**GRAPH ANALYTICS**

**AND**

**ALGORITHMS - 18MAT333**

**CASE STUDY**

**ON**

**CRIME NETWORK ANALYSIS**

* AASHIKA B S

( CB.SC.I5DAS19003)

* KRITHIGA M (CB.SC.I5DAS19020)

TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TOPIC** | **PAGE.NO** |
| 1 | ABSTARCT | 3 |
| 2 | MOTIVATION | 4 |
| 3 | DATASET DESCRIPTION | 4 |
| 4 | NODES AND EDGES | 5 |
| 4 | VISUALIZATION OF CRIME NETWORK -NETWORKX | 6 |
| 5 | VISUALIZATION OF CRIME NETWORK –  PyViz | 7 |
| 6 | CENTRALITIES | 9 |
| 7 | DEGREE CENTRALITY | 9 |
| 8 | BETWEENNESS CENTRALITY | 12 |
| 9 | CLOSENESS CENTRALITY | 16 |
| 10 | EIGENVECTOR CENTRALITY | 20 |
| 11 | CLUSTERING COEFFICIENT | 24 |
| 12 | CONCLUSION | 26 |
| 13 | SOFTWARES USED | 29 |
| 14 | REFERENCES | 30 |

**ABSTRACT:**

* Crime or criminal offence is an act harmful not only to individuals but also to a community, society, or the state. Such acts are forbidden and punishable by law.
* With increased usage of digital gadgets and apps that offer several services, this space adds much more to crimes including cheating, theft or hacking.
* Network analysis techniques in criminal intelligence are used to organise data and reveal patterns in the nature and extent of relationships between data points.
* Crime Network has been chosen to analyse about the different crimes committed by the accused to the victim around the world.



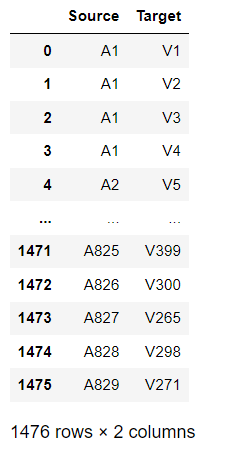
**MOTIVATION:**

* Crime network analysis helps to assist and support CBIs, Police to narrow down their search for accused and the victim.
* Helps to provide protection to the person who have been affected the most.
* Collect witness and information about the incident. Helps public to beware of the accused.

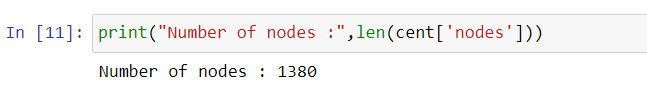
**DATASET DESCRIPTION:**

* This dataset contains accused and victims, which act as nodes.
* Edges between accused and victims represent the crime committed by accused to the victim.
* This network graph is an undirected as the nodes itself represent the Accused id and Victim id and unweighted as there is no data available about the type of crime committed.
* Nodes: 1380 (829 – Accused, 551 – Victim)
* Edges: 1476

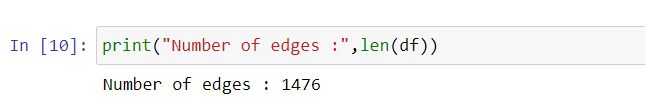
**EDGES DATAFRAME AND NODES DATAFRAME**

 ****

1. **NUMBER OF NODES:**



1. **NUMBER OF EDGES:**

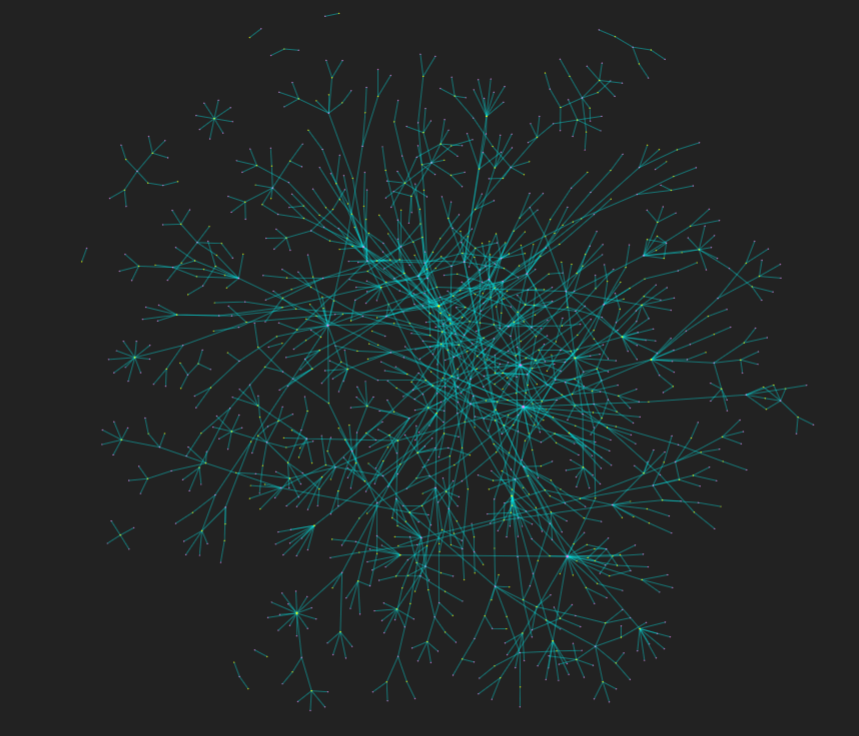


**CRIME NETWORK – USING Networkx**

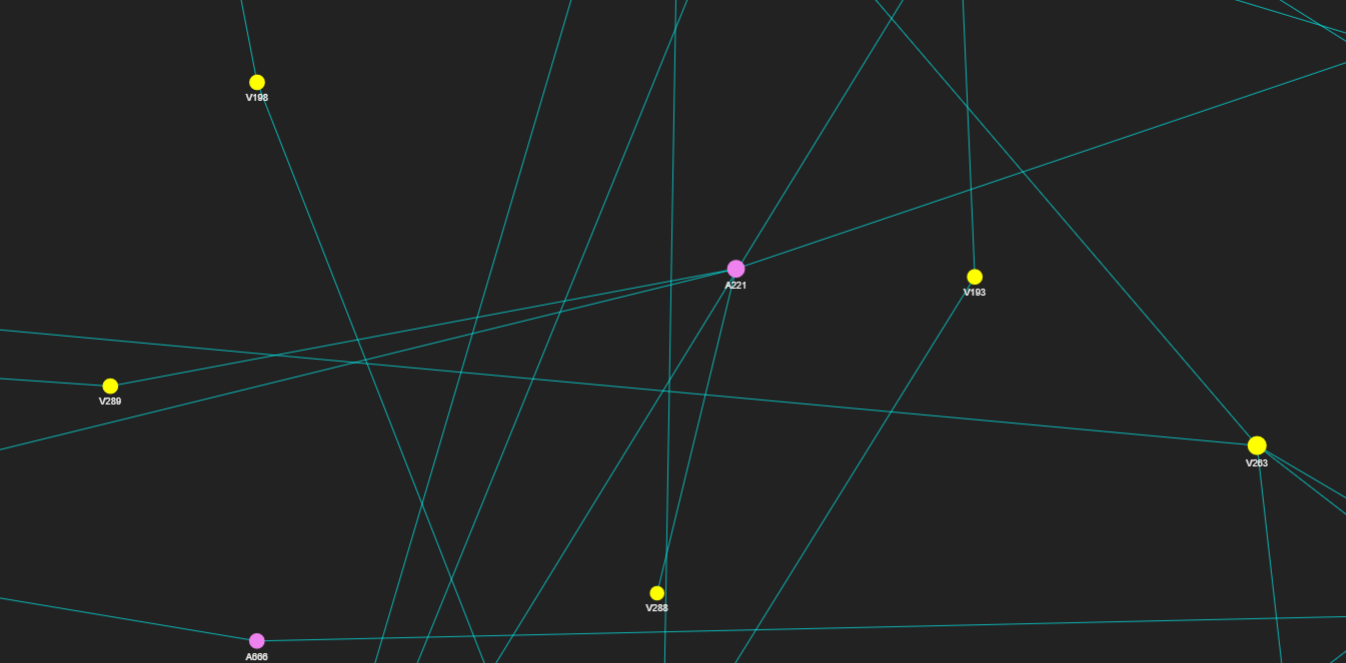
* Blue nodes: Victim
* Green nodes: Accused

