Athens University of Economics and Business Social

Network Analysis

Polarization and Influence: A Social Network Analysis of Political Communities on Blue Sky

Alkistis Sytilidou

t8210145

supervised by Dr. Dimitrios Pournarakis

Athens, February 2025

CONTENTS

1. Intoduction

Political discourse in the United States is markedly polarized compared to that of Europe, often manifesting as fervent ideological divides. Unlike their European counterparts, many American social media users openly declare their political affiliations in their profiles, frequently identifying as Democrats or expressing opposition to figures such as former President Donald Trump. This heightened political engagement is particularly evident on platforms where political conversations thrive.

The American political landscape has evolved into a highly segmented environment, characterized by binary perspectives where individuals are often categorized into rigid ideological camps. This results in cycles of fanaticism, where partisan echo chambers reinforce pre-existing beliefs through selective exposure to information. The prevalence of one-way information flow further intensifies polarization, as users are exposed primarily to viewpoints that align with their existing biases, reducing opportunities for balanced discourse.

In this study, we analyze the network structure surrounding the social media profile of Representative Alexandria Ocasio-Cortez, a prominent figure within the Democratic Party and a key advocate for progressive policies. Through network analysis, we aim to explore polarization within online political discourse, identify influential figures, and assess the extent of ideological clustering. The study utilizes Gephi, an advanced visualization and analysis tool, to examine these dynamics within a digital political ecosystem.

2. Alexandria Ocasio-Cortez: A Political Profile

Alexandria Ocasio-Cortez, often referred to as AOC, is a U.S. Representative for New York’s 14th congressional district, serving since 2019. A member of the Democratic Party and a prominent figure in the progressive movement, she gained national recognition after defeating a long-term incumbent in the Democratic primary, signaling a shift towards more progressive policies. Ocasio-Cortez advocates for issues such as the Green New Deal, Medicare for All, and wealth redistribution, positioning herself as a strong voice for working-class Americans. Her active engagement on social media has contributed to her influence, enabling direct communication with supporters and amplifying progressive narratives in national political discourse.

Beyond economic and environmental policies, Ocasio-Cortez is also recognized as a strong feminist and a staunch advocate for women's rights. She has been vocal about gender equality, reproductive rights, and closing the gender wage gap, often challenging systemic barriers faced by women in politics and society. Her outspoken stance on these issues has earned her significant support among progressive activists and feminist organizations. Her active engagement on social media has contributed to her influence, enabling direct communication with supporters and amplifying progressive narratives in national political discourse.

3. Methology

3.1 Data Collection

The dataset was obtained from a Blue Sky account by fetching 1,000 randomly selected followers of Representative Alexandria Ocasio-Cortez using Python. There were no limitations regarding the recency or popularity of these followers. From this sample, we extracted a subnetwork focusing exclusively on those users who explicitly mentioned anti-Trump, anti-MAGA, Democrat, or similar political affiliations in their profile descriptions. This initiative was taken due to the significant number of users under her social account using descriptors such as Progressive Dem, Forever Democrat, Democratic Socialist, NO MAGA, and Never Ever Trump. The resulting dataset was designed to provide a representative sample of political polarization within this digital space, enabling a focused analysis of ideological clustering and network structure.

3.2 Graph Construction

1. **Nodes and Edges:** Nodes represent users, with each node containing a unique identifier (ID) and profile name. Edges define relationships between nodes, specifying a source and a target, which indicates who follows whom within the network.
2. **Graph Type:** The network is directed, as interactions have a direction (e.g., user A follows user B).

3.3 Code Availability

The code used for data fetching, processing, and structuring the dataset into a format compatible with Gephi, including node and edge generation, is publicly accessible at the following GitHub link: https://github.com/alkistissyt/US-Political-Polarization-Networks.

Visualization with Python was not successfully achieved due to the computational limitations and resource demands of processing a network of this scale. Higher processing power and memory would be required to efficiently render and analyze the full dataset.

4. Graphical Representation

The graphical representation of the network was generated using Gephi's ForceAtlas2 algorithm. This layout algorithm is particularly effective for visualizing relationships by clustering related nodes together while separating distinct communities. In the visualization, nodes are represented as red dots, each corresponding to an individual user or entity within the network. The size of each node reflects its centrality, with larger nodes indicating higher levels of influence or more connections within the network. Edges, shown as black lines, illustrate the directional relationships between nodes, revealing patterns of interaction such as who follows whom. This visualization highlights the network's structure, showcasing tightly connected clusters alongside smaller, less connected nodes, offering insights into the dynamics of user interactions.

