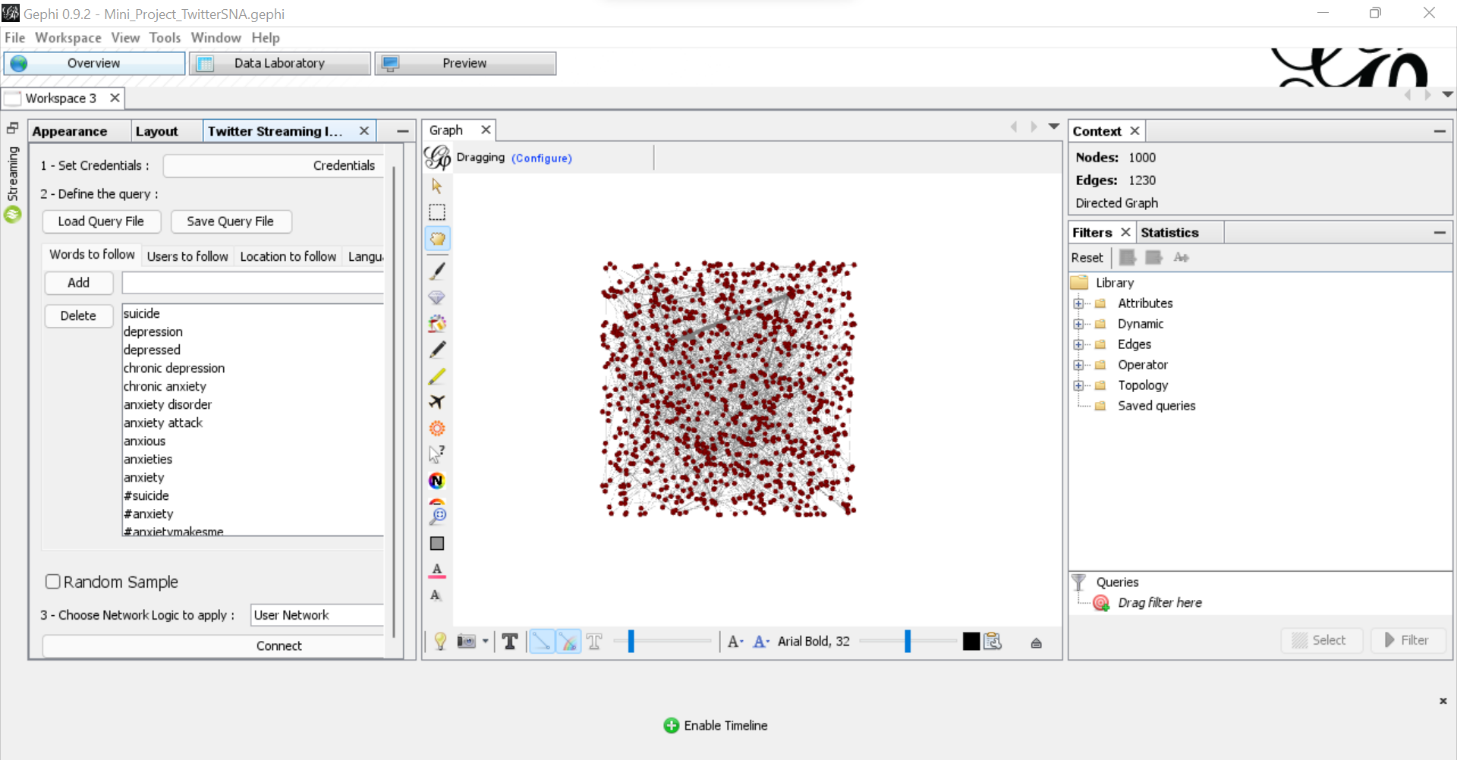
**RESULTS OF SOCIAL NETWORK ANALYSIS OF DEPRESSED USERS ON TWITTER**

###### Global Network Description

This section attempts to describe the network-level measures i.e., establish an overall idea about the data that is being dealt. The dynamic data collection of thousand nodes from Twitter is visualized in Gephi as shown below.