****

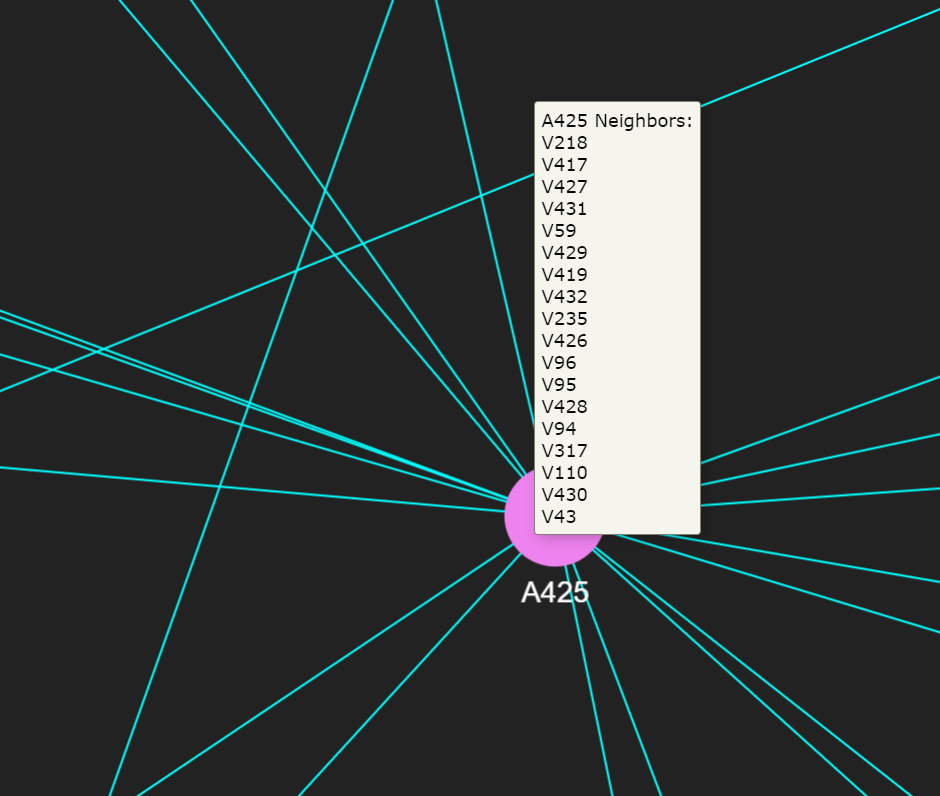
**CRIME NETWORK – USING PyViz**



**ZOOMED VIEW OF NETWORK**



**DISPLAYING THE NEIGHBOURS OF NODE:**



**CENTRALITIES:**

* Centrality gives an estimation on how a node or edge is important for the connectivity or the information flow of the network.
* It is used to measure the importance (or “centrality” as in how “central” a node is in the graph) of various nodes in a graph.
* Each node could be important from an angle depending on how “importance” is defined.

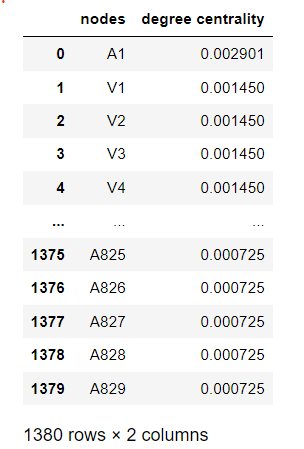
1. **DEGREE CENTRALITY:**

Degree centrality assigns an importance score based simply on the number of links held by each node. The degree centrality of a node in a graph is simply a count of the number of edges that connect to it.

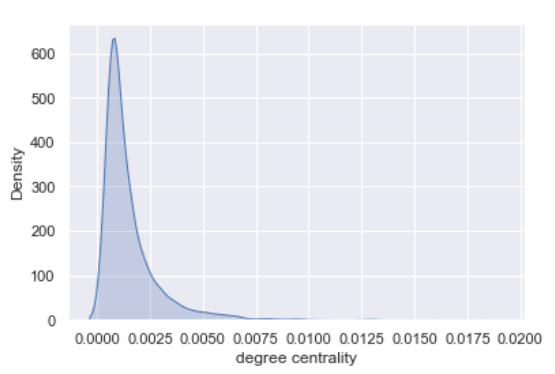
****

where Aij is the ij-th element of the adjacency matrix A of the graph and n is the number of vertices in the graph.

