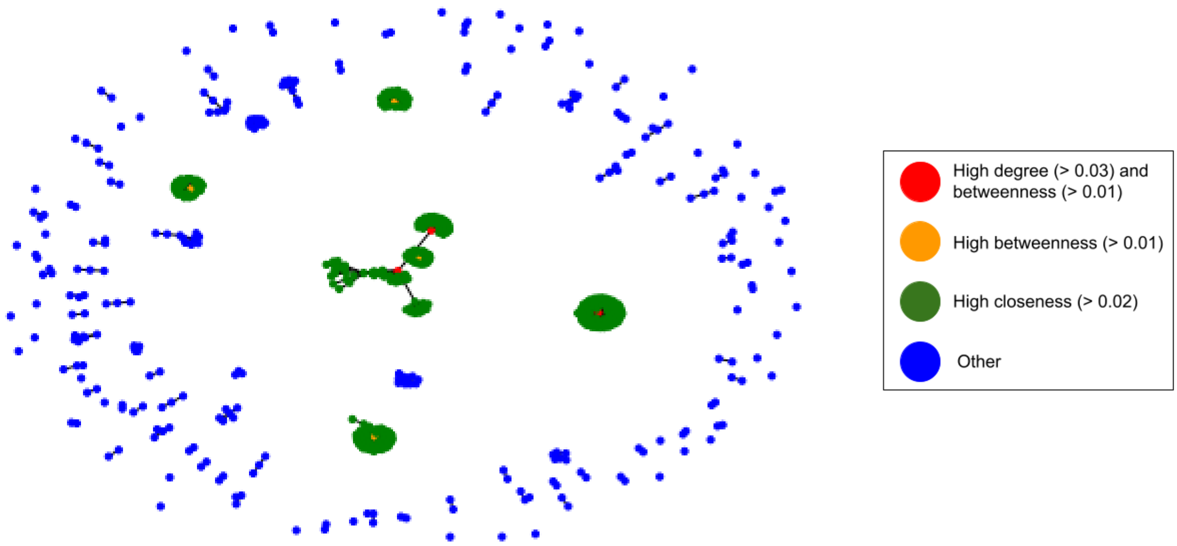
Part 1

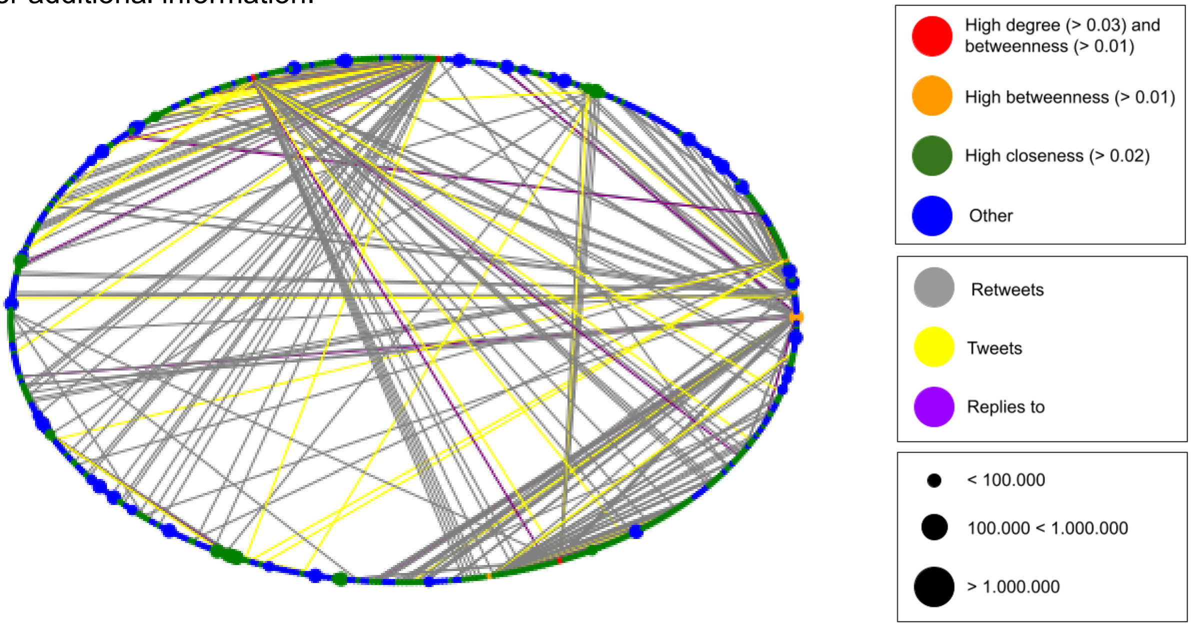
For the first part of the assignment we chose a hashtag and analyzed the corresponding nodes and edges which form a network. We chose to analyze the hashtag #qanon as we were interested in how this network would look. This topic is a sensitive topic as followers of this theory believe in all sorts of extreme conspiracy theories ranging from a secret world order to satanic child abuse. Our expectation for this network was that it would be a polarized crowd network. On the one hand a big group of people that believe in the Qanon theories and tweet about it and on the other hand the people that report on the behavior of the people that follow the theories of Qanon or are opposed to the theories of Qanon.