*Loading of network data into Gephi*

**1.1 Summary of the network measures**

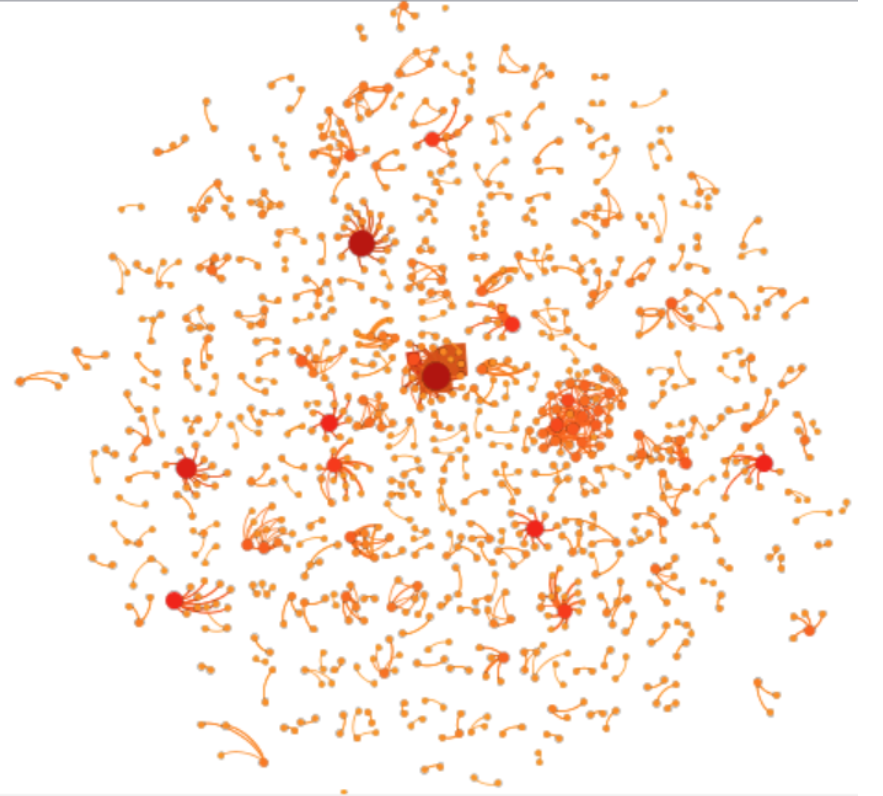
|  |  |
| --- | --- |
| **Network Measure** | **Value** |
| Number of nodes | 1000 |
| Number of edges | 1230 |
| Type of graph | Directed |
| Graph Density | 0.001 |
| Network Diameter | 2 |
| Average Degree | 1.23 |

*Summary of global network measures*

The low graph density value suggests that the users are sparsely connected.

**1.2 Degree Report**

A degree report can give an idea of how users are connected, i.e., an idea of the number of followers and the number of other users they are following. The graph shown in the figure can be modified for a better understanding the connections of these users with anxious tweets.

****

*Degree Distribution*

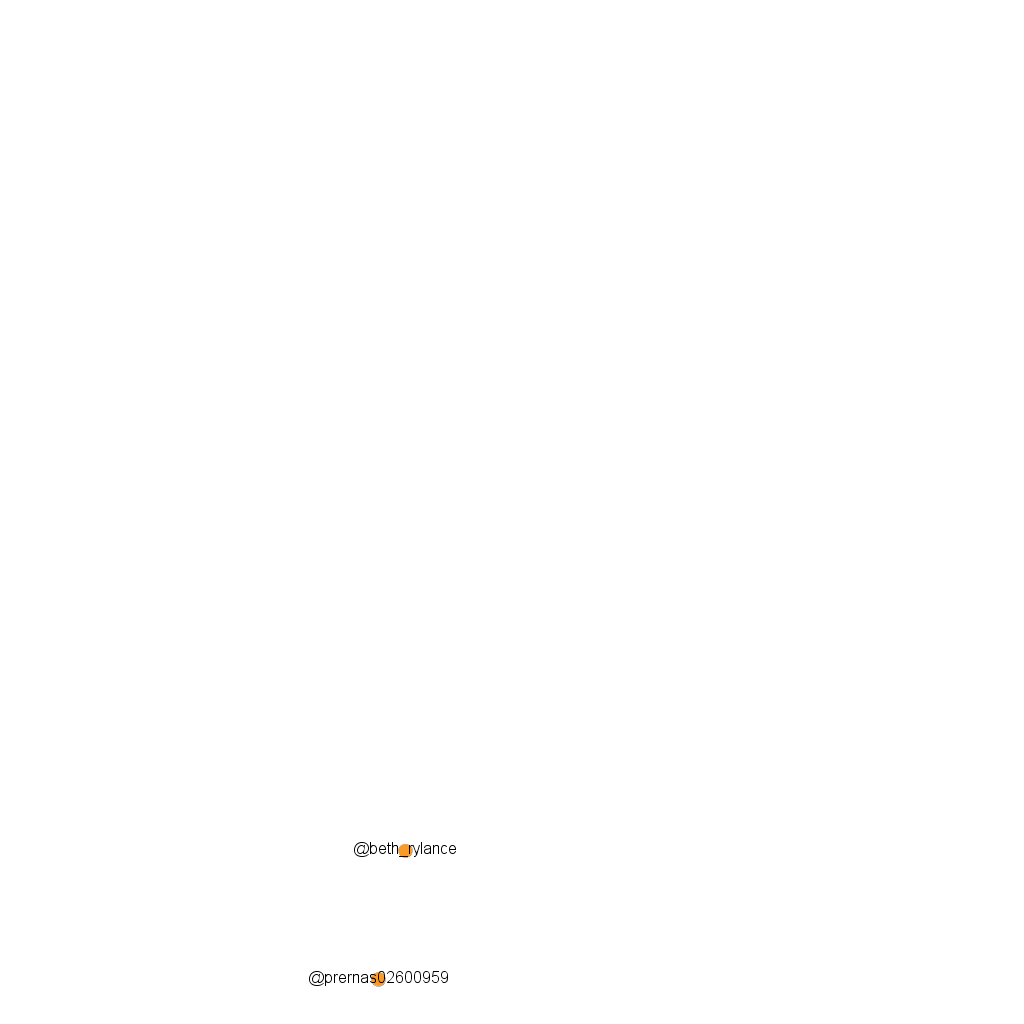
By using the OpenOrd layout, followed by the Noverlap layout, and some colouring and sizing options, we end up with a graph as shown in figure above. The bigger and reddish the nodes are, the higher their degree is. An analyst can apply some filters to identify the most connected and least connected users. Enabling the “show labels” option allows to see the usernames and identify them.