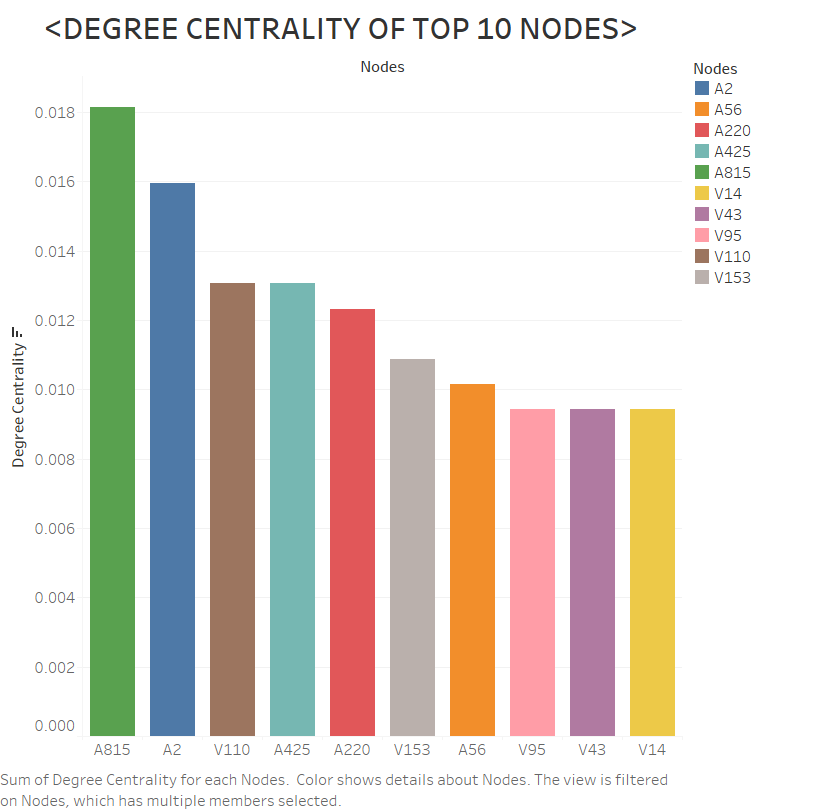
* **DEGREE CENTRALITIES OF NODES:**

****

* **DISTRIBUTION OF DEGREE CENTRALITIES:**

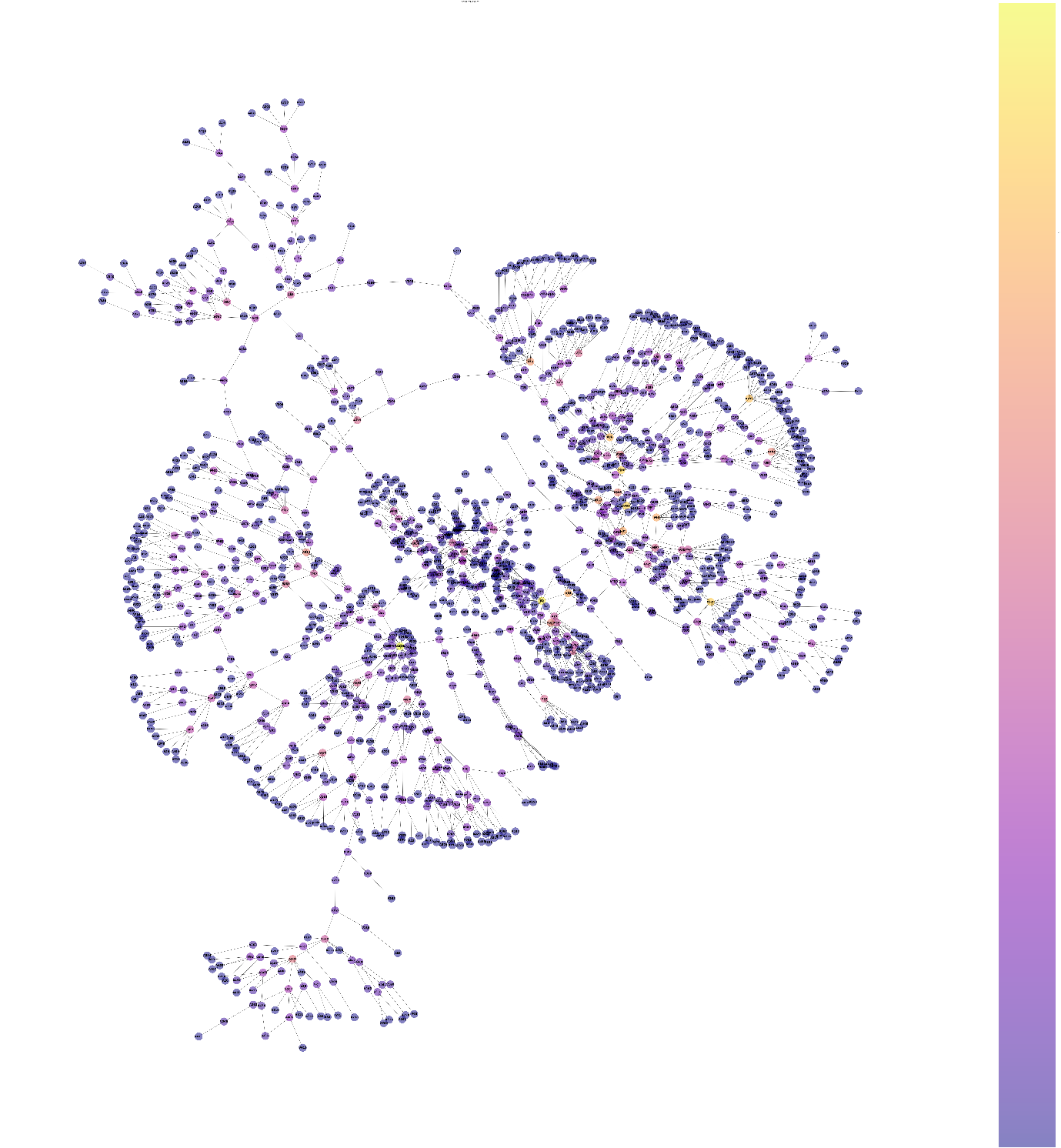
****

* **TOP 10 NODES RANKED ACCORDING TO DEGREE CENTRALITY:**



* **NETWORK REPRESENTATION OF DEGREE CENTRALITY:**

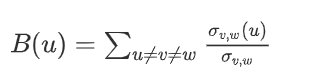
The colour bar represents the degree centralities of nodes. Dark violet colour represents the node which has low degree centrality and light-yellow colour represents the node which has high degree centrality.

****

* **INFERENCE:**

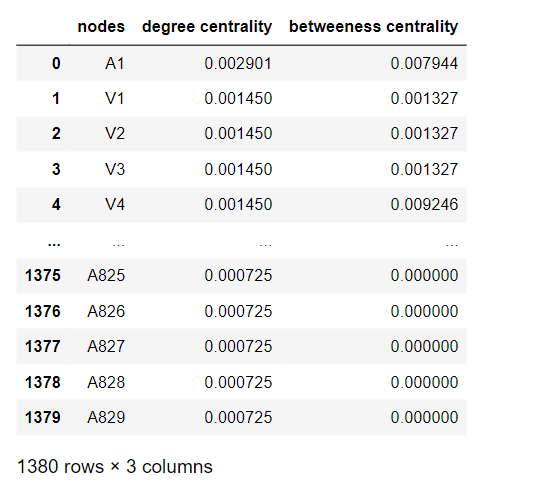
1. Degree centrality tells about the number of edges incident to the vertex. Higher the degree centrality, higher the number of connections a vertex has.
2. Here, ‘A815’ (Accused No. 815) has more degree centrality which implies that, he is the most common accused involved in most of the crimes.
3. Similarly, ‘A2’ is the second most common accused.
4. Among Victims, ‘V110’ is the person who have been affected by most of the crimes.
5. **BETWEENNESS CENTRALITY:**

Betweenness centrality is a way of detecting the amount of influence a node has over the flow of information in a graph. It is often used to find nodes that serve as a bridge from one part of a graph to another.

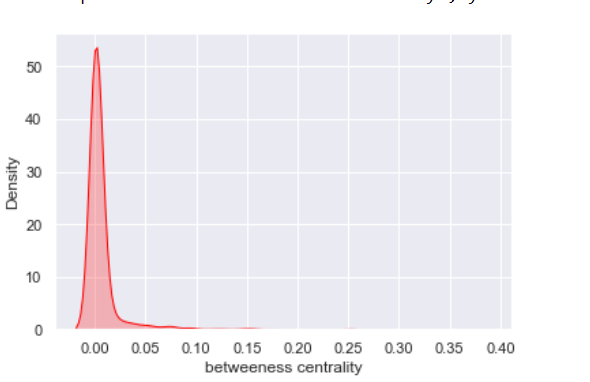


Where B(u) represents betweenness of vertex u v,w(u) represents shortest path between v and w through vertex u and v,w represents the shortest path between v and w.