Contrary to our prediction the network does not look like a polarized crowd network at all. It seems that people that tweet about Qanon consist of closely connected groups of people that are mostly isolated from other groups. There are also some isolated people tweeting about Qanon that do not belong to any group. This description fits closely to Community Clusters.



We calculated the most important people in the network on the basis of degree centrality, as people with high degree centrality are connected with the most people, and betweenness, as people with high betweenness are important nodes that catch a great deal of traffic within the network. People with high degree centrality and betweenness are marked with a red color in our visualization. As you can see, the red dots appear in the middle of the very tight groups of people in all places in the network. People that have high degree centrality but not as high betweenness score are marked with an orange color. As you can see the orange dots appear in tight groups but these are smaller than the groups with the red dots. People with high closeness are marked with a green color in our visualization. As you can see, the green dots appear around all the central red and orange dots for the tight groups in the network. This makes sense as the people that are part of these tight groups are the most connected with one another and thus they will all have high closeness scores. The nodes that are left are marked with a blue color. These nodes do not have a particularly high closeness, betweenness or degree centrality score and are often at the edges of the network with few connections.

The first graph gave us a good overview of the type of network we were dealing with. We also wanted closer inspection of the edges that ran between the nodes. As this is hard to see in a normal graph we chose to draw another graph that included the edges and some other additional information.



The nodes in the second graph follow the same color scheme as the first graph. As you can see the red and orange nodes are mostly surrounded by green nodes which fits with the notion we got from the first graph. As an additional measure we tried to find out if the amount of followers had any connection with a high degree centrality and betweenness scores. We distributed the sizes of the nodes within three categories. Small sized nodes are nodes with users with less than 100.000 followers, the medium sized nodes are users between 100.000 and 1.000.000 followers and the large nodes are users with more than 1.000.000 followers. It turns out that the number of followers does not have a great effect on the degree centrality or closeness score as the largest nodes appear in both nodes with high and low degree centrality, closeness and betweenness scores across the graph. The last we visualized in this second graph are the edges which give us more information about the type of interaction between users. The lines with a gray color represent retweets, the yellow color represents tweets and the purple color represents replies. The majority of the interactions appear to be retweets. The retweets also appear very frequently around the tight groups with red and orange dots in the center of green clusters. This leads us to the conclusion that the tight groups are formed by a user whose tweet got retweeted by many other users. The user whose tweet got retweeted are the red and orange dots and the people that retweeted are the other green dots.

Before close analysis of the tweets, we expected to find many tweets coming from Qanon followers however most of the tweets we found were critical of Qanon or news articles about events that had to do with followers of Qanon. A possible reason for the finding could be that the Qanon followers have left twitter as tweets that contain fake news or conspiracies are removed from twitter. Perhaps if we analyzed another platform such as Parler we could encounter more Qanon followers. Another explanation could be that the actual followers of Qanon were simply not in the dataset of 566 users that we worked with.

Part 2

Linguistic status
To calculate the linguistic status measure we decided to study the proportions of I, we and you words to the total amount of I, we and you words used and also used the proportions of these pronouns to the total amount of words of the user. As I words are associated with lower social status, these words would count for less social gain than the we and you words.

Network centrality
As an additional centrality measure we have chosen to calculate the load scores for the users in the network. This is the fraction of all shortest paths that pass through that user. The load score is a good indicator for important users in the network as it indicates that a great deal of the twitter traffic passes through that user.

