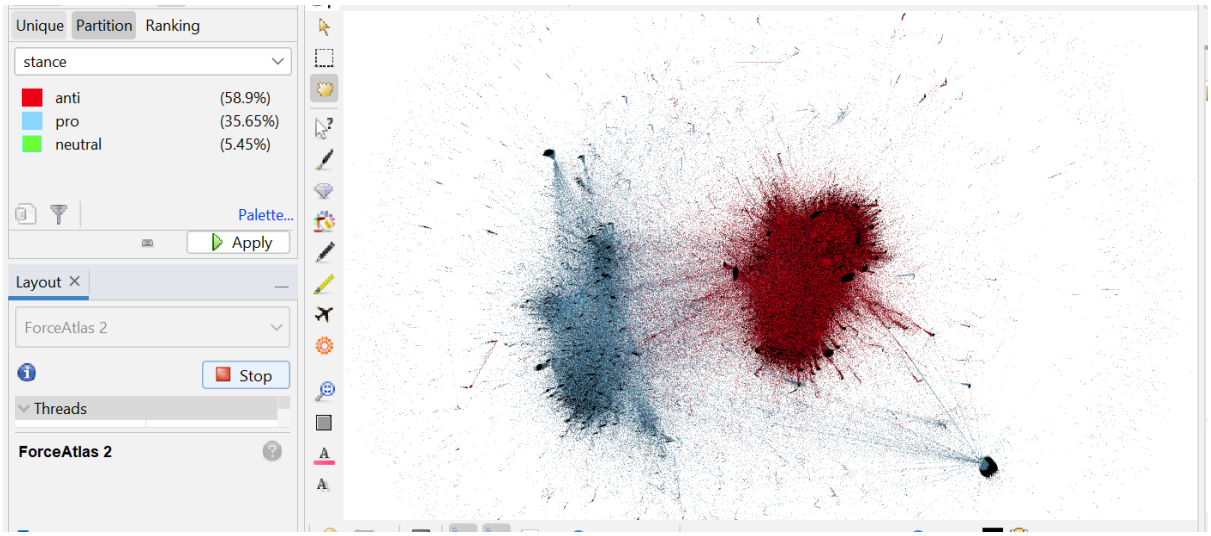
Retweet graph based on stance
Colour Scheme: Pro: Blue, Anti: Orange, Neutral: Purple
Nodes: 373,438, Edges: 892,074

As can be seen from the above picture, the dataset is split into two groups of pro vax users (blue) and anti vax users (orange). It is obvious that pro users retweet other pro users and anti users retweet other anti users. There are some faint lines between the two colours in the graph, as some users retweet users on the opposite side of the debate, either in an attempt to reason with the other side or to mock them. This only occurs a small amount of the time, the vast majority of the users simply retweet users on the same side as themselves.



Retweet graph based on accepting or rejecting quality
Colour Scheme: Neither: Yellow, Accepting: Blue, Rejecting: Orange
Nodes: 373,438, Edges: 892,074

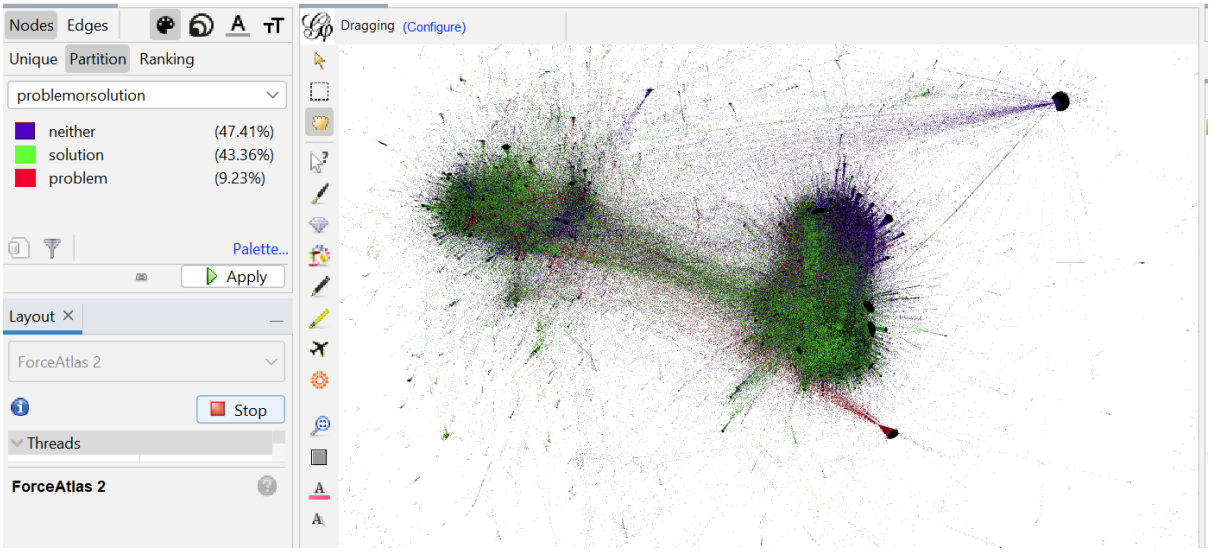
More than half of the dataset users do not use accepting or rejecting hashtags. If we ignore the yellow colour in the graph, we can see that users who use rejecting hashtags (in orange) retweet other users who also use rejecting hashtags, and the users that use accepting hashtags mostly retweet other users who use accepting hashtags. There is some overlap in the middle where certain accepting users retweet other rejecting users and vice versa. Looking at the yellow colour we can see that both accepting and rejecting users retweet users who use hashtags that are neither accepting or rejecting