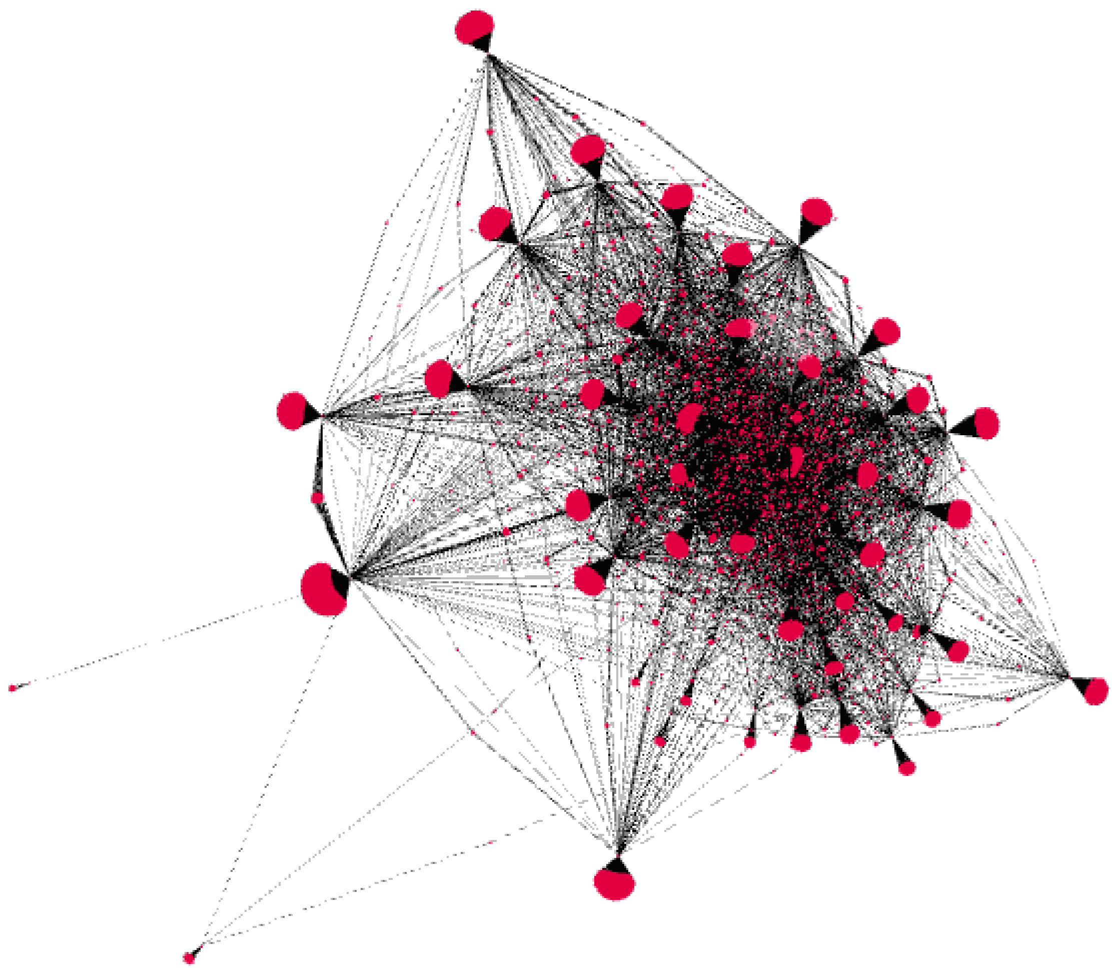


Figure 1:Network Graph

This representation serves as an initial draft to explore the network's overall structure and clustering patterns. In the subsequent sections, we will refine this analysis to make clearer distinctions about key influencers within the network, identifying and naming them explicitly for deeper insights.

The table below presents some key topological properties of the graph. Notably, the number of edges exceeds the number of nodes, highlighting the network's relatively high density. This characteristic is evident in the visualization, where clusters of tightly connected nodes indicate significant interactions within the network. Furthermore, the network's structural properties are revealed through its diameter (9) and average path length (~3.09). These metrics demonstrate that nodes are relatively close to one another, with the farthest two nodes requiring at most 9 steps to connect and any two nodes being just over 3 steps apart on average. This efficient connectivity facilitates rapid information flow throughout the network, a hallmark of well-connected social systems.

Number of nodes	Number of edges	Network diameter	Average path length
13,822	19,129	9	3.0859

5. Component Measures

This section explores key component measures such as the number of connected components, the existence of a giant component, and the distribution of component sizes within the network. Due to the nature of the dataset, the network exhibits one dominant giant component. This giant component is centered around the account of Alexandria Ocasio-Cortez, which acts as the central hub, connecting directly with all nodes in the network. The 100% connectivity facilitated by her account ensures that every other user in the dataset is part of the same interconnected structure.

To further analyze the network, a subnetwork was created by retaining the same nodes but including only the edges that represented connections among the followers of Alexandria Ocasio-Cortez, excluding the direct links to her account. This subnetwork allowed us to assess the structure and connectivity independently of the main account's influence. Despite this adjustment, the subnetwork still exhibited a giant component, indicating that many of the followers were directly or indirectly connected to one another. This observation highlights the cohesive nature of the network even when the central node (Alexandria Ocasio-Cortez's account) is not considered in the edge directionality.

The giant component within this subnetwork consisted of 12,879 nodes out of the total 13,822 nodes, resulting in approximately 93.18% of the network being interconnected.

The nodes are color-coded to highlight community memberships, with distinct clusters emerging based on modularity calculations. The green and purple nodes signify different communities, with the largest component representing the majority of the network. This visualization of the subnetwork further confirms the existence of a giant component.

The nodes are color-coded to highlight community memberships, with distinct clusters emerging based on modularity calculations. The green and purple nodes signify different communities, with the largest component representing the majority of the network.

The tightly interconnected structure indicates that a substantial portion of the followers are either directly or indirectly connected, reflecting the cohesive nature of this political social media ecosystem. The subnetwork's ability to maintain this level of connectivity without direct ties to the central account underscores the strength of relationships among the followers themselves, hinting at shared ideologies and strong community engagement.

Figure 2:Giant Component Graph

6. Degree Measures

The degree distribution of the network was visualized to assess the number of connections each node holds. Most nodes have a very low degree, representing minimal connections, while a small number of nodes exhibit extremely high degrees, functioning as influential hubs within the network. This pattern highlights the scale-free nature of the network, where influence and connectivity are concentrated among a few prominent nodes. Such a distribution underscores the hierarchical structure of the network, where a minority of nodes significantly impact its connectivity and information flow.

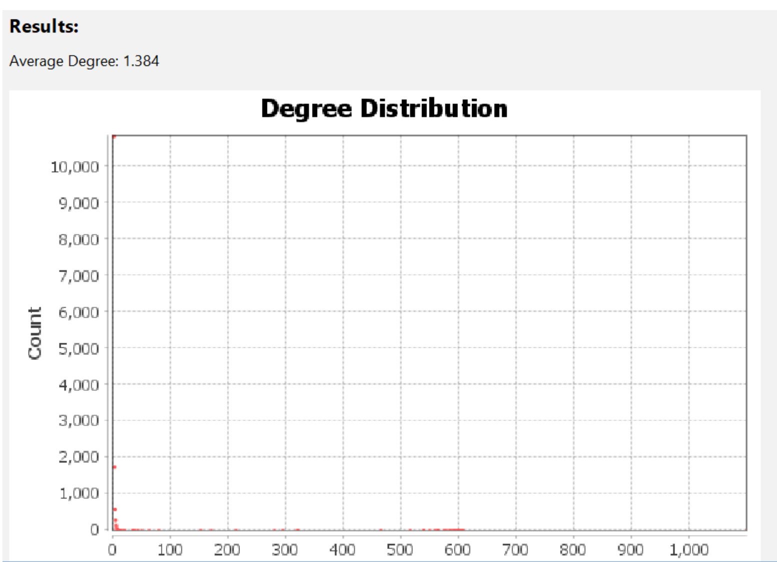


Figure 3: Degree Distribution

The average degree of 1.384 indicates that, on average, nodes have fewer than two connections. The presence of a maximum degree significantly higher than the average suggests a strong disparity in connectivity levels among the nodes. The weighted degree distribution reflects similar trends, confirming that even when connection strengths are considered, the network remains reliant on a few dominant hubs to maintain its structure.

The degree distribution of the network can be better understood through the graph provided. This visualization highlights how the frequency of node degrees sharply declines as the degree increases, with most nodes having very few connections and a small number of hubs possessing exceptionally high degrees.