1. Users with the highest degree: These are the seven users who are the most connected to other anxious users in this network. The highest degree value observed for this network is 34, for the user @luvlyea. If the count increases, the vulnerability of being exposed to mentally depressed tweets will increase.



*Users with most degree*

1. Users with the lowest degree: The lowest degree identified in this network is 0 for two users, which means that they users are not connected to other depressed nodes considered. One can consider @beth\_rylance and @prernas02600959 to be in a safe zone and not exposed to such tweets of this network.



*Users with least degree*

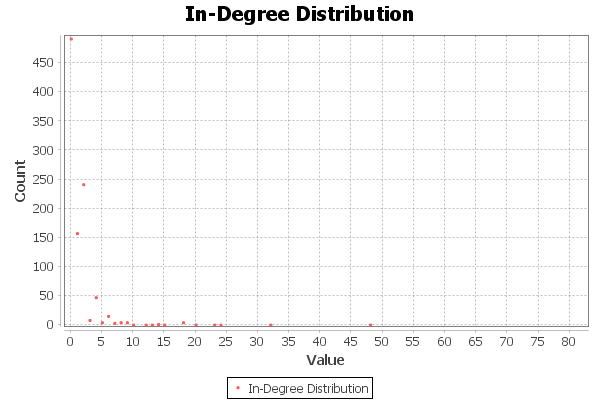
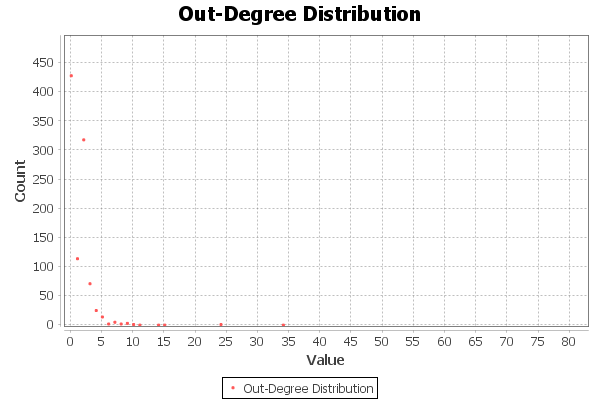
It is also important to note that the value of average degree is close to 1, which means that on an average, the nodes in this network are connected to one other user. A small value shows that they are not connected to too many, but hardly one depressed user.

###### 2. Node Level Measures

Through this section, we shall have a closer inspection of the nodes of the network and uncover some interesting trends about his network. An important part of node-level measures in Social Network Analysis is attributed to the concept of Centrality measures, where the potential and influence of individual nodes can be discovered. Another concept called Modularity allows us to identify the tendency of forming groups within a network. It must be kept in mind that we are looking at tweets that mentioned words like anxiety, suicide and depression, so the results will be different from randomly collected tweets. Another thought to bear in mind is that not all tweets that mention, say, ‘depression’ would mean mental distress, as the word in few cases can be used in other contexts as well. The Data Laboratory window can help us get a look into the tweets on the users considered.