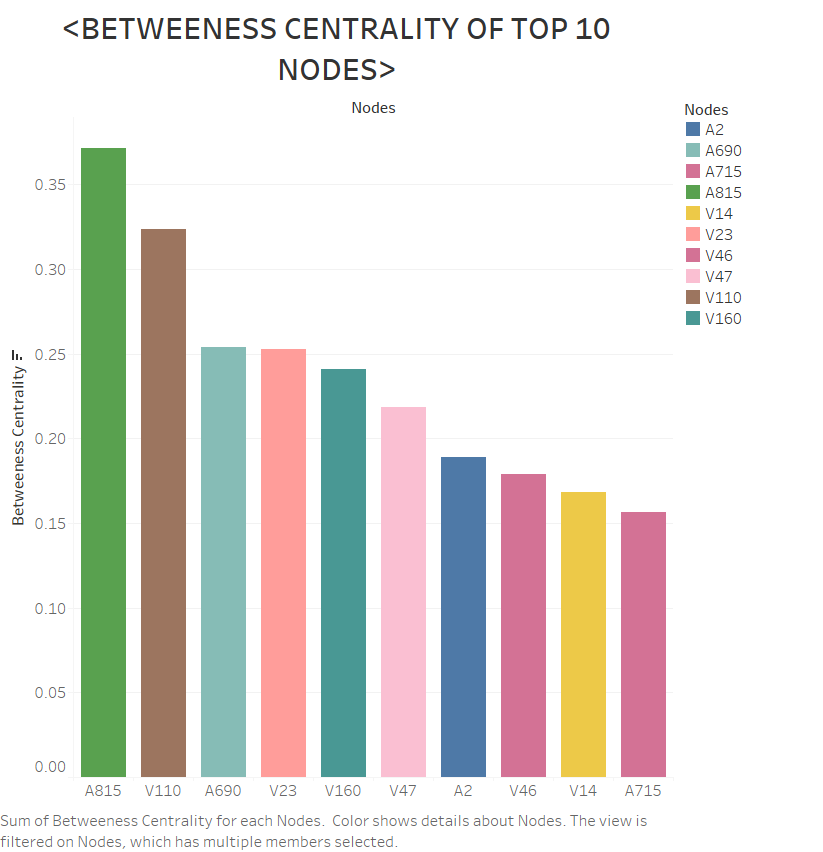
* **BETWEENESS CENTRALITIES OF NODES**

****

* **DISTRIBUTION OF BETWEENESS CENTRALITIES:**

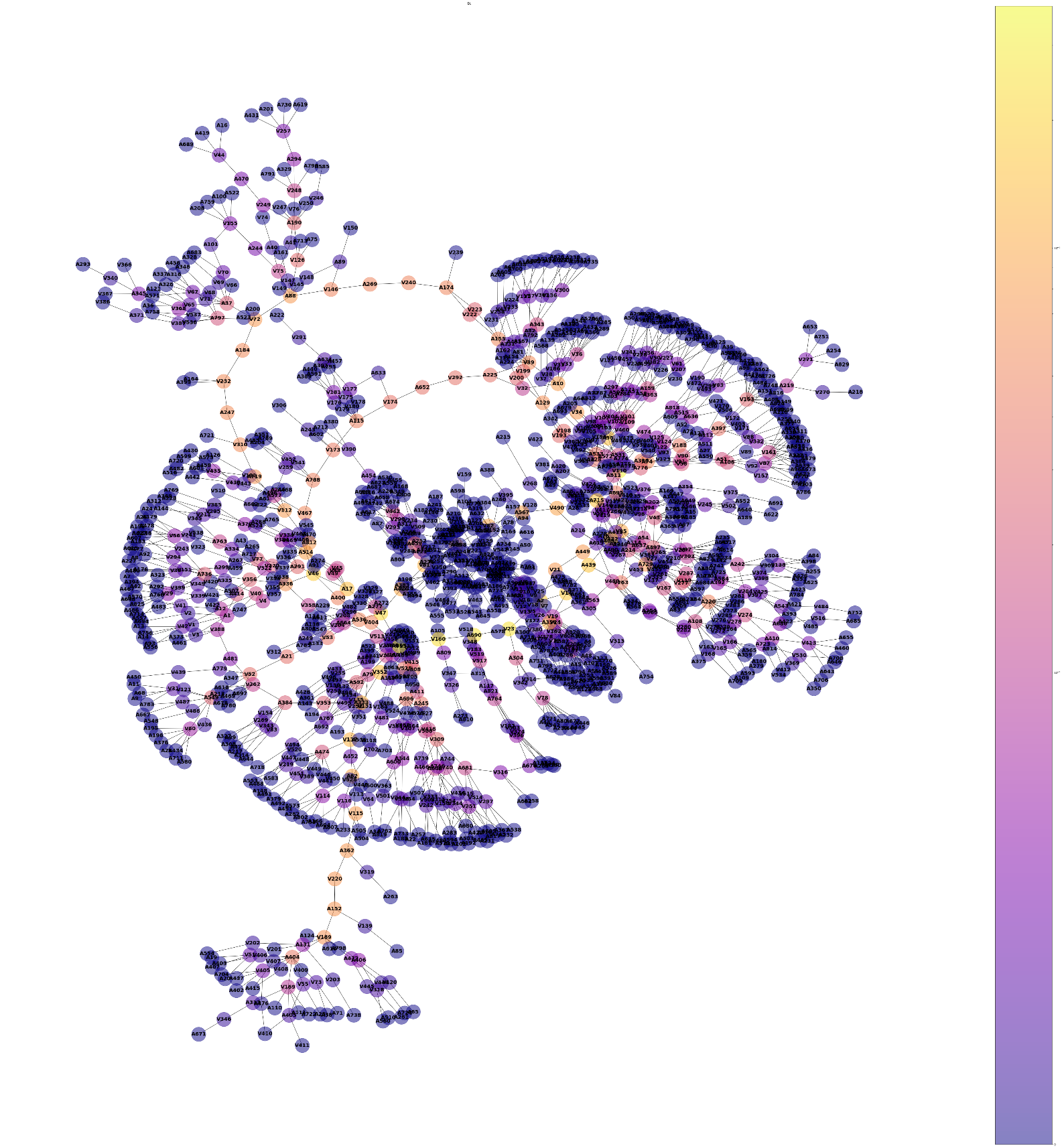
****

* **TOP 10 NODES RANKED ACCORDING TO BETWEENNESS CENTRALITY:**



* **NETWORK REPRESENTATION OF BETWEENNESS CENTRALITY:**

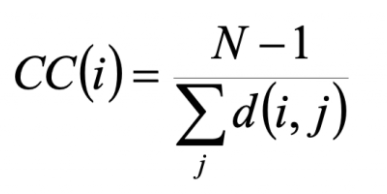
The colour bar represents the betweenness centralities of nodes. Dark violet colour represents the node which has low betweenness centrality and light-yellow colour represents the node which has high betweenness centrality.

****

* **INFERENCE:**

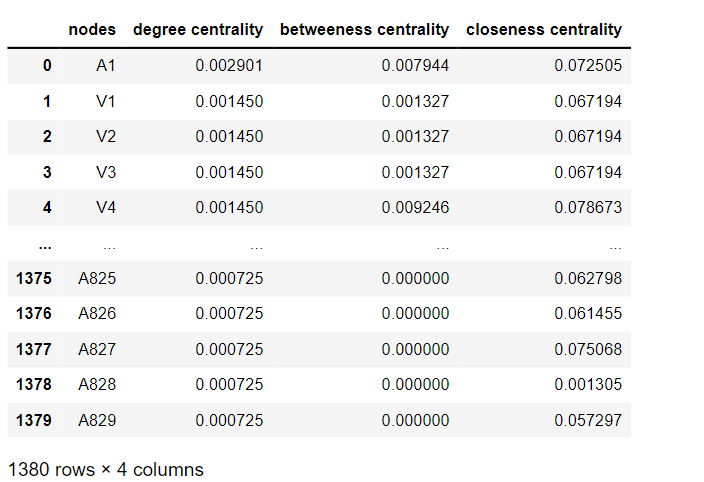
1. Betweenness centrality tells us about how a particular node serves as a bridge between the shortest path of two other nodes.
2. It is done by identifying all the shortest paths and then counting how many times each node falls on one.
3. Here ‘A815’ is the main accused who serves as a bridge between all the type of crimes.
4. Victim ‘V110’ is the most affected victim, who has been a bridge in all kinds of crime.
5. **CLOSENESS CENTRALITY:**

Closeness centrality is a way of detecting nodes that are able to spread information very efficiently through a graph. The closeness centrality of a node measures its average farness to all other nodes. Nodes with a high closeness score have the shortest distances to all other nodes.

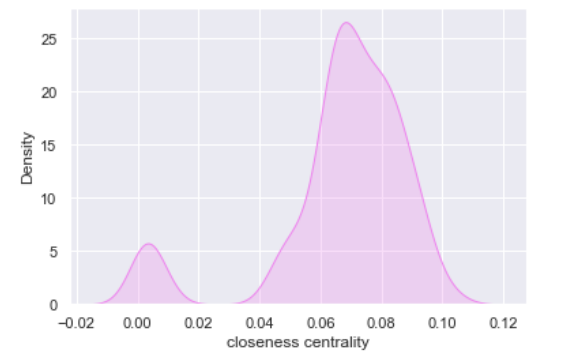


where N is number of nodes and d(i,j) is the length of shortest path between i and j (ij).

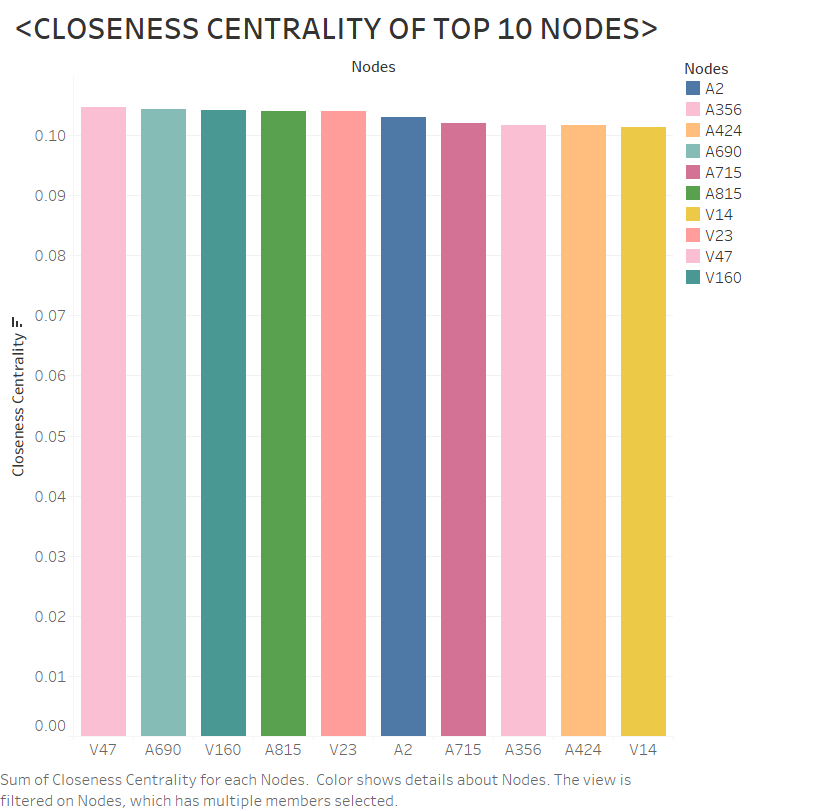
* **CLOSENESS CENTRALITIES OF ALL NODES:**