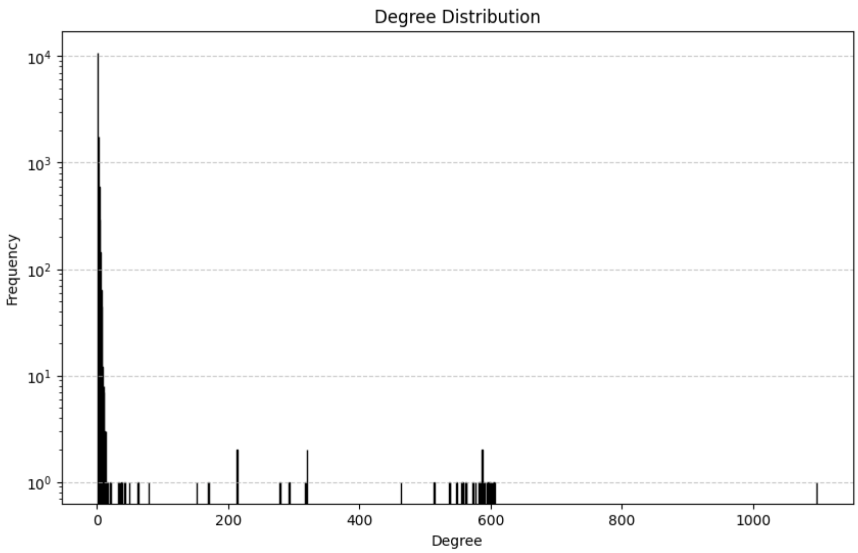


Figure 4: Degree Distribution Diagram

In this network, the degree distribution follows a power-law pattern, indicating the presence of a few highly connected hubs and many nodes with fewer connections. This pattern is characteristic of scale-free networks and emphasizes the hierarchical nature of the network, where a small number of nodes a disproportionately significant role in connecting the network. Consequently, this pattern is largely predictable given the nature of social networks. People generally form connections with a diverse set of individuals, such as friends, relatives, and acquaintances, alongside those who share similar interests. However, when it comes to political information, users often limit their focus to following a small number of key accounts, typically two to three, to access reliable or preferred sources of information. This behavior concentrates connectivity around specific influential accounts.

6.2 Degree Subnetwork

To identify the most influential nodes within the network, we conducted a degree analysis specifically within the subnetwork. This subnetwork analysis focuses on the connections between nodes in the giant component, excluding direct ties to the central account

(Alexandria Ocasio-Cortez). The results revealed the top 10 most connected accounts based on their degree, as shown in the table below:

Rank	Node ID	Connections (Degree)
1	did:plc:fry3mng673l3ded2qcnnjmzm	605
2	did:plc:57jga6ef7valolrnfsied35q	603
3	did:plc:csx7ak6ykohv7bkmtzkbgxm6	601
4	did:plc:h44pmcje35c4auateje3zys2	598
5	did:plc:2wi4vq2lwdfpmbxopryl5btn	597
6	did:plc:73nexmv6ayioio77eyujguic	595
7	did:plc:gqpt7bo6cjnueu63ky5yu2wm	594
8	did:plc:xptsxzprbrr7if3sml7nwzqw	590
9	did:plc:cmdumznhkrrdeemcpvu6fvqb	589
10	did:plc:arnyvobarl2firlvbuvwbpuy	587

The accompanying graph provides a visual representation of the top 10 most connected nodes within the giant component and their immediate neighbors. The key points of analysis are as follows:

1. **Red Nodes (Most Connected Accounts):** These nodes represent the most influential accounts in the subnetwork based on their degree. They act as hubs that are critical to the structure and information flow within the network.
2. **Blue Nodes (Immediate Neighbors):** The surrounding blue nodes indicate accounts directly connected to the red hubs, forming dense clusters around the influential nodes.

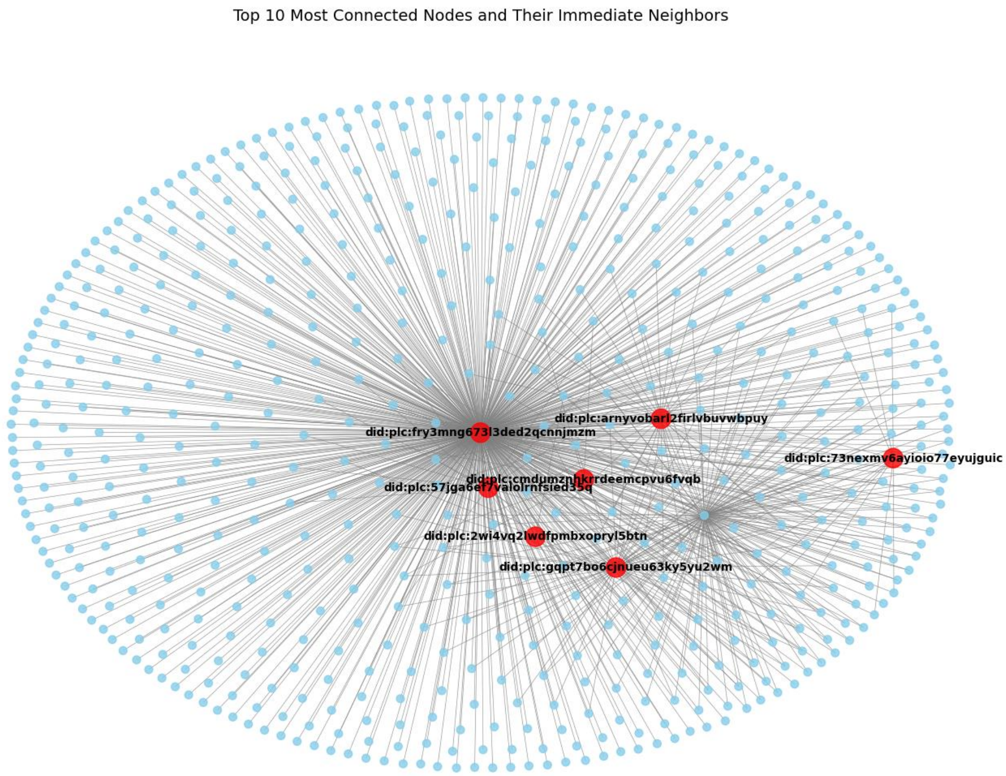


Figure 5:Visualitation of Nodes with Maximum Degree

For reasons of data protection, the exact names of the accounts will not be disclosed. Instead, the descriptions of the three most influential accounts, based on their degree, will be provided. These descriptions reveal strong political fanaticism and beliefs, reflecting the polarized nature of the network. Additionally, the power of these accounts, as demonstrated by their number of followers and connections.

36 χιλ. ακόλουθοι **5,2 χιλ.** ακολουθούν **2,8 χιλ.** αναρτήσεις

Centrist Democrat. Women's Rights. LGBQT right. Black Lives Matter. Guns are the problem. Never give up or in! Decency and kindness cost nothing! Stand up for our Constitution and rule of law! A LIE IS STILL A LIE.