**2.1 Degree Centrality**

Since the overall degree report is already studied, the in-degree and out-degree of the nodes can now be examined. If the plots for both the parameters are examined, a conspicuous trend can be observed in both of them: most of the nodes that mention words that indicate a low mental state have a very low in-degree and out-degree to other such nodes. Most people (around 450 in count) tend to a zero value, while the other nodes may have a greater value but populate at the lower degree values on the scale. In social media terminology, less in-degree and out-degree indicate less followers count and following count respectively.



*In-degree and Out-degree Distribution*

It is also interesting to observe that the users that had high degree are also the same ones that have a high in-degree value. Now, we can change our view about their vulnerability – since the high degree value is due to incoming nodes, their followers are the ones who would be affected by anxious tweets. One such tweets by these seven users (figure 4.6), can influence their whole networks.



*Users with high in-degree*

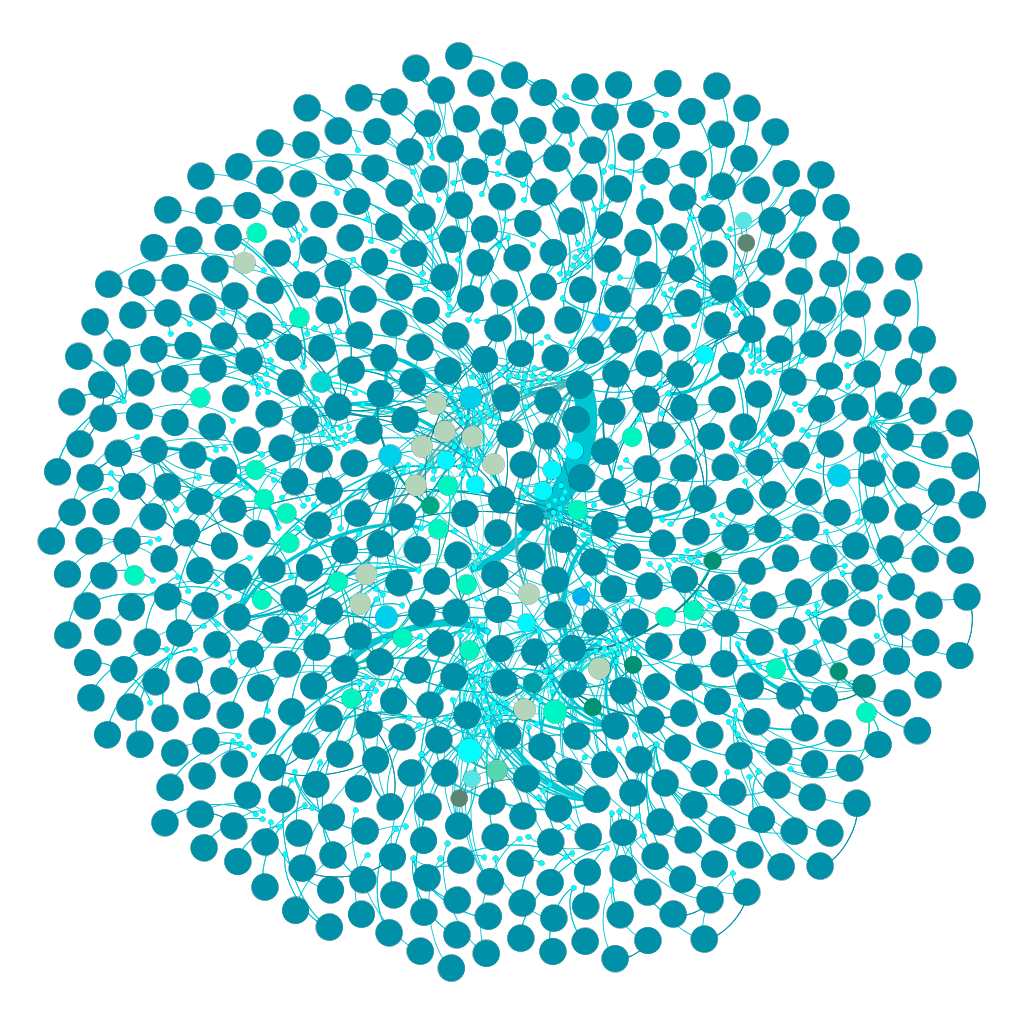
As for out-degree, the following figure shows the users with the highest value. The users @mistissrf, @healthyfellow and @leadcoalition, each follow twelve other users from this anxiety network. This is a very small number compared to the thousand overall, and an analyst can decide if they want to consider these users vulnerable only if the number is high.