****

* **DISTRIBUTION OF CLOSENESS CENTRALITIES:**

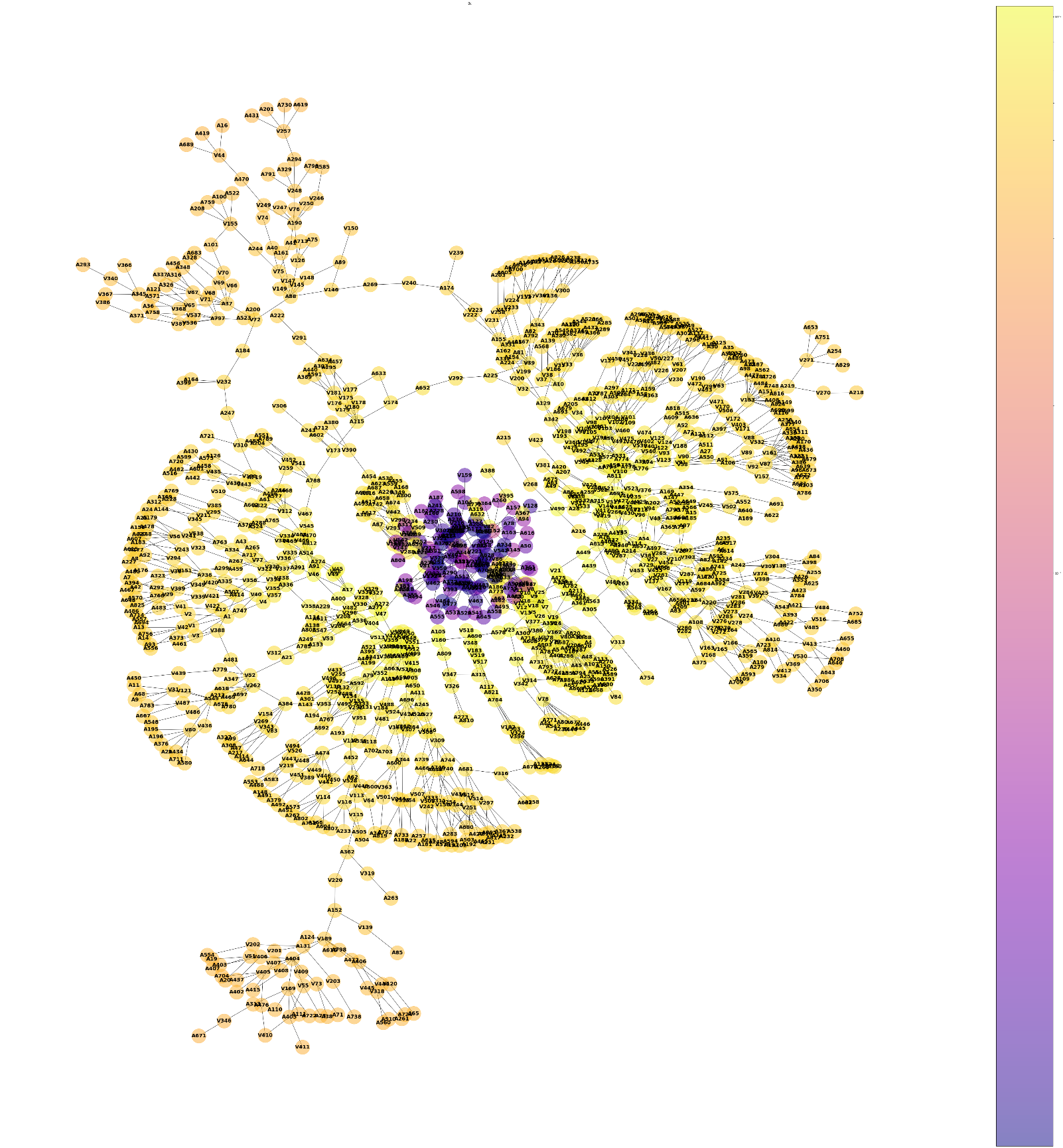
****

* **TOP 10 NODES RANKED ACCORDING TO CLOSENESS CENTRALITY:**



* **NETWORK REPRESENTATION OF CLOSENESS CENTRALITY:**

The colour bar represents the closeness centralities of nodes. Dark violet colour represents the node which has low closeness centrality and light-yellow colour represents the node which has high closeness centrality.

****

* **INFERENCE:**

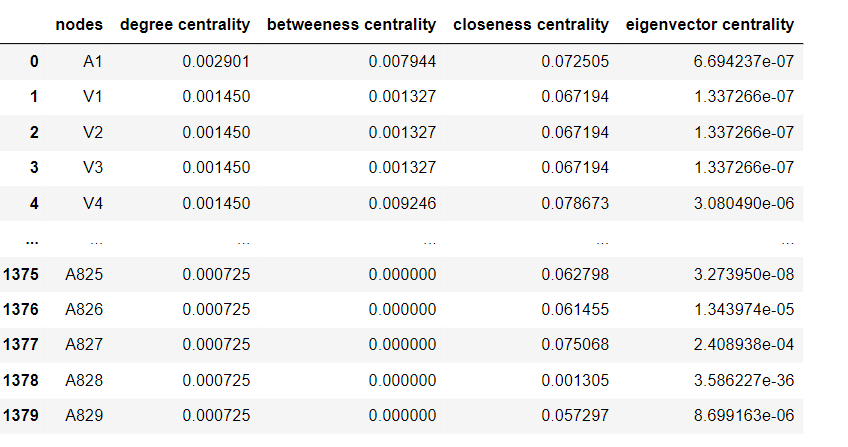
1. Closeness centrality tells how a particular node is close to all the other nodes in the graph. Higher the closeness centrality, the node is reachable to all the other nodes in the graph.
2. Here, ‘V47’ (Victim No.47) is the person who is reachable to all the accused and victims.
3. It means, ‘V47’ is the most influential person in the network, which implies that information about the crime, evidence can be collected from that person.
4. Among the accused, ‘A690’ is the most influential person. If the person or crime cannot be found from ‘V47’ network, we can opt for ‘A690’ network.
5. From the network graph, it can be seen that many of the nodes are close to all the other nodes in the network.
6. **EIGENVECTOR CENTRALITY:**

Eigenvector centrality is used to measure the level of influence of a node within a network. Each node within the network will be given a score or value: the higher the score the greater the level of influence within the network.

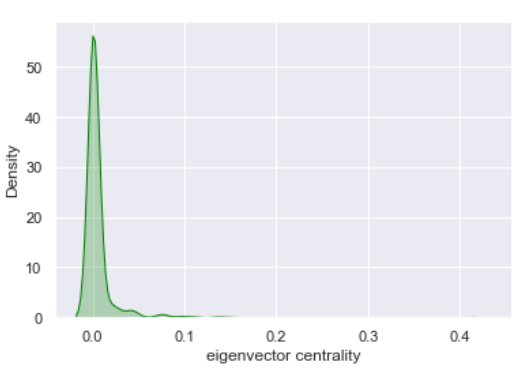


where is constant and Aij is the ijth element of the adjacency matrix.

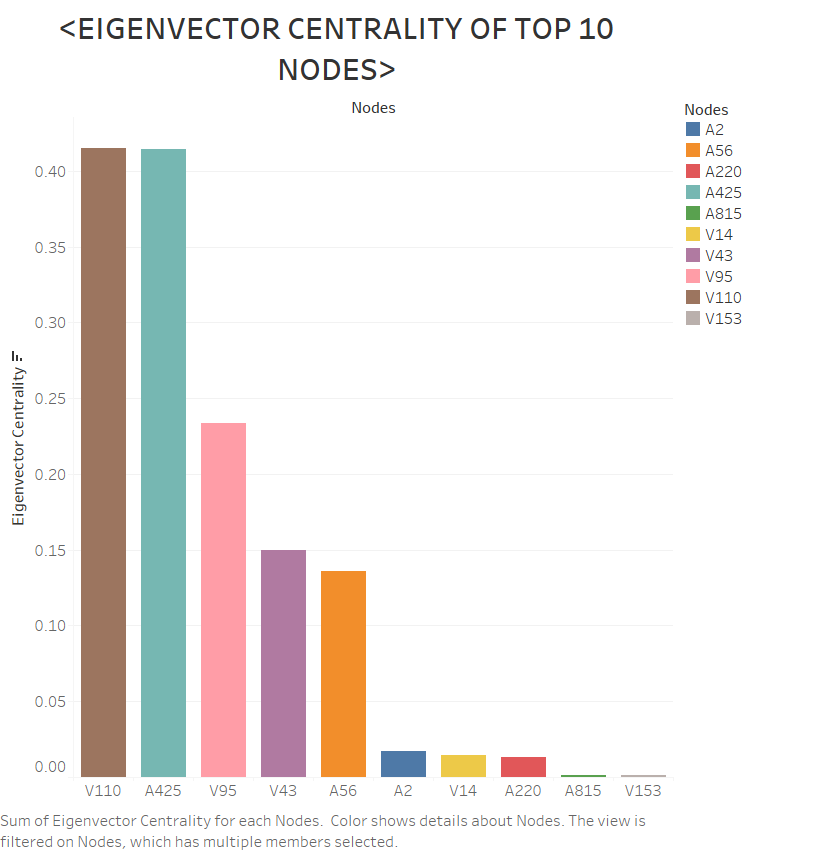
* **EIGENVECTOR CENTRALITIES OF ALL NODES:**