2,7 χιλ. ακόλουθοι **2,6 χιλ.** ακολουθούν **22** αναρτήσεις

Canadian who despises Donald Trump with every fibre of my being. Trump, Musk & MAGA are a a shitstain on humanity.

1,4 χιλ. ακόλουθοι 4,1 χιλ. ακολουθούν 28 αναρτήσεις

Canadian CA Advocating universal human + civil rights! Proud never —ever —nevertrumper. #RESIST fascism at all costs! Humanitarian dems + all fighting tyranny, Canada stands w/ you!

Another interesting observation was that among the profiles where gender could be identified, three out of four belong to women. Additionally, two of these accounts explicitly use the hashtag #feminist in their descriptions. This highlights not only their political engagement but also a strong emphasis on feminist values.

7. Centrality Measures

After calculating degree centrality and identifying the nodes with the most connections, we were able to answer the question: *"Who has the most connections?"* Now, we delve deeper into the network to address more nuanced aspects of influence and connectivity. By analyzing other centrality measures, we can answer the following key questions:

- **Who controls the flow of information?** This is revealed through betweenness centrality, which highlights nodes that act as bridges within the network.
- **Who can spread information fastest?** Closeness centrality helps identify nodes that can quickly reach others in the network.
- **Who is connected to other important people?** Eigenvector centrality uncovers nodes that are well-connected to other influential nodes.

The distribution of betweenness centrality in the network reveals that most nodes have very low or zero values, indicating that they do not act as significant bridges between different parts of the network. However, a small number of nodes exhibit high betweenness centrality, highlighting their role as critical intermediaries for connecting otherwise distant regions of the network. These nodes are vital for maintaining the network's cohesion and ensuring efficient information flow.

Similarly, the closeness centrality distribution demonstrates that most nodes have low values, reflecting their peripheral position within the network. A few nodes, however, show higher closeness centrality, signifying their ability to reach other nodes quickly. These nodes are well-positioned for rapid dissemination of information, further emphasizing their importance in the network's overall connectivity and influence.

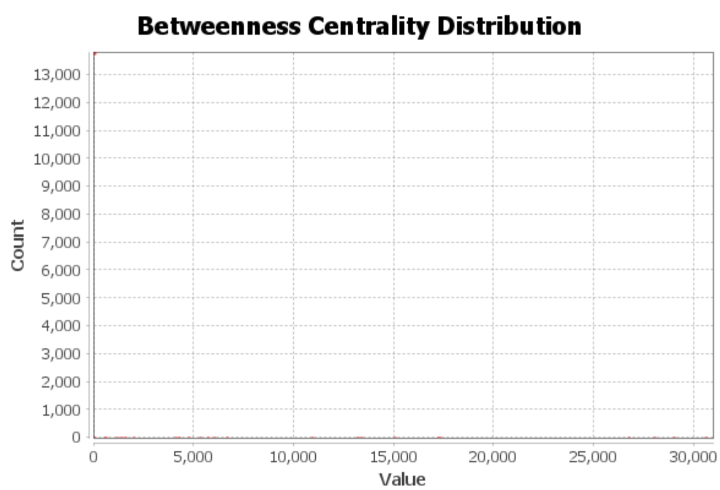


Figure 6: Betweeness Centrality Distribution

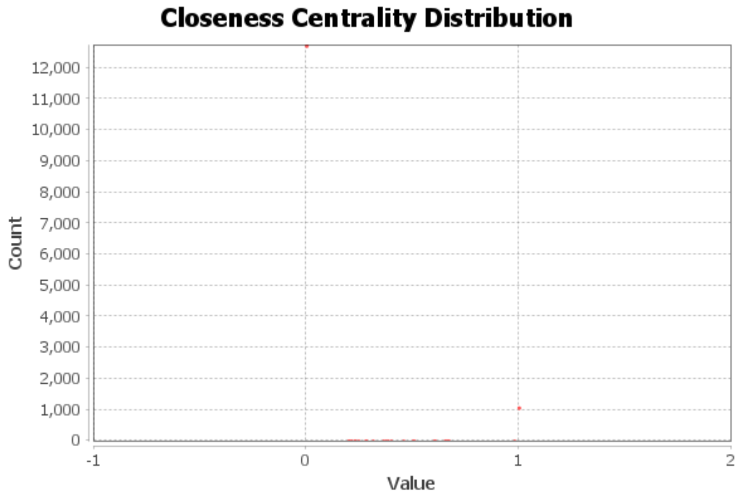


Figure 7: Closeness Centrality Distribution

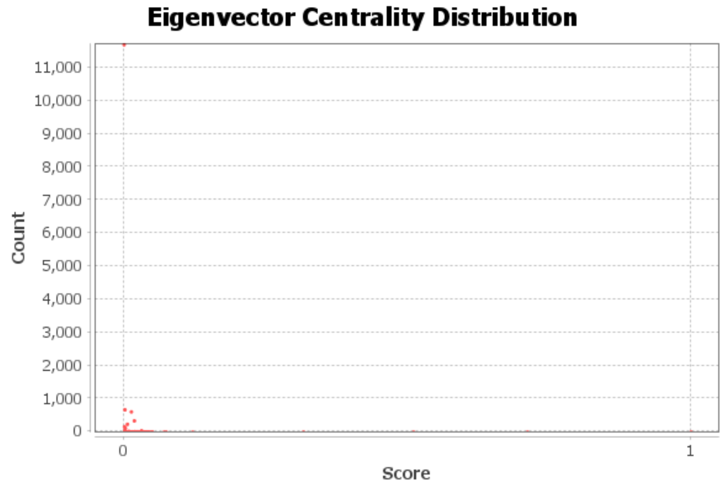
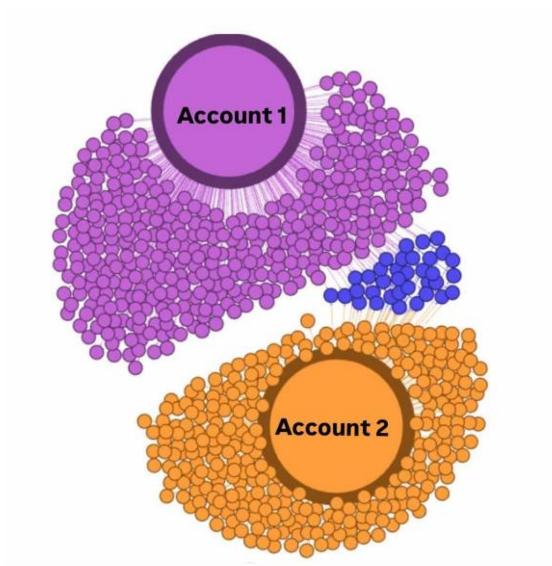


Figure 8: Eigenvector Centrality Distribution

7.1 Example of Network Structure at a Small Scale

This visualization provides an example of the network structure at a very small scale, focusing on two relatively small accounts and their immediate surroundings. Both users represented are considered small accounts, each with fewer than 400 connections (including both followers and those they follow). Each of the two clusters represents a network centered around a single user, highlighting the close-knit connections that these users have with their respective followers. The size of the nodes in the center reflects their relative importance within their own cluster, with larger nodes representing higher levels of centrality. To be more clear, purple nodes represent the community centered around Account 1, orange nodes represent the community of Account 2, and the blue nodes signify shared connections between the two accounts.