*Users with high out-degree*

**2.2 Closeness Centrality**

Closeness centrality is the average shortest distance from a given starting node to all other nodes in the network. However, for this centrality measure to be applied, we need the network to be a fully connected component. In that, we do not need a complete graph, but a network where there is only one individual component. So, if we attempt to rank all the nodes in our network from the least to high centrality, and partition the colours based on the values, we would get something as shown in below.

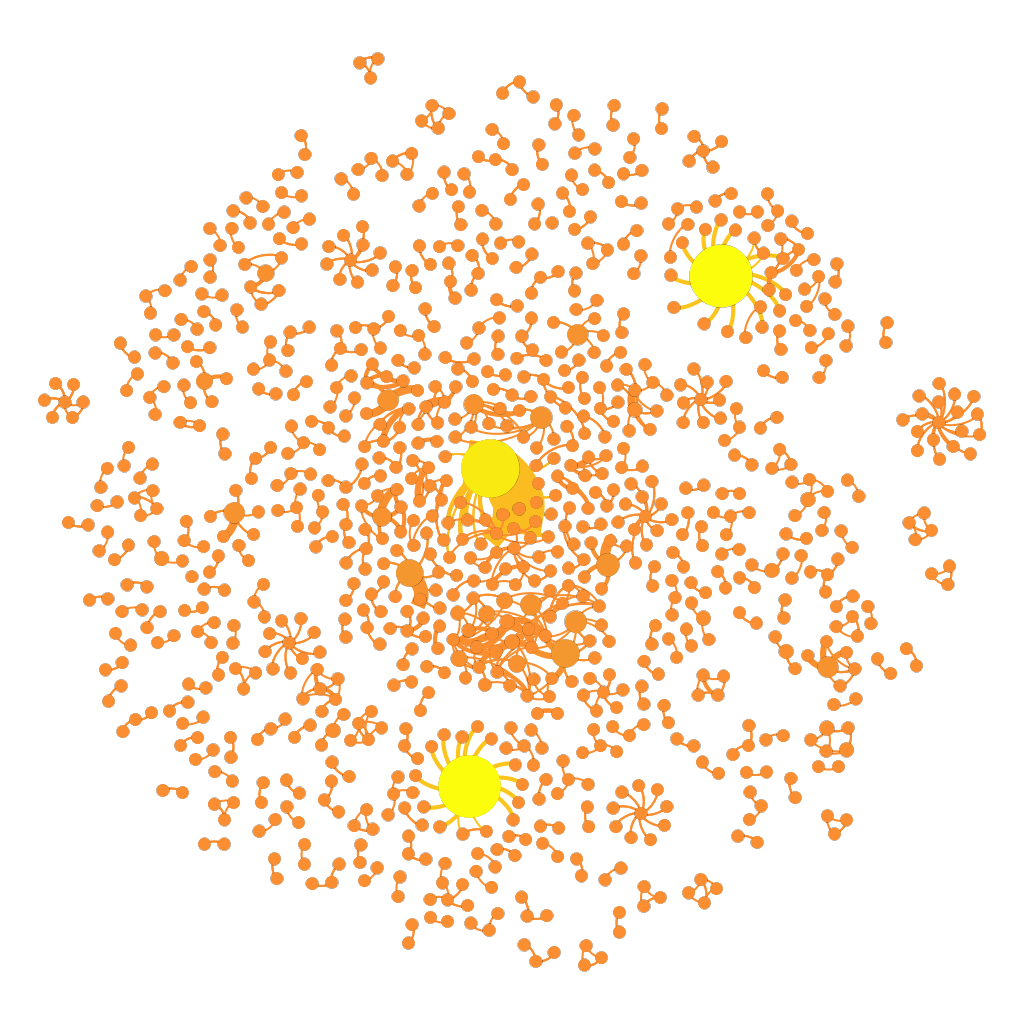
­

*Closeness Centrality Distribution*

Here, most nodes are of the same size and colour; these are the ones with closeness centrality value as 1. In addition to that, there are as many as 320 communities (which will be discussed shortly). That implies that there are multiple components in the graph, hindering us from calculating the maximum closeness centrality value of the whole network. On closer observation, one can infer that most components are made of two nodes, and hence the value 1 for them, they are easily accessible as they are not very much connected.