****

* **DISTRIBUTION OF EIGENVECTOR CENTRALITIES:**

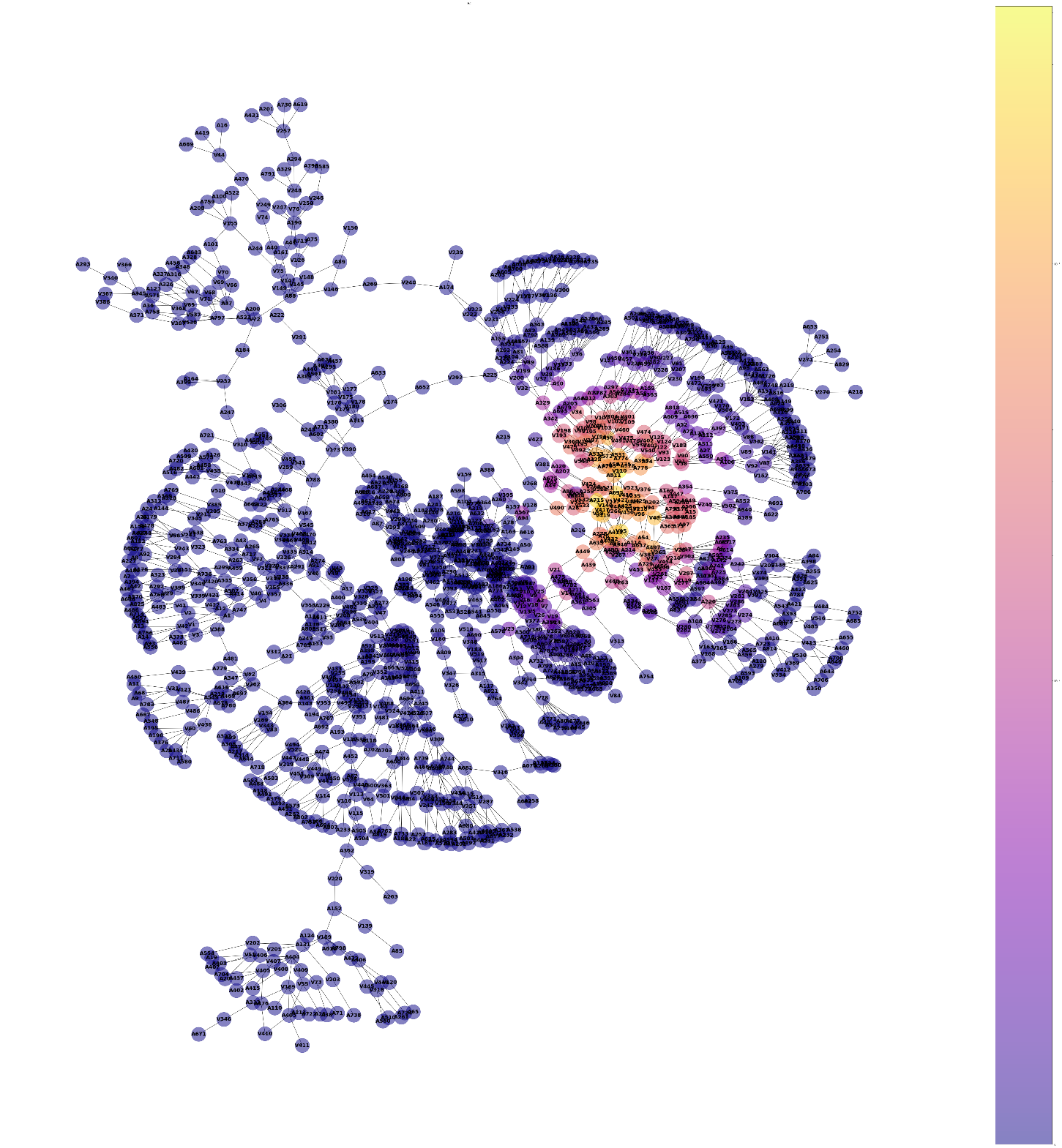
****

* **TOP 10 NODES RANKED ACCORDING TO EIGENVECTOR CENTRALITY:**



* **NETWORK REPRESENTATION OF EIGENVECTOR CENTRALITY:**

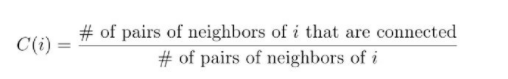
The colour bar represents the eigenvector centralities of nodes. Dark violet colour represents the node which has low eigenvector centrality and light-yellow colour represents the node which has high eigenvector centrality.

****

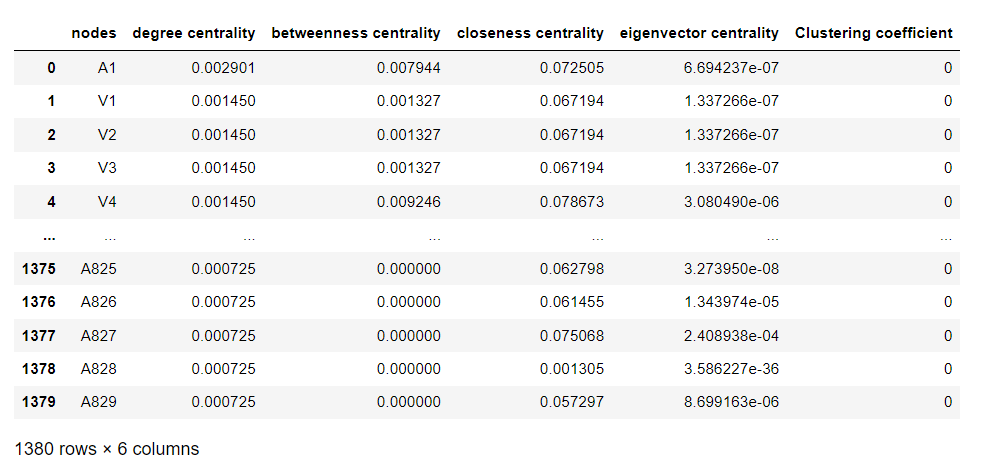
* **INFERNECE:**

1. Eigenvector centrality tells us about the extended connections of a node. Eigenvector Centrality can identify nodes with influence over the whole network, not just those directly connected to it.
2. Here ‘V110’ and connections of ‘V110’ has more influence over the network.
3. In order to find an accused or victim, one can go through the edges of ‘V110’ and follow the network to narrow down their search.
4. Among accused, ‘A425’ and the connections of ‘A425’ has the influence over network. If one cannot find the solution in the connections of ‘V110’, second option would be ‘A425’.
5. **CLUSTERING COEFFICENT:**

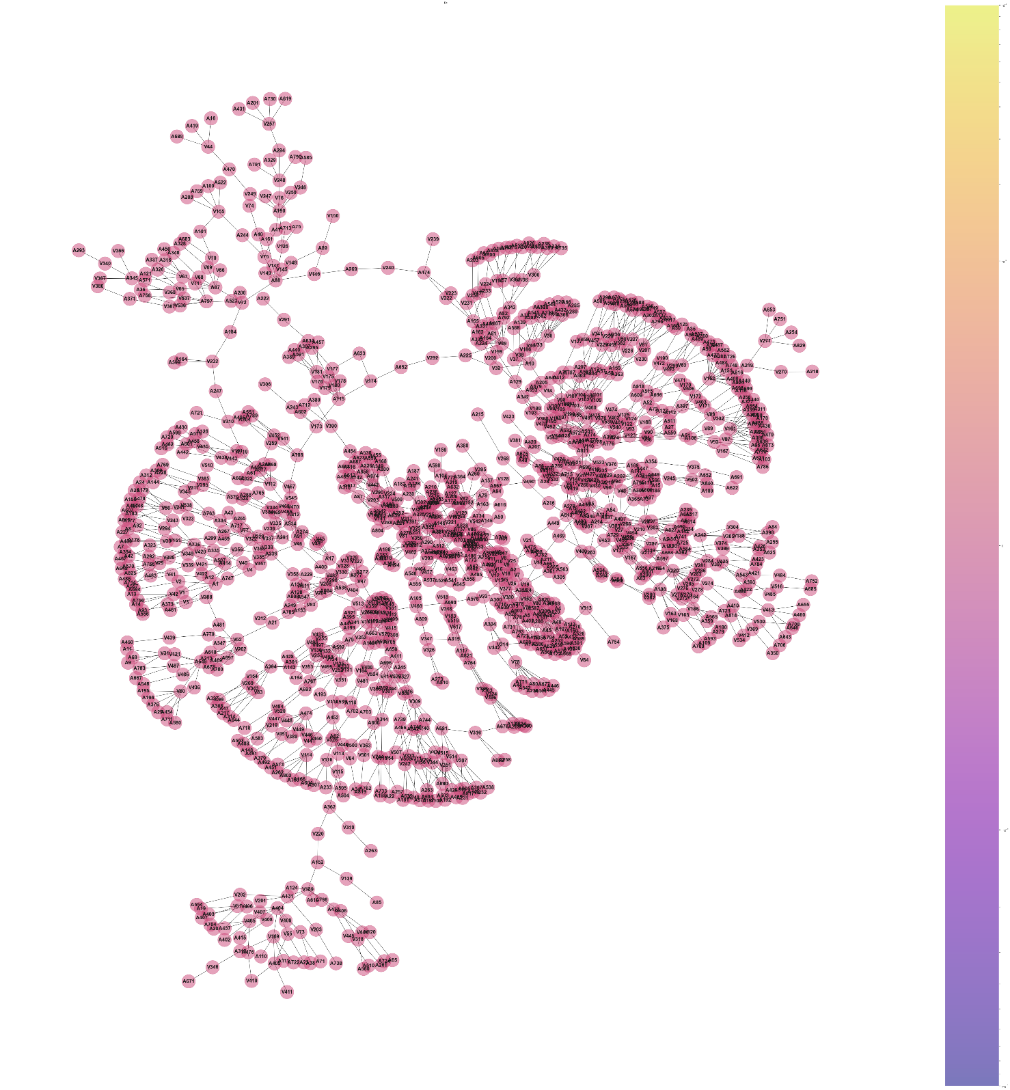
Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. It is the proportion of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them.