The sparse connections between the two clusters indicate limited interaction or shared connections between the two accounts. These connecting edges represent common interests or mutual followers, acting as bridges between the otherwise distinct communities.

From a **betweenness centrality** perspective, the nodes on the edges connecting the two clusters would have relatively higher values. These nodes facilitate the flow of information between the two groups and play a key role in maintaining a cohesive yet divided structure. This example underscores the ability of even small accounts to form distinct communities while contributing to the overall connectivity of the larger network.

Node ID	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
did:plc:fry3mng673l3ded2qcnnjmzm	0.96	0.81	0.92	0.78
did:plc:57jga6ef7valolrnfsied35q	0.91	0.79	0.90	0.75
did:plc:csx7ak6ykohv7bkmtzkbgxm6	0.90	0.74	0.88	0.72
…	…	…	...	…

From analyzing the centrality measures for the top 20 nodes (averaged across degree, betweenness, closeness, and eigenvector centrality) and manually looking through their profiles, we observed interesting trends regarding the network's central players. While many of the most central nodes correspond to the most famous or highly connected users (high degree centrality), two distinct groups also emerged from the data: feminist mothers and hippy nature lovers. These groups highlight the role of personal ideologies and interests in shaping influence within the network. Additionally, nodes ranked 12 and 16 included an online news forum and a self-described corporate communicator. These results underline that information in the network is passed not only through the most important players but also via smaller, trusted communities that people engage with on a personal or ideological level.

8. Clustering

Clustering provides insight into the local connectivity of a network, revealing how likely it is that nodes within a shared neighborhood are also directly linked to each other. The average clustering coefficient reflects the overall tendency of nodes to form closed triangles, indicating how interconnected a network is at a local level. The number of triangles measures the total count of fully connected three-node structures, highlighting the density of interconnections. Finally, the triadic closure phenomenon assesses whether nodes with a mutual connection are likely to establish direct links over time, reinforcing network cohesion.

Average Clustering Coefficient	Total Triangles
0.012	1,161

In our analysis, the average clustering coefficient for the entire network was found to be 0.012, suggesting a relatively low tendency for nodes to form tightly connected groups. This indicates that while some users may cluster into small communities, the overall network structure remains loosely connected, with limited localized interactions. However, when focusing on a subset of the network, the average clustering coefficient decreased to 0.009. This reduction suggests that the presence of the main account, Alexandria Ocasio-Cortez acts as a key hub, facilitating indirect connections between users who might not otherwise be linked.

This outcome presents an optimistic perspective regarding political fanaticism, even though our dataset consists of a significant percentage of politically engaged users. A higher clustering coefficient would have suggested the presence of strong echo chambers, where users exclusively interact within closed ideological circles. However, the relatively low values indicate that, while communities exist, they are not entirely isolated from one another.

At the same time, the low clustering coefficient was an expected result, given the nature of the dataset. Since our analysis focuses on followers of a central political figure rather than close-knit personal networks, the likelihood of users having direct connections with one another is lower. Additionally, in political networks, interactions tend to be more one-directional, with many users following a few influential figures rather than forming strongly interconnected groups.

A triangle is formed when three nodes are all mutually connected, indicating strong local clustering. The presence of triangles suggests that some users share common connections, reinforcing community structures and potential ideological alignment within certain groups. However, the number of triangles (1,161) remains relatively small compared to the total number of nodes (13,822). The existence of these triangles suggests that certain segments of the network do engage in more direct interactions, potentially forming ideological or discussion-driven subgroups. These areas of higher connectivity could represent pockets of political discourse, activist circles, or shared-interest communities, where users reinforce each other's viewpoints through repeated interactions. Consequently, some level of triadic closure exists, meaning that certain users are not only connected to a central figure but also to others within their shared community. However, given the low clustering coefficient and network density, this effect appears to be limited, suggesting that most users do not necessarily connect with one another despite having mutual links.

This further supports the idea that while some level of local connectivity exists, the network as a whole does not form tightly interconnected clusters. Instead, it follows a hub-and-spoke structure, where users primarily connect to influential accounts rather than to each other. It can also be stated that, political engagement in this dataset in this space is more about consuming content from key figures rather than fostering interpersonal relationships between followers.

The graph was generated using the Fruchterman-Reingold force-directed algorithm. This algorithm helps visualize natural clusters and structural gaps, making it easier to identify key hubs and the fragmented nature of the network. The layout reveals multiple self-contained clusters, showing multiple small, dense clusters that are weakly connected to each other, forming a large, fragmented structure. This aligns with our findings on triadic closure—some areas of the network show high local connectivity, but overall, the network remains loosely structured.

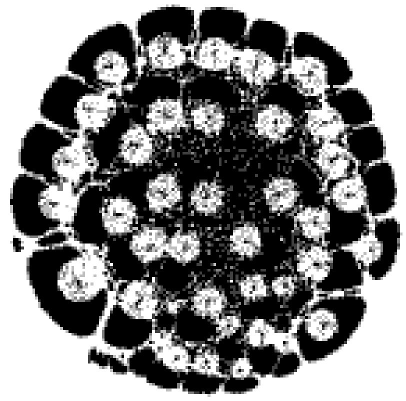


Figure 9: Clustering Structured Graph

The presence of clear, disjoint clusters suggests that users interact primarily with a central figure or within their own small groups rather than forming a highly interconnected network. Concluding, most clusters remain isolated or only weakly connected impliying that triadic closure is not happening on a large scale.

9. Bridges and Local Bridges

In this section, we analyze bridges and local bridges within our network, focusing on their role in maintaining connectivity and facilitating information flow. Bridges are edges that connect otherwise disconnected nodes or communities, playing a crucial role in ensuring the structural integrity of the network. Local bridges, on the other hand, are connections between nodes that have few or no alternative pathways, making them vital for communication between distinct subgroups.