**2.3 Betweenness Centrality**

A handy benefit to betweenness centrality is that we do not require a fully connected graph or a component. A similar visualization as that of the previous measure, by ranking the nodes according to their value of betweenness centrality is shown in figure. The large, yellow nodes are the ones with the highest value. The next following figure gives a closer look at the users.



*Betweenness Centrality Distribution*

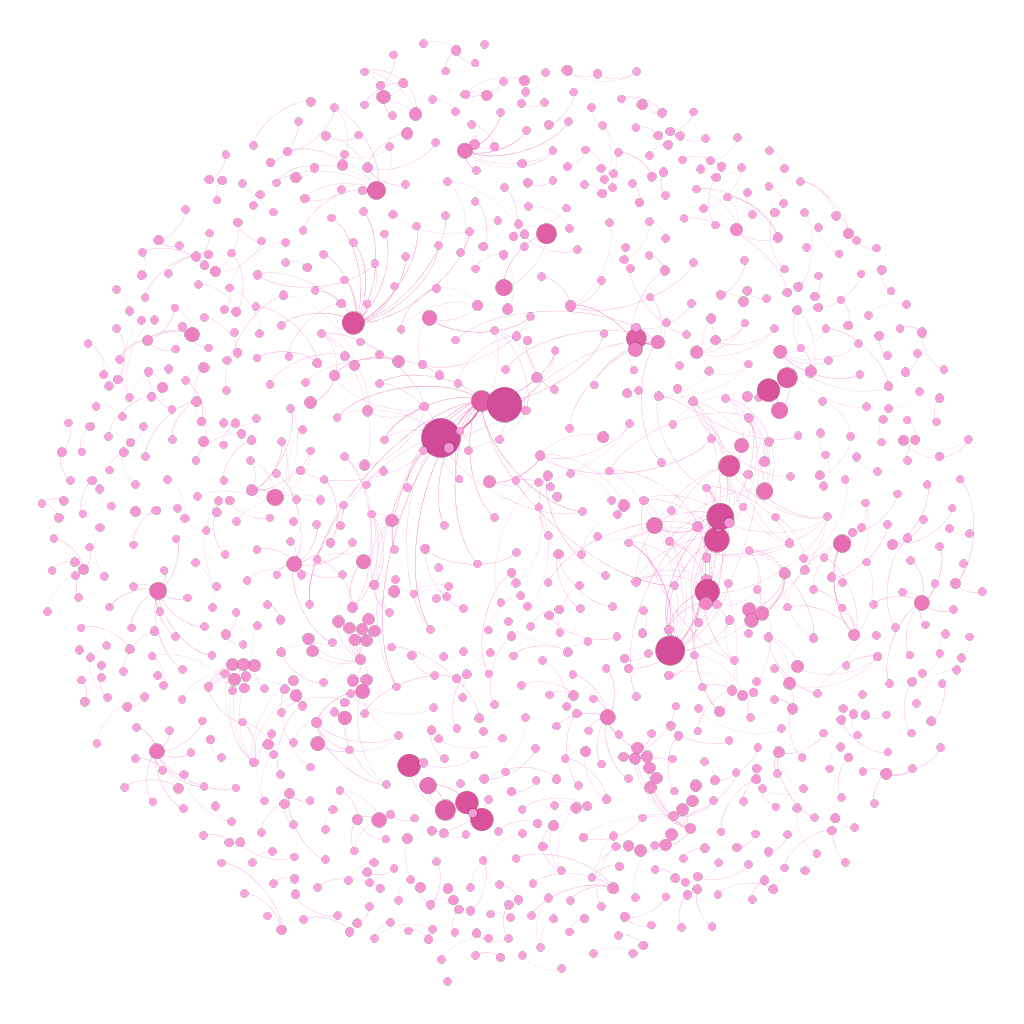


*Users with most Betweenness Centrality*

The users @lead\_coalition, @keitholbermann and @healthyfellow are important controllers of power and information because they bridge different components and hold the ability to control the dissemination of data across the network. Hence, the more anxious they are, the more they induce such ideology in their networks through their tweets. These kinds of users must be identified first when trying to prevent deteriorating mental health, and also on the contrary, use them for spreading positivity.