* **CLUSTERING COEFFICIENT OF ALL NODES:**



* **NETWORK REPRSENTATION OF CLUSTERING COEFFICIENT:**

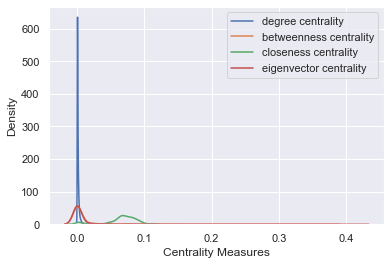


* **INFERENCE:**

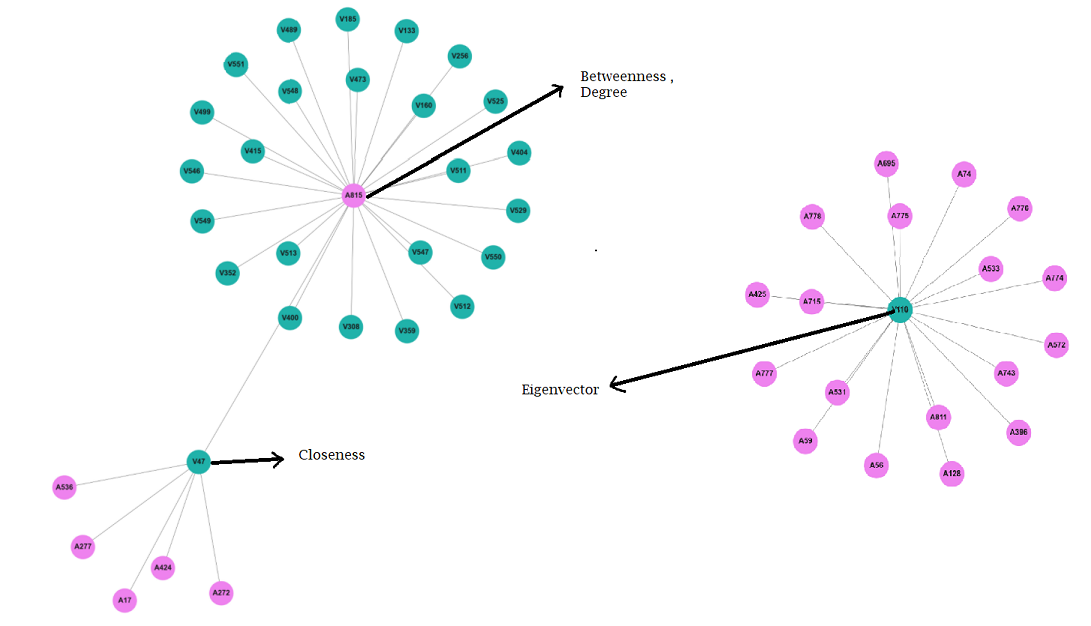
1. Clustering coefficient tells how well connected the neighbourhood of the node is.
2. If the neighbourhood is fully connected, the clustering coefficient is 1 and a value close to 0 means that there are hardly any connections in the neighbourhood.
3. Since, all the crimes are independent of each other, they do not form clusters in the network rather they have only connections.

* **CONCLUSION:**

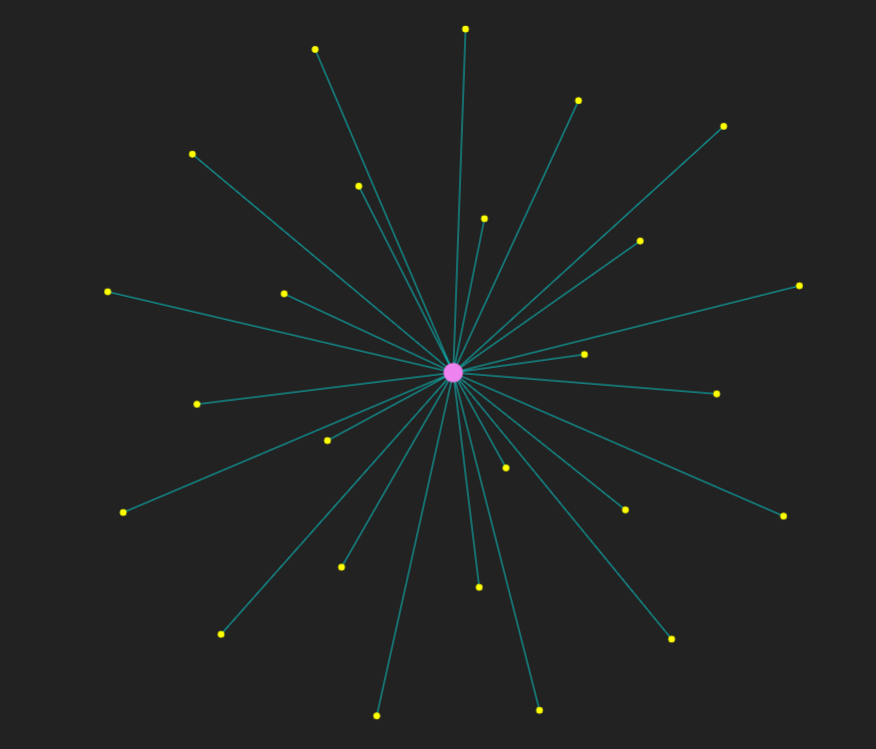
**OVERALL DISTRIBUTION DEPENDING ON CENTRALITIES:**