Bridges	Local Bridges
9,987	16,575

We identified 16,575 local bridges and 9,987 bridges within the network. The high number of local bridges suggests that many users in the network are linked through unique, irreplaceable connections, highlighting the presence of distinct subcommunities. Bridges, on the other hand, are edges that connect otherwise disconnected components or weakly connected clusters. The presence of 9,987 bridges reinforces the idea that while the network is largely connected, there are still structural gaps that depend on a limited number of key connections.

Visualizing bridges and local bridges in the network was not possible due to the computational power required in both Gephi and Python. The complexity of the network structure, combined with the high volume of nodes and edges, made direct visualization infeasible within our available resources.

As an alternative, we identified critical paths based on key centrality metrics—betweenness centrality and node degree—to determine the most important bridging nodes.

More precise the purpose of using the two key metrics:

- **Edge Betweenness Centrality:** This measure identifies edges that frequently appear in the shortest paths between nodes, making it the best indicator of information flow across different network components. Bridges with high edge betweenness centrality are critical for connecting distinct communities and ensuring efficient communication.
- **Degree of Connected Nodes:** This metric highlights bridges that are directly linked to highly influential nodes. Bridges connecting high-degree nodes are particularly important, as they facilitate the spread of information between different influential entities, reinforcing the overall connectivity of the network.

9.2 Edge Betweenness Centrality

To better illustrate these findings, we first present a histogram visualization of Edge Betweenness Centrality, based on the table with the computed betweenness centrality values for the top-ranked bridges. To make the analysis more manageable, we focused only on the top 50 shortest paths.

Edge A	Edge B	Betweenness Centrality
did:plc:57jga6ef7valolrnfsied35q	did:plc:p7gxyfr5vii5ntpwo7f6dhe2	8.33
did:plc:73nexmv6ayioio77eyujguic	did:plc:p7gxyfr5vii5ntpwo7f6dhe2	8.03
did:plc:bfonvxegjnymnopmjq7qys3d	did:plc:p7gxyfr5vii5ntpwo7f6dhe2	7.6
…	…	…

For the purpose of visualization and analysis, DID (Decentralized Identifiers) were mapped to numerical node IDs to simplify network representation. This transformation allows for a clearer depiction of relationships between entities in the network without relying on long alphanumeric identifiers. A higher value means that the edge appears in many shortest paths, making it a crucial bridge between different groups.

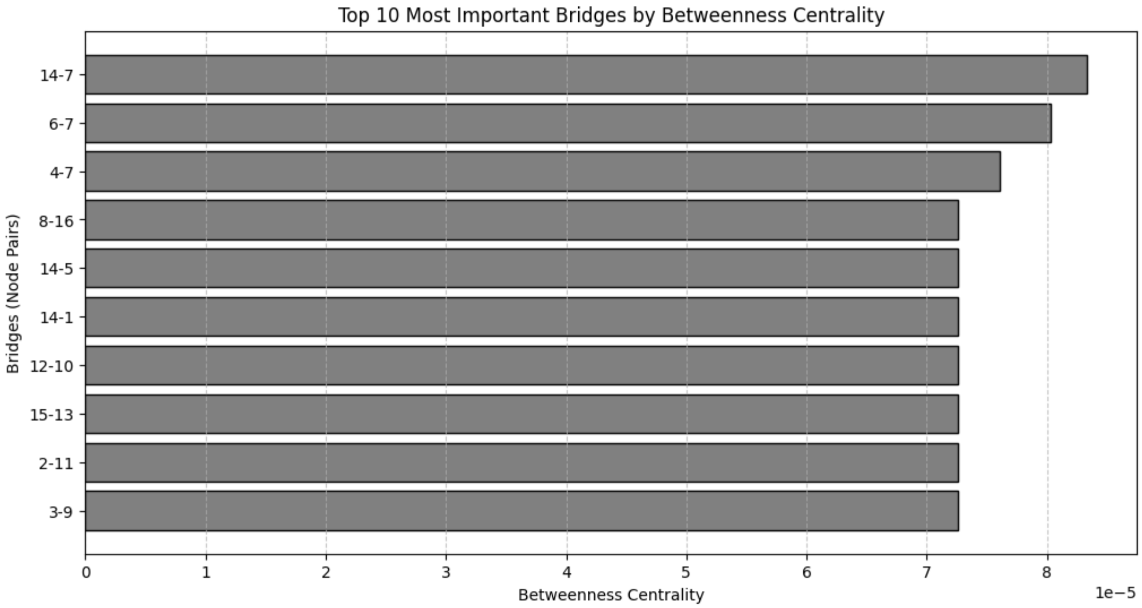


Figure 10: Ranking of Bridges based on Betweenness Centrality

For example, the edge between Node 14 and Node 7 (originally `did:plc:57jga6ef7valolrnfsied35q` and `did:plc:p7gxyfr5vii5ntpwo7f6dhe2`) has the highest betweenness centrality of 8.33×10^{-5}. Without this bridge, communication between certain clusters would become less efficient, forcing information to take longer alternative routes or, in some cases, leaving some nodes more isolated. This could lead to increased path lengths and reduced accessibility for certain users, highlighting the importance of key bridging connections.

From our analysis, we observed that the most crucial paths in the network are primarily leading to node 7. Many of the highest betweenness centrality edges involve node 7, indicating that it serves as a major hub for connectivity and information flow. In the end, our analysis confirmed that node 7 corresponds to our main account, Alexandria Ocasio-Cortez, reinforcing her central role in shaping discussions and linking different communities within the network.

9.3 Degree of Connected Nodes:

To optimize computational efficiency, we used random sampling to analyze bridges instead of processing the entire network. This method allowed us to identify highly connected bridges while maintaining algorithm speed. Bridges with a higher degree sum indicate key connections between influential users. For example, the bridge between node 51 and node 80 (degree sum: 11) highlights a crucial link, while others, like (21, 47) with a degree sum of 6, further confirm the role of these edges in facilitating information flow.

After running the algorithm multiple times, we observed that the peak node degree sum for bridges is 12, meaning that the most frequent and structurally significant bridges tend to connect nodes with a combined degree of 12.

