SOCIAL NETWORK ANALYSIS - MINI PROJECT REPORT

205001069
Nishaanth R

205001085
Sabarivasan V

205001086
Sai Shashaank R

205001097
Shajith Hameed

Analysis of social network - The Crisis Journal network

1. INTRODUCTION

W. E. B. Du Bois founded The Crisis in 1910, the official magazine of the NAACP, covering issues of race and social injustice. It was highly influential during 1910-1922, addressing various aspects of African American life, including women's suffrage, education, children, labor, homes, vacations, and the war. The magazine played a crucial role in the Harlem Renaissance, featuring renowned authors such as Du Bois, Jessie Fauset, Charles Chestnutt, Georgia Douglas Johnson, and Claude McKay. Started as a literary and arts movement, The Crisis continues to be published today.

2. ABOUT THE VISUALIZATION TOOL USED - “GEPHI”

Gephi is an open-source network analysis and visualization tool that simplifies the exploration of complex networks, including social networks. With its intuitive interface, users can import network data, apply various metrics and statistics, and visualize the network's structure and dynamics. Gephi's capabilities, including community detection and data filtering, make it a valuable resource for researchers, analysts, and data scientists seeking insights from network data.

3. ABOUT THE DATASET

The dataset shows a group of people connected to a magazine called "The Crisis," started by W. E. B. Du Bois in 1910. Each person involved, like Du Bois, Jessie Fauset, and Charles Chestnutt, is shown as a node, and the lines between them indicate their relationships. This helps us see who worked together, influenced each other, or had conflicts from 1910 to 1922. The dataset is useful for studying how people in this magazine were connected, understanding their collaborations, and finding out about important events.

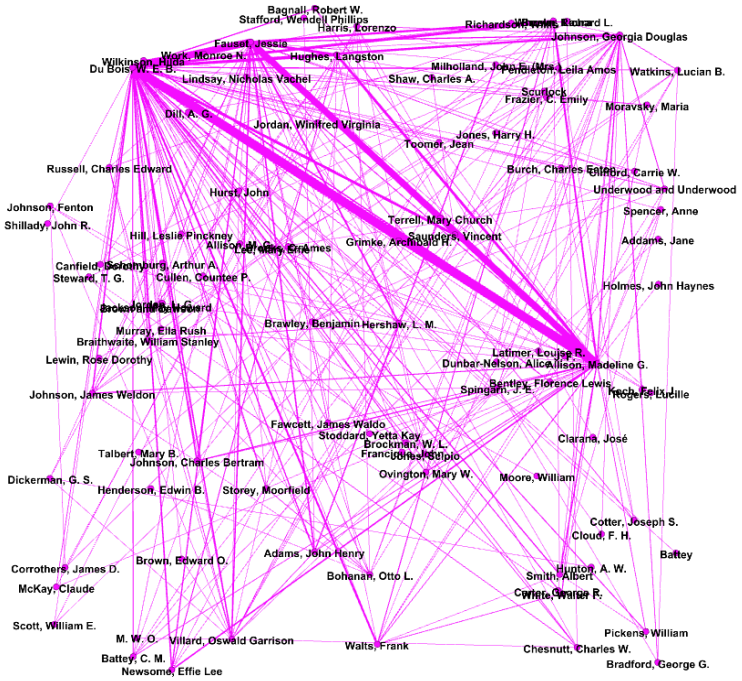
Statistics:

- Number of Nodes : 96
- Number of Edges : 273; weighted; undirected

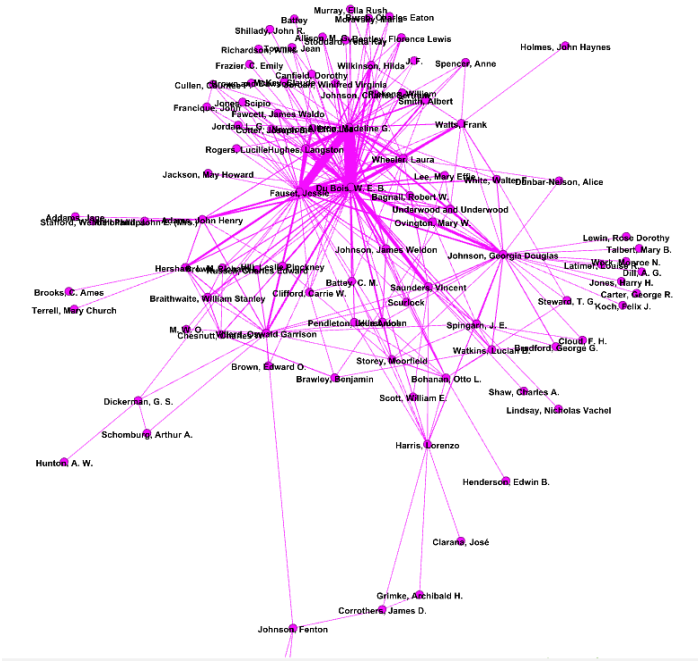
4. ANALYSIS

a) VISUAL REPRESENTATIONS OF THE NETWORK

i) Random Layout

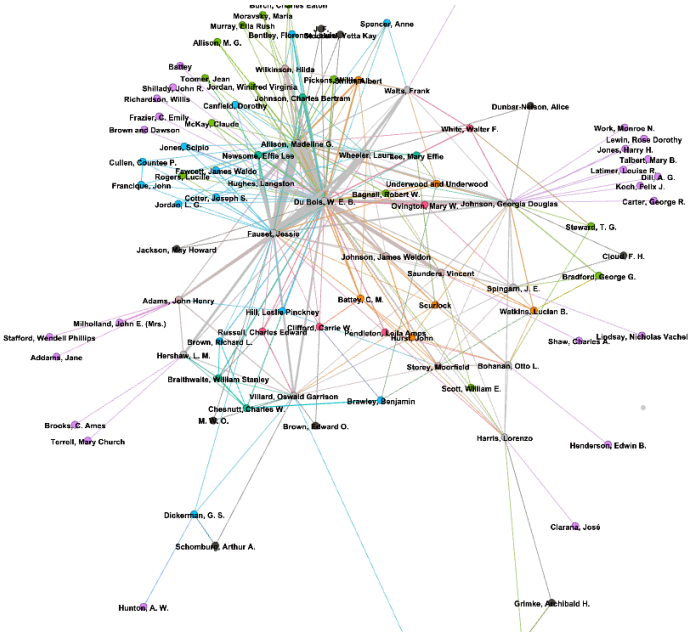


ii) Force Directed Layout: Force Atlas



b) DEGREE, DEGREE CENTRALITY

Degree centrality is a measure in network analysis that quantifies a node's importance by counting its direct connections (degree) within the network. The below graph represents the nodes with colors based on their degree value.



Inference: From the graph, we can tell that the biggest and boldest nodes are the ones that are popular in the network. In this network, a major international humanitarian organization is actively involved in disaster relief efforts. This node "HumanitarianOrg" exhibits high degree centrality.It is directly connected to a multitude of other nodes, representing partnerships, collaborations, and joint initiatives with various NGOs and local organizations. Its high degree centrality signifies extensive collaboration in crisis response. The average degree of this network is 5.688. This means that the amount of average nodes that connect to each other is 5.688.

c) NETWORK DIAMETER

Network diameter is the maximum distance (shortest path) between any two nodes in a network. It represents the longest geodesic path in the network and is a measure of the network's overall size or reach.

Inference: The diameter of this network is 5. This means that the farthest node is parted by 5 nodes.