High status
User: rayn3ll
Betweenness: 0
Closeness: 0
Degree: 0.003
Load: 0

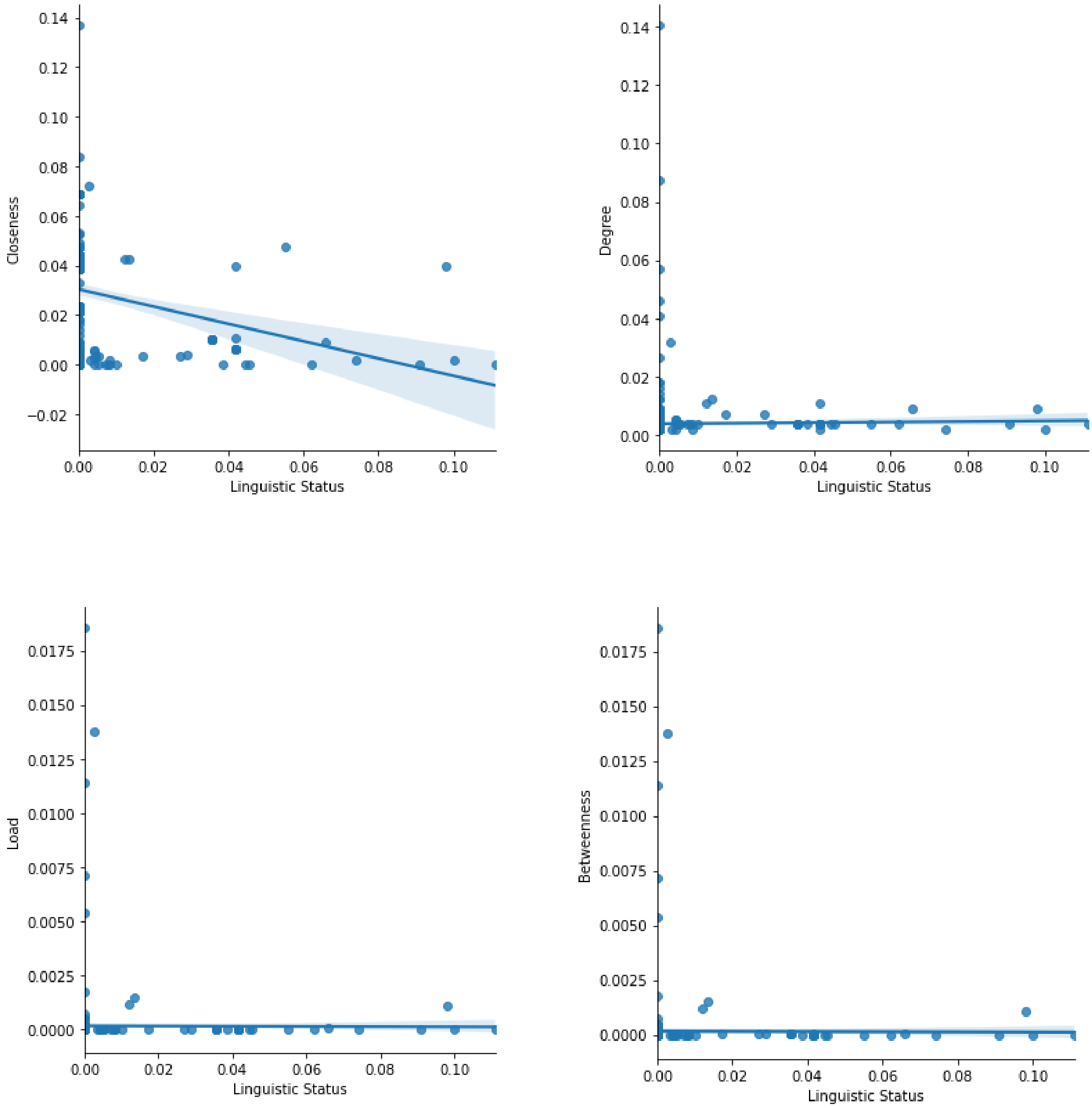
The user *rayn3ll* does not fit the prediction of Pennebaker. We would expect that a user with a high linguistic status would have a high social status as well. This would translate itself to high degree, closeness, load and betweenness scores. This does not appear to be the case as all the scores across the board are remarkably low.

Low status
As we are working with a dataset that does not contain a great deal of tweets for every user it does occur that users did not use I, we and you words at all. This caused all of these users to have a linguistic status score of 0. As we still want to report about these people we took a random sample out of this group:

User: basile061
Betweenness: 0
Closeness: 0.021
Degree: 0.004
Load: 0

The user *basile061* does not fit the prediction of Pennebaker either. We would expect that a user with low linguistic status would have low social status as well. This would translate itself to low degree, closeness, load and betweenness scores. While the scores are not particularly high, these are higher than the person with the highest linguistic status within our network. This should not be the case if the linguistic status would match the social status.

Status vs centrality



The graphs above show the correlation between the four centrality measures we used in our research. As can be observed load, betweenness and degree do not show any correlation with linguistic status. There does seem to be some weak negative correlation between closeness and linguistic status.

One of the major reasons linguistic status did not function as optimal as it could was that the tweets that we worked with were not complete. NodeXL was not able to load the entirety of the tweets but only the first portion. This means that we calculated the linguistic status with incomplete data, thus probably missing a great deal of I, you and we words that could potentially be in the hidden part of the tweets, resulting in an incomplete measure. Another major reason for the suboptimal measure of linguistic status was that the majority of the interactions in the network consisted of retweets. This means that a great deal of users did not write a tweet themselves but copied the same message over and over. When one of the retweeted tweets did not contain any information about linguistic status this would echo for the entirety of the cluster that retweeted that tweet resulting in similar low scores for many uses. Because of the aforementioned reasons we cannot give a good conclusion about whether or not this measure is useful for measuring the social status of a user. Although the scientific reasoning for this measure seems good we doubt if this measure would be useful in analyzing a social media network. Mainly due to the fact that many users do not write their tweets but retweet them. To add this, the measure proposed by Pennebaker was based on written text, however the use of language on social media platforms is very different from regular written text so we are unsure how applicable this measure is in this domain.
Our research would suggest that using the network centrality measures gives us a better representation of the present social structures, therefore we would suggest using the network centrality measures to analyze social structures on social media platforms.