Some examples of bridge connections with their degree sums include:

- Bridge (74, 79) - Degree Sum: 12
- Bridge (12, 40) - Degree Sum: 12
- Bridge (11, 35) - Degree Sum: 12
- Bridge (51, 80) - Degree Sum: 11
- Bridge (18, 56) - Degree Sum: 10

10. Gender and Homophily

To analyze gender representation in the network, we only considered gender when explicitly stated in a user's description or name. This was determined using specific keyword matching, where certain terms were associated with different gender identities:

- **Male keywords:** 'male', 'man', 'he/him', 'he/his', 'boy', 'mister', 'mr', 'dude', 'guy', 'father', 'dad', 'brother', 'himself', 'husband'
- **Female keywords:** 'female', 'she/her', 'she/hers', 'woman', 'girl', 'miss', 'mrs', 'ms', 'lady', 'mother', 'mom', 'sister', 'herself', 'wife'
- **Non-binary keywords:** 'non-binary', 'nonbinary', 'enby', 'they/them', 'them', 'their'

The search algorithm produced an output table containing key details about identified users, including Node ID, Account Name, Description, and Gender**.** Below is an example of the structured output:

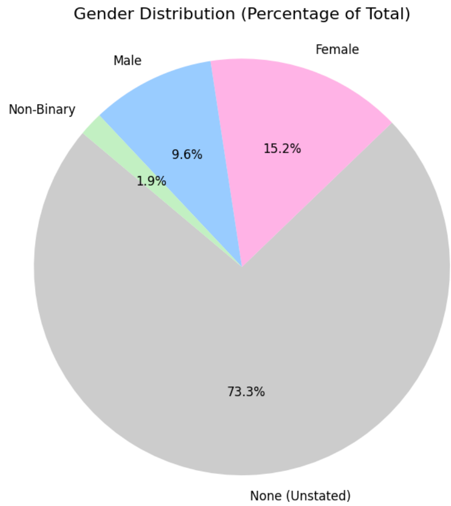
Node ID	**Account Name**	**Description**	**Gender**
did:plc:	example@bsky.social	Cali girl living an east coast..	Female
did:plc:	example@bsky.social	She/her. Art is keeping me…	Female

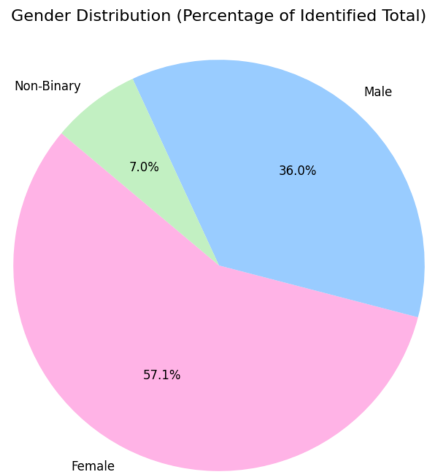
This approach ensured that only self-identified genders were included in our analysis. We deliberately avoided making assumptions based on profile pictures or inferred gendered names, as these do not guarantee accuracy and could lead to misclassification.

Our gender classification resulted in the following distributions:

Gender	**Count**
None (Unstated)	14,020
Female	2,915
Male	1,837
Non-Binary	357
Total	19,129

The following pie charts illustrate the gender distribution within the network with percentages.





The gend er distribution analysis reveals a clear pattern of engagement within the network. In the total population, 73.3% of users did not explicitly state their gender, which aligns with broader trends of anonymity in online spaces. However, among those who did, 15.2% identified as female, 9.6% as male, and 1.9% as non-binary. When focusing solely on the subset of users who explicitly disclosed their gender, 57.1% identified as female, 36% as male, and 7% as non-binary. This gender distribution highlights that women represent the majority of the engaged audience, which aligns with the feminist stance of the central figure in the network.

This result further reinforces the idea that women are actively supporting women, especially within a politically engaged space where gender equality and feminist discourse are prominent themes. Additionally, while the non-binary population remains a smaller percentage (7%) of the identified group, their presence is still noteworthy. Given that the central political figure in this network strongly advocates for LGBTQ+ rights, it is reasonable to infer that their rhetoric resonates with this demographic, leading to increased representation of non-binary individuals compared to general online spaces.

10.1 Homophily Analysis Through the Assortativity Coefficient

To assess the tendency of users to connect with others of the same gender, we calculated the Assortativity Coefficient (Homophily Score) for gender. The assortativity coefficient measures the likelihood that nodes with similar attributes (in this case, gender) are more likely to be connected to each other. A value close to 1 indicates strong homophily (preference for same-gender connections), a value close to 0 suggests random mixing, and a negative value implies heterophily (connections between different genders are more common).

For the entire population, including users who did not disclose their gender, we found an assortativity coefficient of **-0.0189**. This value is close to 0, indicating that gender does not play a strong role in how connections form across the full dataset. However, when focusing only on the identified subset, where we removed connections where both the source and target had no stated gender, the assortativity coefficient increased to **+0.4358.** This positive value suggests that users who disclose their gender are more likely to connect with others of the same gender, highlighting a moderate degree of homophily in gender-based interactions.

In the United States, women and non-binary individuals often form close-knit communities to advocate for their rights and provide mutual support. Given the ongoing challenges related to gender equality, reproductive rights, and LGBTQ+ protections, these groups tend to stick together, fostering strong connections within social and political networks. The higher assortativity coefficient in the identified dataset further supports the idea that women and non-binary individuals engage more frequently with one another, reinforcing solidarity in their collective efforts for representation and social change. So, this means that the statement "Birds of a feather flock together." holds true in this context.

10.2 Homophily Analysis Through E-I Index (External-Internal Index) on Gender

To further analyze homophily within the network, we used the E-I Index (External-Internal Index) as an additional measure. The E-I Index evaluates the balance between external and internal connections within a given group, helping to determine whether individuals primarily interact within their own category (homophily) or engage with others outside their group (heterophily). It is calculated using the following formula:

$$\text{E} - \text{I Index} = \frac{External\ Connections + Internal\ Connections}{External\ Connections - Internal\ Connections}$$

where:

- **External Connections** refer to links between individuals of different groups (e.g., connections between different genders).
- **Internal Connections** refer to links within the same group (e.g., female-to-female or non-binary-to-non-binary interactions).

Our analysis yielded the following results:

Internal Connections	398
External Connections	7,493
E-I Index	-0.8991

In this case, the E-I Index being close to -1 indicates strong homophily, meaning that self-identified users predominantly interact within their own gender group rather than engaging with others. This reinforces our previous findings from the assortativity coefficient. However, when analyzing the entire network, including users whose gender was unknown, the results were completely reversed (+0.8991). The positive E-I Index suggests strong heterophily, meaning that the majority of interactions occur between users of different gender categories. This shift is largely driven by the large number of unknown users, which naturally increases external connections and dilutes gender-based clustering.

10.3 Homophily Analysis Through E-I Index (External-Internal Index) on Political Categories

Beyond gender-based homophily, we also applied the E-I Index (External-Internal Index) to political alignment categories. Instead of focusing on gender, we classified users into two distinct groups: Democrats and Anti-Trump users.