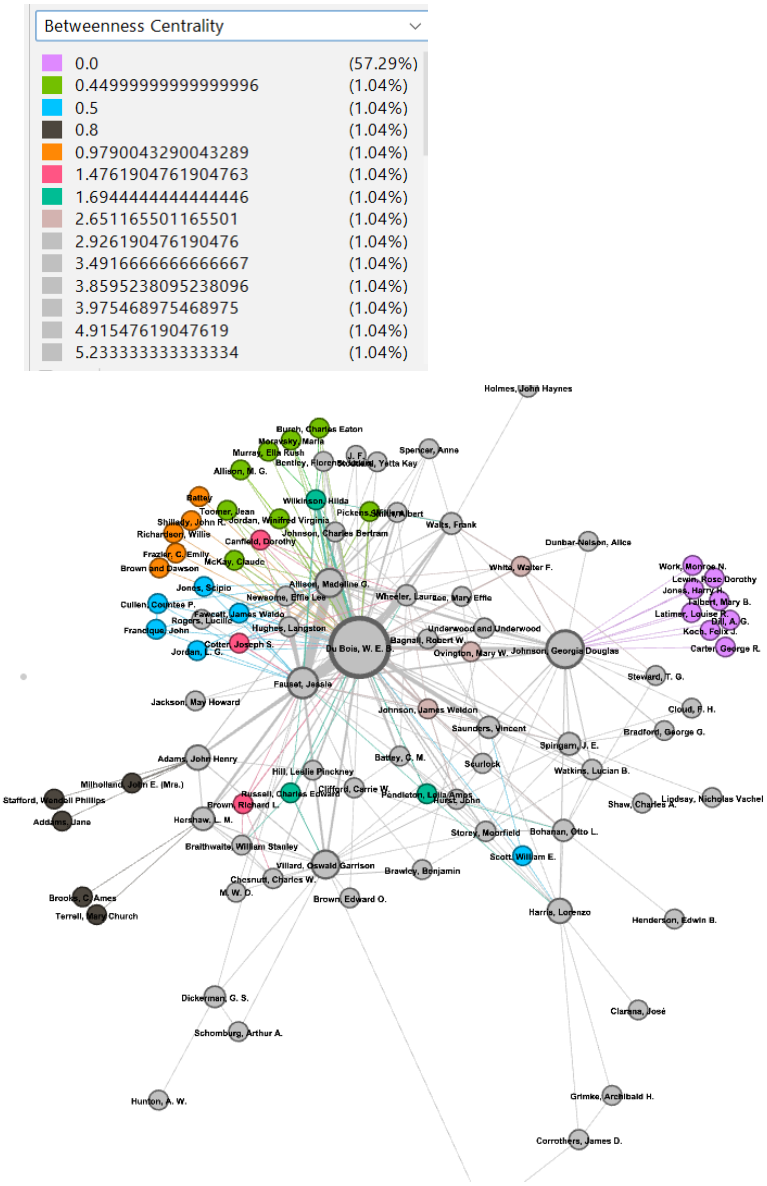
d) GRAPH DENSITY

Graph density is a network analysis metric that quantifies the proportion of actual connections in a network compared to the total possible connections. It reflects how interconnected the nodes are in the network.

Inference: For this network, the graph density is 0.06 which means this network is little dense, resulting in the closeness of the member.

e) BETWEENNESS CENTRALITY

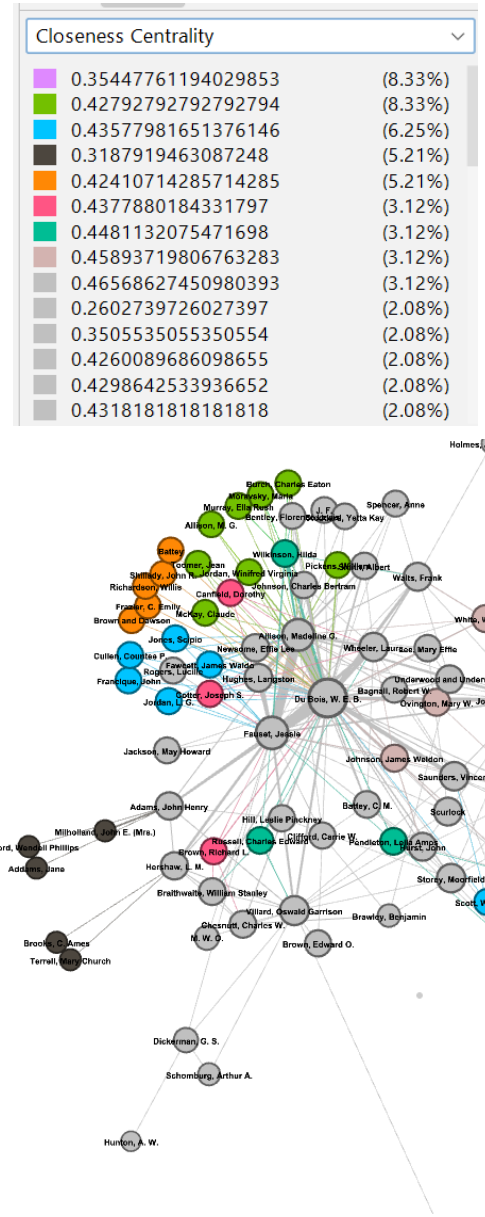
Betweenness centrality is a network analysis metric that quantifies a node's importance by measuring the number of shortest paths that pass through it. This measurement is used to determine which node is more likely to be the path of information of other nodes to connect with another node. Also to define where the communication will break apart.