Retweeted graph based on stance
Colour Scheme: Anti: Red, Pro: Blue, Neutral: Green
Nodes: 373,438, Edges: 892,074

Similar to the retweet graph, there appears to be two separated groups of users who get retweeted by users of the same stance (anti vax stance in red and pro vax stance in blue).

We can conclude from the retweet graph and the retweeted graph that twitter users mainly interact with users of the same opinion as themselves, with very little dialogue between the two sides of the debate.



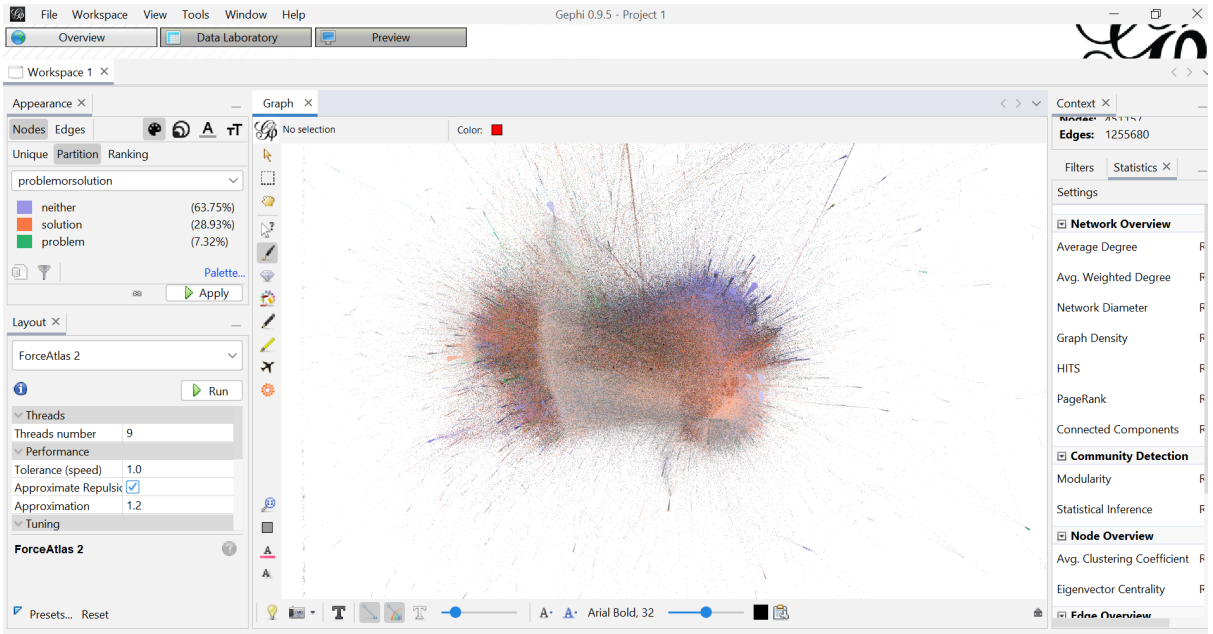
Retweeted graph based on problem or solution
Colour Scheme: Neither: Blue, Solution: Green, Problem: Red
Nodes: 373,438, Edges: 892,074

The percentages and graph show us that most users are talking about solutions rather than problems. Clearly both sides of the vaccine debate are talking about things like vaccines, masks, social distancing but we can assume that anti users are speaking negatively and rejectingly about these solutions and pro users are speaking in the opposite manner.