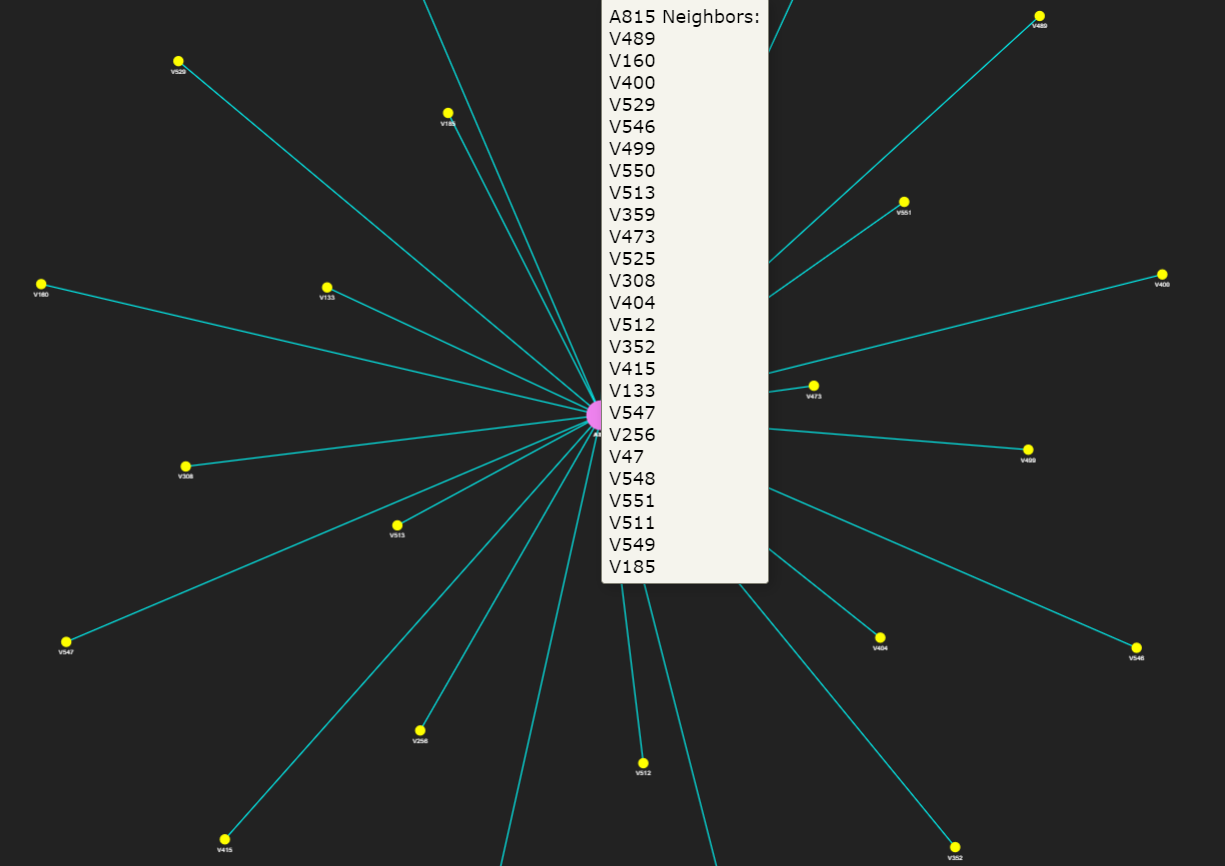
****

**OVERALL NETWORK DEPENDING ON CENTRALITIES:**

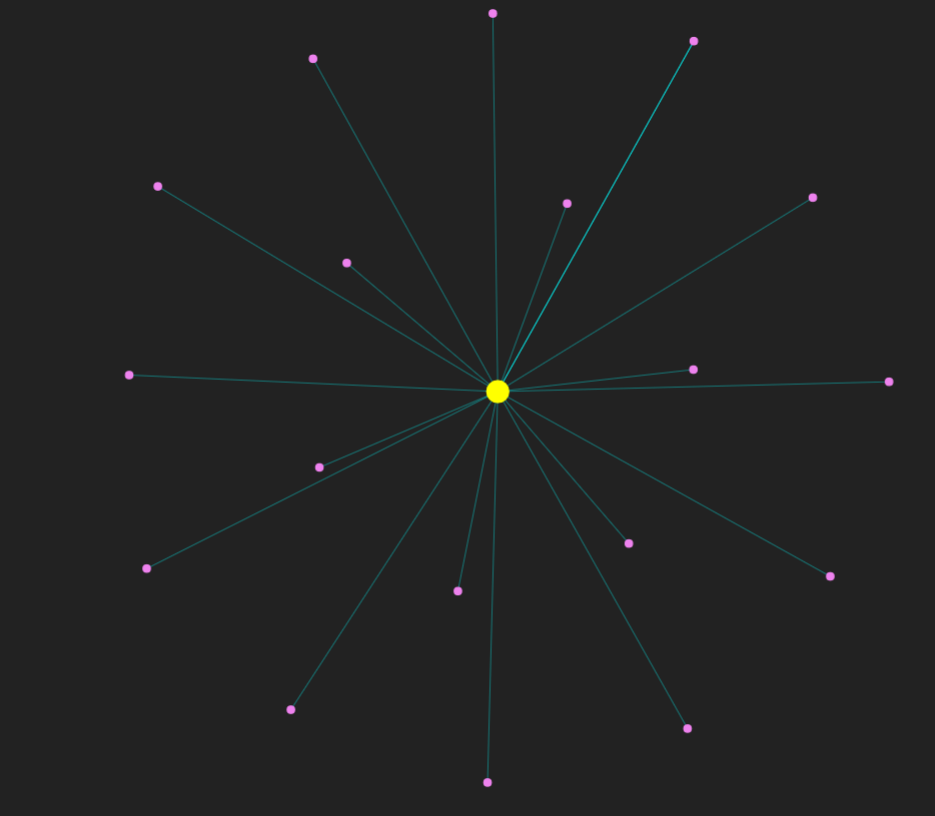
****

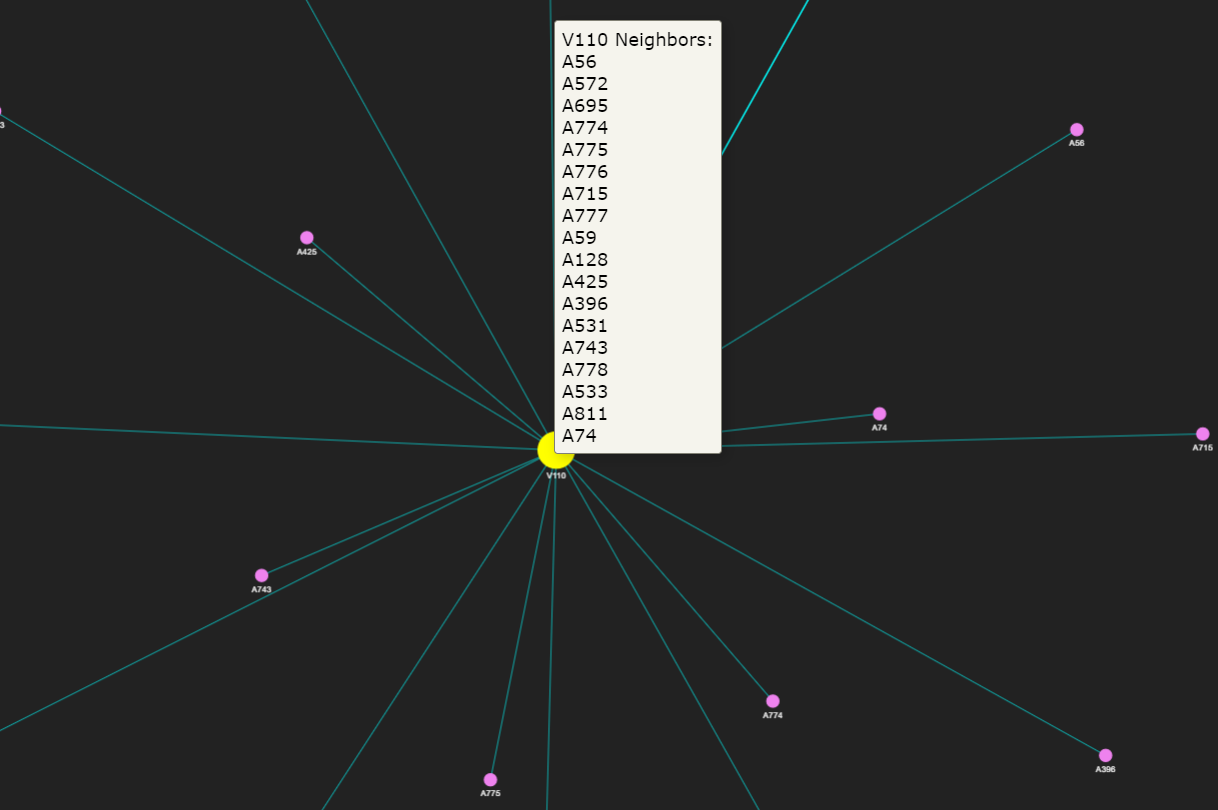
1. ‘A815’, ‘V110’, ‘V47’ are the most influential persons in this crime network. With the help of these three persons, investigation about the crime, accused, victim can be found easily.
2. ‘A815’ is the most common accused in all of the crimes





1. ‘V110’ is the most common victim in all of the crimes.

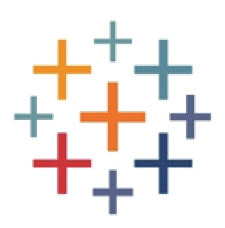
****

****

Similarly, for any given crime network, with the help of centrality measures, criminal cases can be solved easily, which would be very much helpful for criminal intelligence.

**SOFTWARE USED:**

* Tableau used for our data’s statistical analysis (investigating the patterns, trend and relationship between themselves) and graphical analysis (depiction of data using charts, figures and graphs).
* Jupyter notebook used for the python coding to get the information of data, to calculate different centralities, coefficients and many more.
* HTML used to create interactive network graph.

**REFERENCES:**

* [Wikipedia](https://www.wikipedia.org/)
* [Analytics Vidhya - Learn Machine learning, artificial intelligence, business analytics, data science, big data, data visualizations tools and techniques. | Analytics Vidhya](https://www.analyticsvidhya.com/)
* [Python Programming Language - GeeksforGeeks](https://www.geeksforgeeks.org/python-programming-language/)

THANK

YOU!!