Inference: From the graph, we can see that the nodes with purple and other colors with different sizes are most likely to bridge the other nodes to communicate. Reporters like DuBois, Faucet, Georgia and Mandelin act as a bridge between different clusters or communities within the network. They frequently interact with various stakeholders, including government agencies, NGOs, and local communities, to gather and share real-time information about the crisis.

f) CLOSENESS CENTRALITY

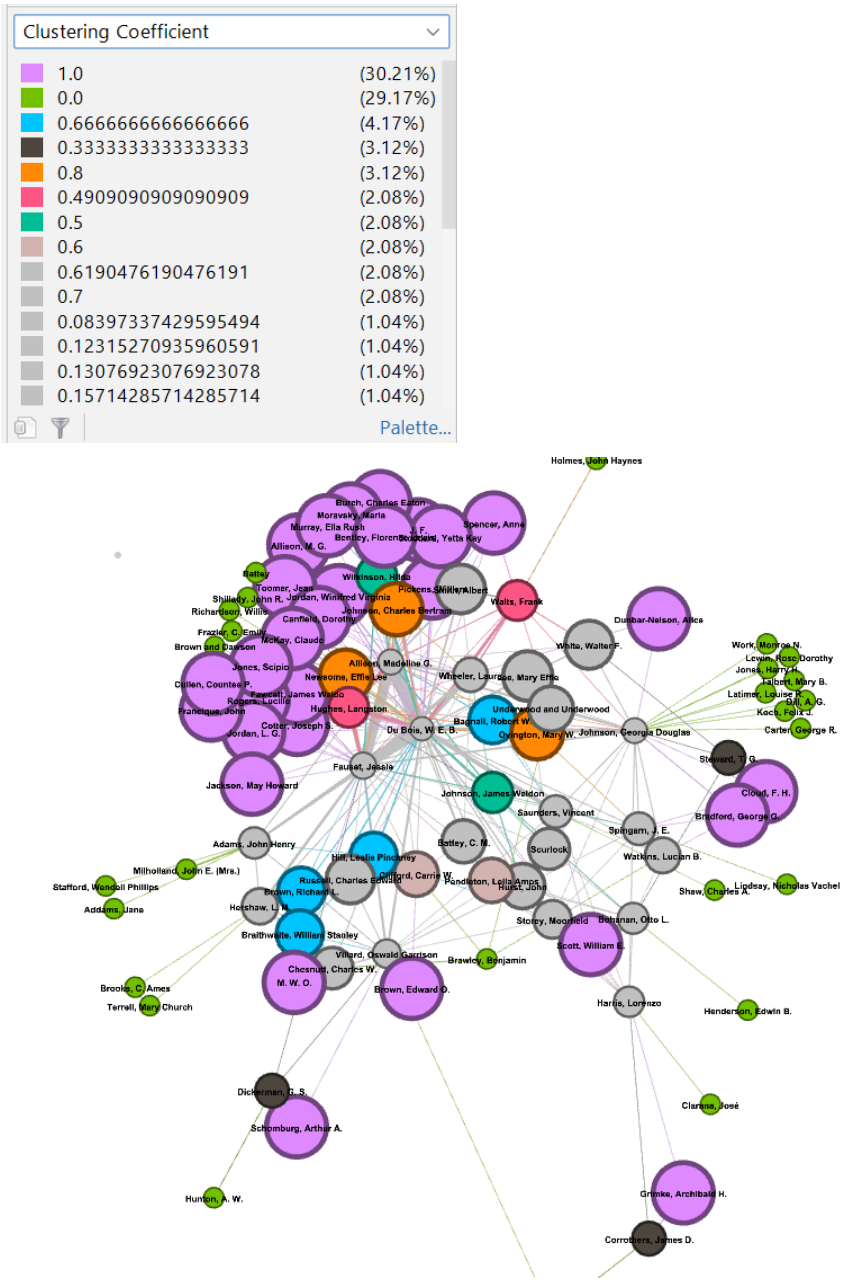
Closeness centrality is a network analysis metric that quantifies how quickly a node can reach all other nodes in the network. Nodes with higher closeness centrality are more centrally located and have shorter average distances to other nodes. This measurement is used to measure how easy is the node to reach (shortest path) other nodes.