Mentions/Mentioned:

Mentions graph based on stance
Colour Scheme: Anti: Purple, Pro: Orange, Neutral: Green
Nodes:451157, Edges: 1255680

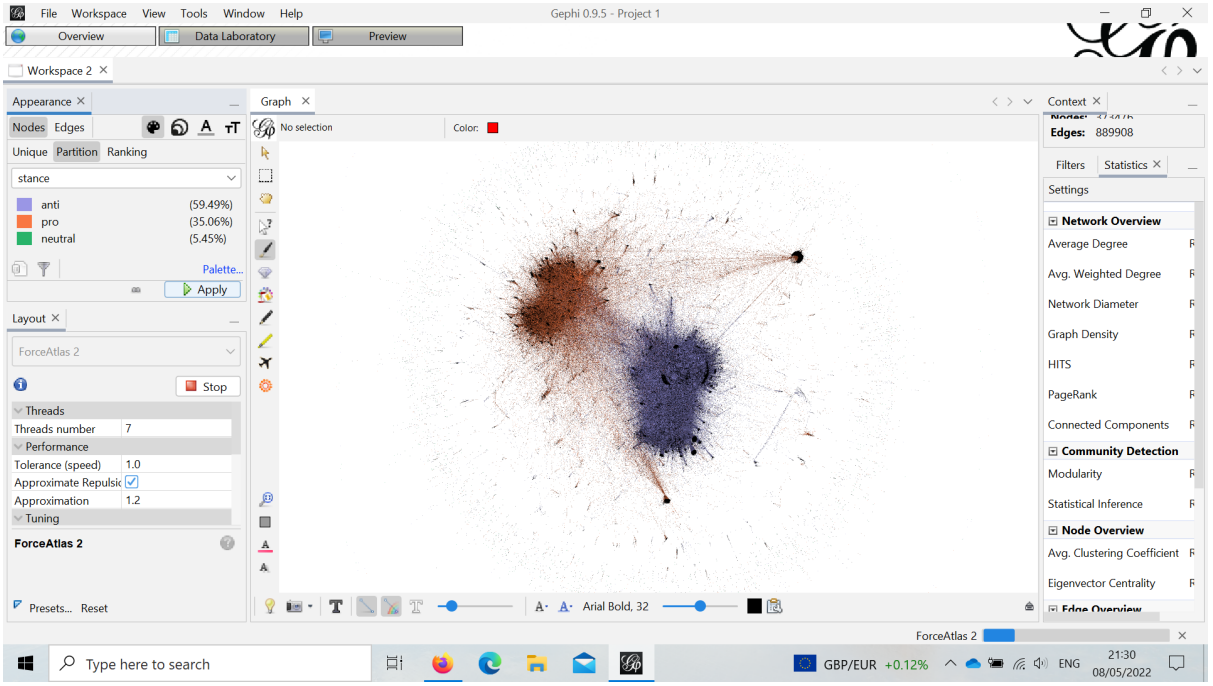
From this graph, we can clearly see that anti users mention each other significantly more than pro users mention each other. We also see a very limited amount of pro/anti users mentioning each other so it seems that the people tweeting are not interested in talking to each other, there seems to be a bit of an echo chamber.



Mentions graph based problem/solution
Colour Scheme: Anti: Purple, Pro: Orange, Neutral: Green
Nodes:451157, Edges: 1255680

We see some interesting results when we analyse the mentions graph and take into account if the users are discussing problems/solutions.

We can clearly see that solution oriented users tend to mention each other the majority of the time but we can also see a limited amount of solution oriented users talking to problem oriented users. For a significant number of users, it wasn't possible to conclude whether they were more focused on problems/solutions but these users most often mention solution oriented users.