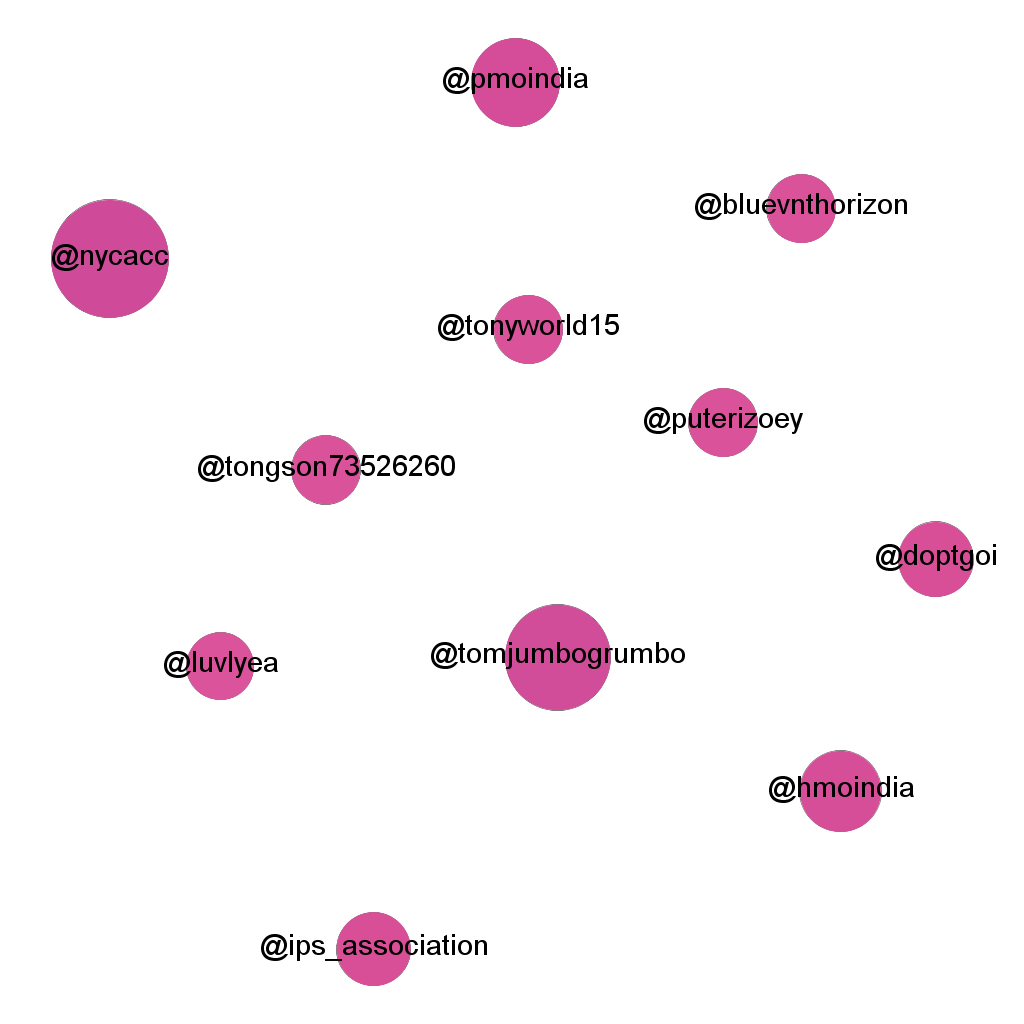
**2.4 Eigenvector Centrality**

A high eigenvector score means that a node is connected to many nodes who themselves have high scores, so that connecting to some vertices has more benefit than connecting to others. Their importance propagates out to the nodes to which they are connected. So, if we take a look at the graph for that in figure, we realize that there are a few influential people in the network who tweeted about their anxiety and depression.



*Eigenvector Centrality Distribution*

After filtering out the nodes and performing rearrangement, the users with high eigenvector centrality are as shown in figure below.