Inference: From the graph, we can see that the nodes with similar colors are the fastest nodes to reach other nodes. Reporters working in government emergency response agencies are centrally located within the network, having short paths to a large portion of other nodes. This indicates that the government agency can quickly receive and disseminate information, enabling a rapid response to emerging crises or disaster situations.

g) CLUSTERING COEFFICIENT

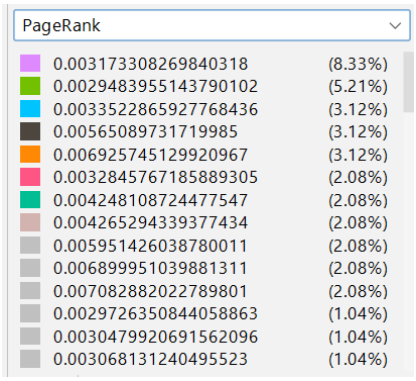
The clustering coefficient is a network metric that quantifies the degree to which nodes in a network tend to form tightly connected clusters or groups. It measures the likelihood that neighbors of a node are also connected to each other.

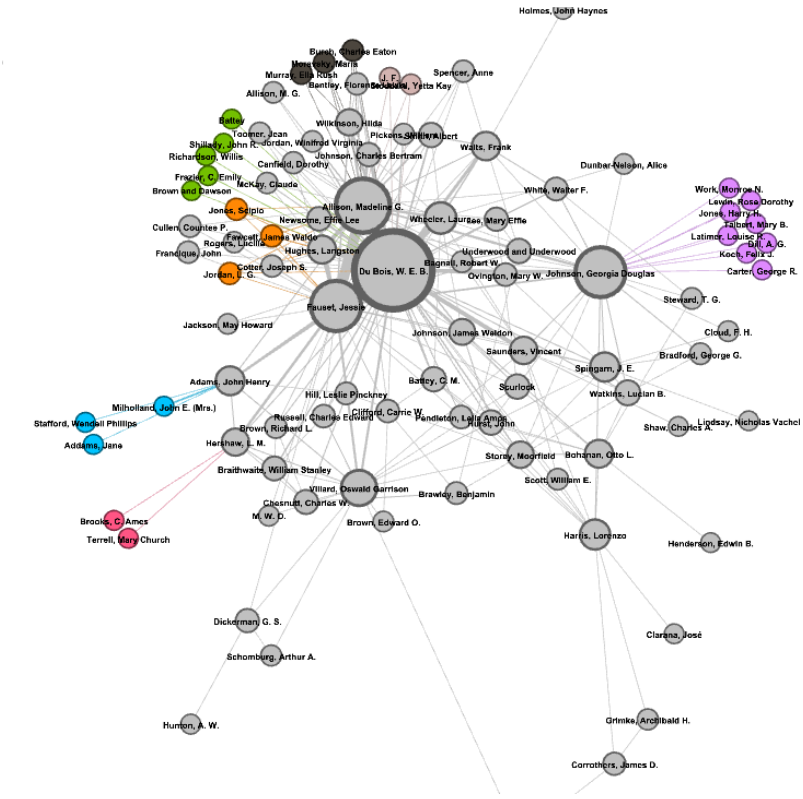


Inference: From the graph, we can see that the nodes with purple and other colors in which its neighbors are strongly interconnected, forming close-knit clusters or groups. The average clustering coefficient of this network is 0.681.

h) PAGE RANK

The metric assigns each node a probability that is the probability of being connected with the other node. The page rank values are the values in the eigenvector that have the highest corresponding eigenvalue of a normalized adjacency matrix A.
Parameters: Epsilon = 0.001 ; Probability = 0.85