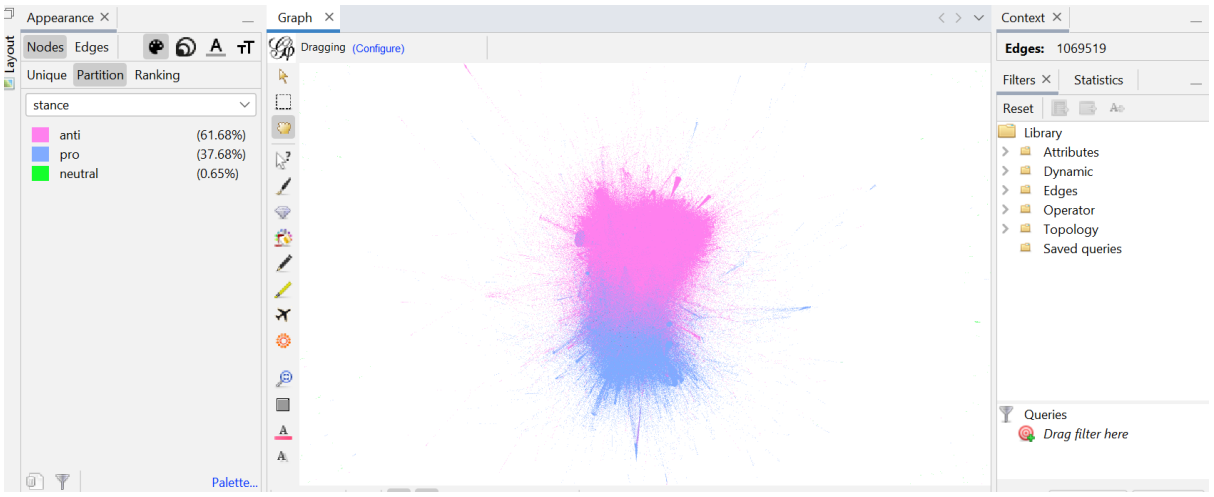


Mentioned graph based on stance
Colour Scheme: Anti: Purple, Pro: Orange, Neutral: Green
Nodes:373476, Edges: 889908

We can clearly see that anti users tend to be mentioned by each other the majority of the time and vice versa for the pro users. We do see a limited amount of interaction between the groups however only to a insignificant extent.

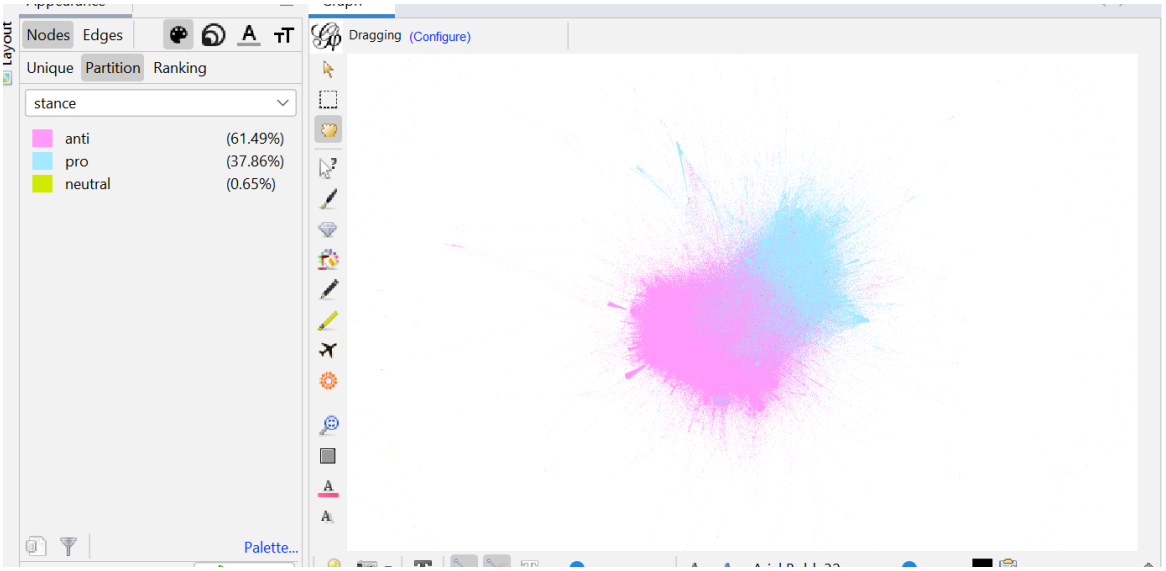
Hashtag to Users/ Users to Hashtags:

The following graphs contain TwitterUser nodes and Hashtag nodes. There are edges between the users and the hashtags they use. Based on Hashtag to Users graph using Stance



Hashtag to Users based on stances

Colour Scheme: Anti: Pink, Pro: Blue, Neutral: Green
Nodes:320,244, Edges: 1,069,519



Users to Hashtag graph based on Stance
Colour Scheme: Anti: Pink, Pro: Blue, Neutral: Green
Nodes:320,244, Edges: 1,069,519

Using the algorithm Force Atlas 2 in Gephi, it is clear that when twitter users and hashtags in the dataset are separated by stance (pro/anti vax (blue/pink)) that the dataset is split into two almost distinct groups, where the anti users use anti stance hashtags and the pro vax users use pro stance hashtags. There is some slight overlap between the two groups as some pro vax users use antivax hashtags mockingly, and vice versa. In other scenarios, a hashtag is used by both sides of the debate and an accurate stance is not possible to obtain.