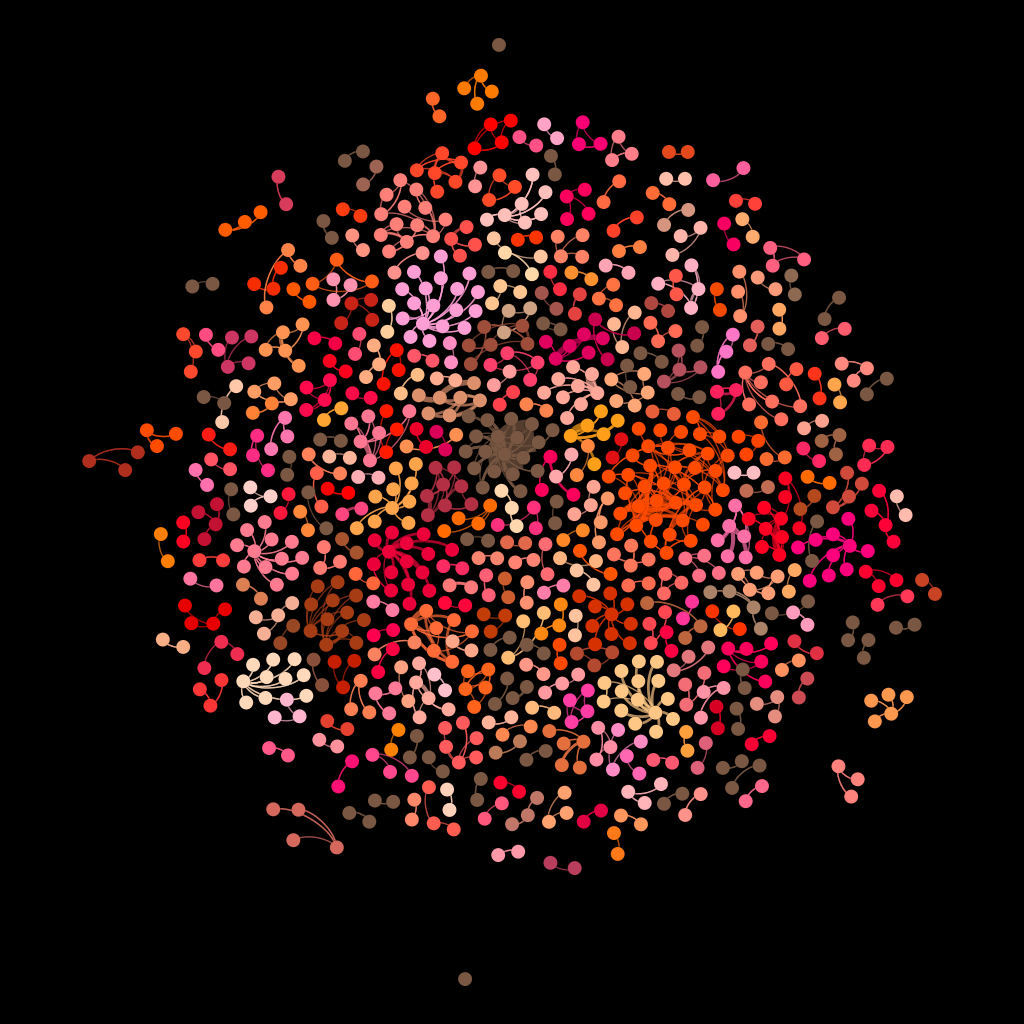


*Users with high Eigenvector Centrality*

Eigenvector centrality seems be identical to in-degree centrality because we might attach importance to those who have many followers. But even if a user may low degree, they might exhibit high eigenvector centrality. This is due to the status they hold. Since they are powerful actors in a network, their followers widely echo their notions and ideologies, directly or indirectly. These actors must be careful in expressing opinions.

###### 3. Modularity

People tend to group into clusters on social media platforms. But figure below shows that there are too many communities, with each colour representing a community. This is typical and expected of anxiety networks because such users don’t tend to socialize much. So, the users either be a part of small groups or have small degree value. In the network considered, there are 320 different communities, where anxious users are not very much connected to others in the same lot, but there are some clusters found.



*4.13 Modularity Analysis*

**CONCLUSION**

In this Social Network Analysis of users who tweeted about anxiety, depression, suicide etc., an overall idea was establish about the network, while also delving into the behaviour of particular users. The data was collected dynamically by using a Twitter API key obtained through a Twitter developer account. The overall degree distribution, in-degree and out-degree distributions, Closeness Centrality, Betweenness Centrality, Eigenvector Centrality and Modularity were considered and drawn into graphs. The outcomes that shed light on some specific users exhibiting particular characteristics and how that can impact and affect their networks. This was done by the usage of Gephi, a visualization software that enabled the analysis through real-time data collection, various layouts, colour coding and sizing options, statistical applications and filtering. As a result, the considered network was analysed based on the tweets collected.

**FUTURE SCOPE**

Gephi allows sophisticated options for visualization and analysis. The tool is also often paired with Python and R to bring out the results of the statistical measures. However, the analysis can be extended by drawing more information about the users. It is understandable that social media websites do not entertain the disclosure of personal information, but analysts who work for these platforms can, under the law by following the protection of privacy, draw more insights by segregating the user nodes in terms of age, employment etc. Again, this is not to intervene with the public’s data, but to understand the behaviour of social networks better and take situation-specific measures. Furthermore, as Social Network Analysis is still growing, more algorithms, layouts and measures can be brought under study to produce advanced graphs.