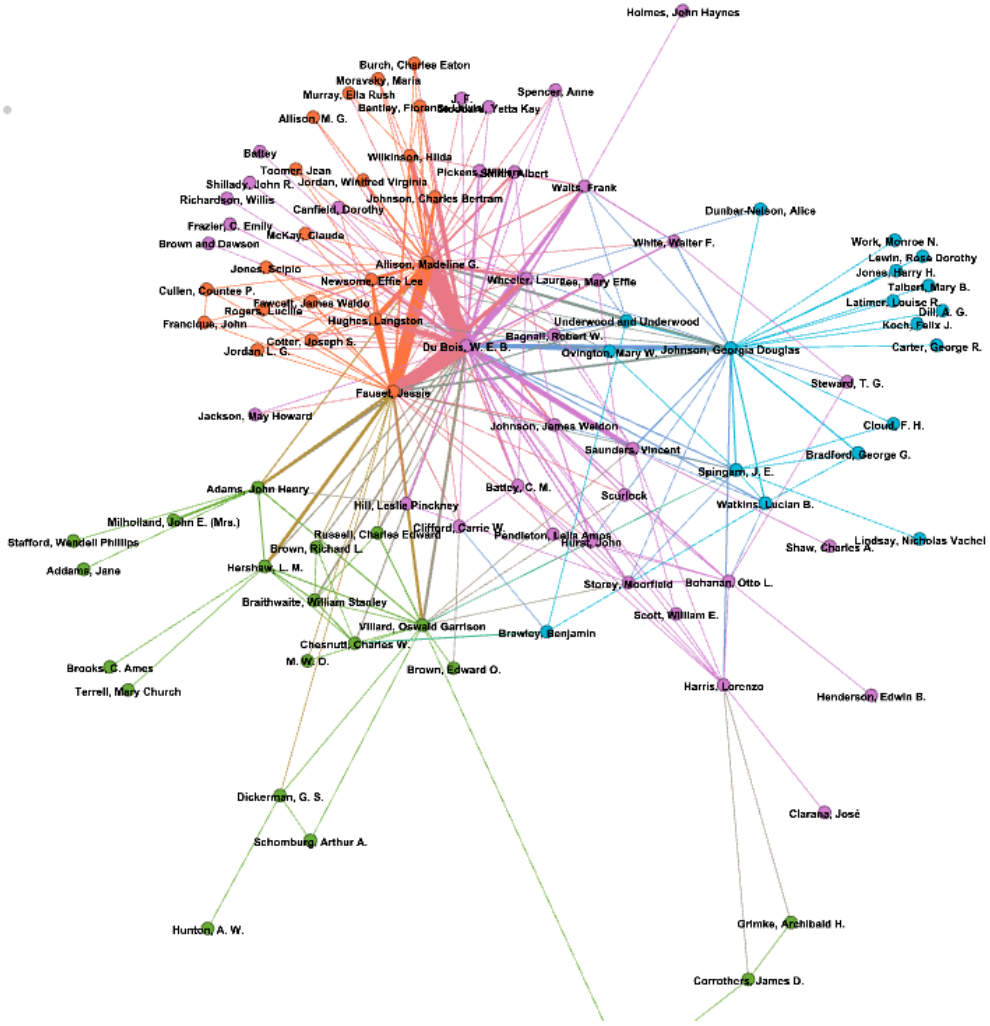




Inference: Humanitarian influencers with similar colors have a high PageRank in the network, indicating widespread influence similar to how important web pages are in the PageRank algorithm. This influencer or organization is recognized globally for its impactful contributions and leadership in humanitarian efforts. Due to its high PageRank, content shared by those humanitarian influencers, it is viewed as valuable and trustworthy. When this node shares updates, guidelines, or calls for action during a crisis, the information tends to reach a broad audience, fostering trust and credibility.

i) COMMUNITY DETECTION

The community detection algorithm used by Gephi is the “hierarchical cluster algorithm” which aims to identify the number of communities in the network for different settings provided. Gephi also provides a feature to colorize nodes to visualize the communities better.



Inference: Different communities have been detected with the help of varying colors by applying modularity and statistical inference which is used for community detection.

5. CONCLUSION

In conclusion, the analysis of The Crisis Network has provided some valuable insights. By exploring metrics like degree centrality, betweenness centrality, and community structure, we've gained a deeper understanding of the roles and relationships within the network.This analysis highlights the potential for applying social network analysis to uncover hidden patterns and inform decision-making in various social contexts.