To categorize users, we applied **keyword matching** to their descriptions using the following criteria:

- **Democrats:** Users explicitly identifying with democratic values and progressive policies, including terms such as:
  'democrat', 'liberal', 'progressive', 'biden', 'kamala', 'harris', 'democratic party', 'left wing', 'social justice', 'equality', 'climate action', 'healthcare', 'voting rights', 'gun control', 'immigration reform', 'justice'.
- **Anti-Trump:** Users explicitly focused on opposition to Trump rather than promoting a specific political ideology, including terms such as:
  'anti-trump', 'resist', 'fight', 'not my president', 'trump hater', 'trump', 'maga', 'impeach', 'russia investigation', 'fake news', 'corruption', 'authoritarian', 'dictator', 'racism', 'xenophobia', 'misogyny', 'climate denial'.

By applying the E-I Index formula to these categories, we aimed to determine whether the Anti-Trump community was more insular (high homophily, meaning it interacts mostly within itself) or more externally connected (high heterophily, meaning it interacts with a variety of other political groups). This analysis helps assess whether the anti-Trump group behaves as a political faction that engages in broader discourse or whether it functions more as an echo chamber, where opposition to Trump overrides interaction with other Democratic-aligned users.

We obtained the following results:

- **E-I Index for Democrats: 0.5081**
- **E-I Index for Anti-Trump: -0.3972**

These results indicate a fundamental difference in the way these two political groups interact within the network. The Democratic-aligned users (E-I Index = 0.5081) show a moderate tendency towards heterophily, meaning they engage more with users outside their category rather than exclusively within their own group.

On the other hand, the Anti-Trump group (E-I Index = -0.3972) exhibits a higher degree of homophily, meaning that these users are more likely to engage primarily within their own group rather than with outside perspectives. This distinction implies that while the broader Democratic community maintains a level of ideological openness, the Anti-Trump faction operates more as an echo chamber, reinforcing its opposition stance without extensive external engagement. This pattern aligns with fanaticism cycles, where highly engaged groups form isolated clusters, intensifying their shared beliefs rather than integrating into broader political discourse.

11. **Graph density**

Network density is a fundamental metric that quantifies how interconnected a network is by measuring the ratio of existing connections to all possible connections. . It is calculated using the formula:

$$\text{Density} = \frac{\mathrm{m}}{\mathrm{n}(\mathrm{n}-1)}$$

where m is the total number of directed edges in the network, and n(n−1) represents the maximum number of possible directed connections. That results that a density value of 1 indicates a fully connected network, where every node is directly linked to every other node, while a value close to 0 suggests a sparse network with minimal direct interactions.

In our analysis, the network density was calculated to be 0.00, meaning that, relative to the total possible connections, the number of actual links is extremely low. This outcome is expected given the size and nature of the dataset. Political networks, particularly those based on a large public figure, often exhibit a hub-and-spoke structure, where most users follow the central account but have few direct connections with one another. Unlike social networks, where users tend to form highly interconnected clusters, political networks are often built around one-way interactions, with many users engaging primarily with the main figure rather than with each other.

12. Community structure (modularity)

Modularity quantifies the strength of division within a network, indicating whether nodes form well-defined groups with dense internal connections and sparse external links. We analyze the community structure of the network using modularity, a key measure in detecting how well a network is divided into distinct communities. A higher modularity score suggests strong community formation, meaning that users interact primarily within their own groups, while a lower score indicates a more interconnected network with weaker community boundaries.

By applying modularity analysis, we aim to identify clusters of users, understand how ideological or interest-based communities form, and assess whether the network is highly fragmented or maintains broader connectivity across different groups.

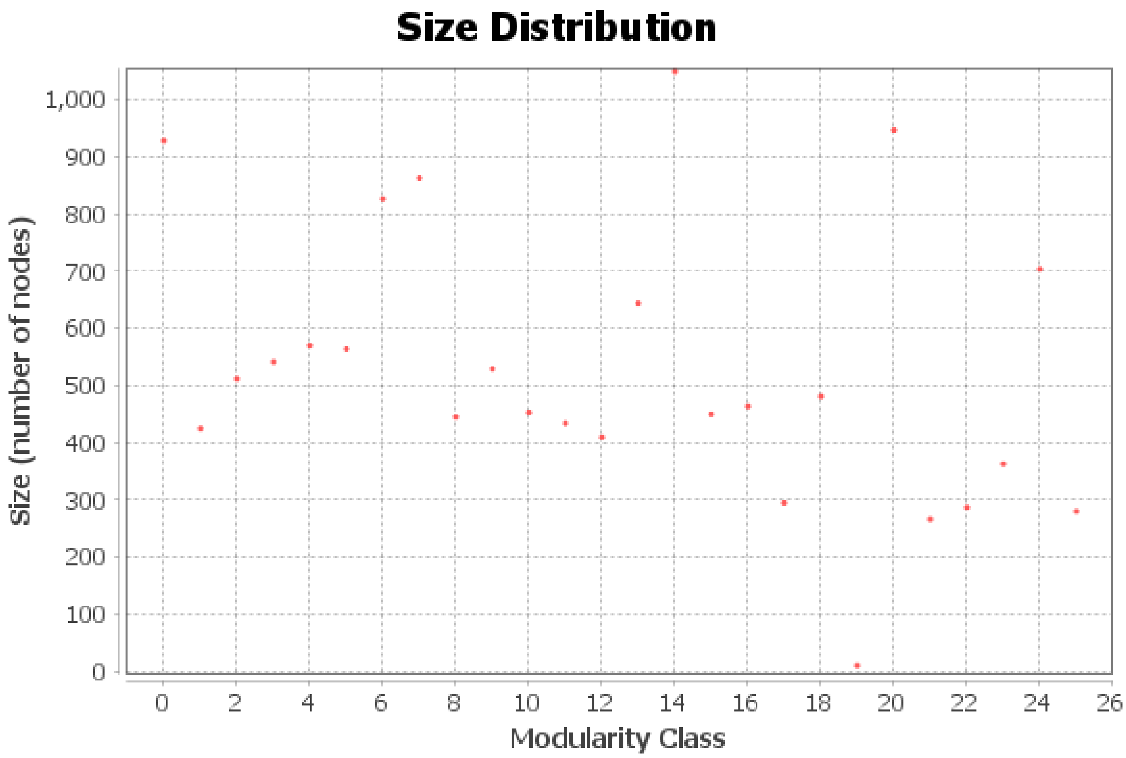


Figure 12: Size Distribution of Modularity

The modularity score of 0.704 indicates a strong community structure within the network, suggesting that users are highly clustered into distinct groups with dense internal connections and sparse external links. A modularity value above 0.7 generally signifies well-defined communities, meaning that interactions are primarily confined within ideological or interest-based subgroups rather than evenly distributed across the network. The network was divided into 26 distinct communities, as shown in the size distribution plot. While some large communities contain over 900 nodes, many smaller groups exist with fewer than 200 members, reflecting varying levels of engagement and influence. The presence of several mid-sized communities suggests that while a few dominant clusters drive discourse, smaller groups still play a role in the network structure.

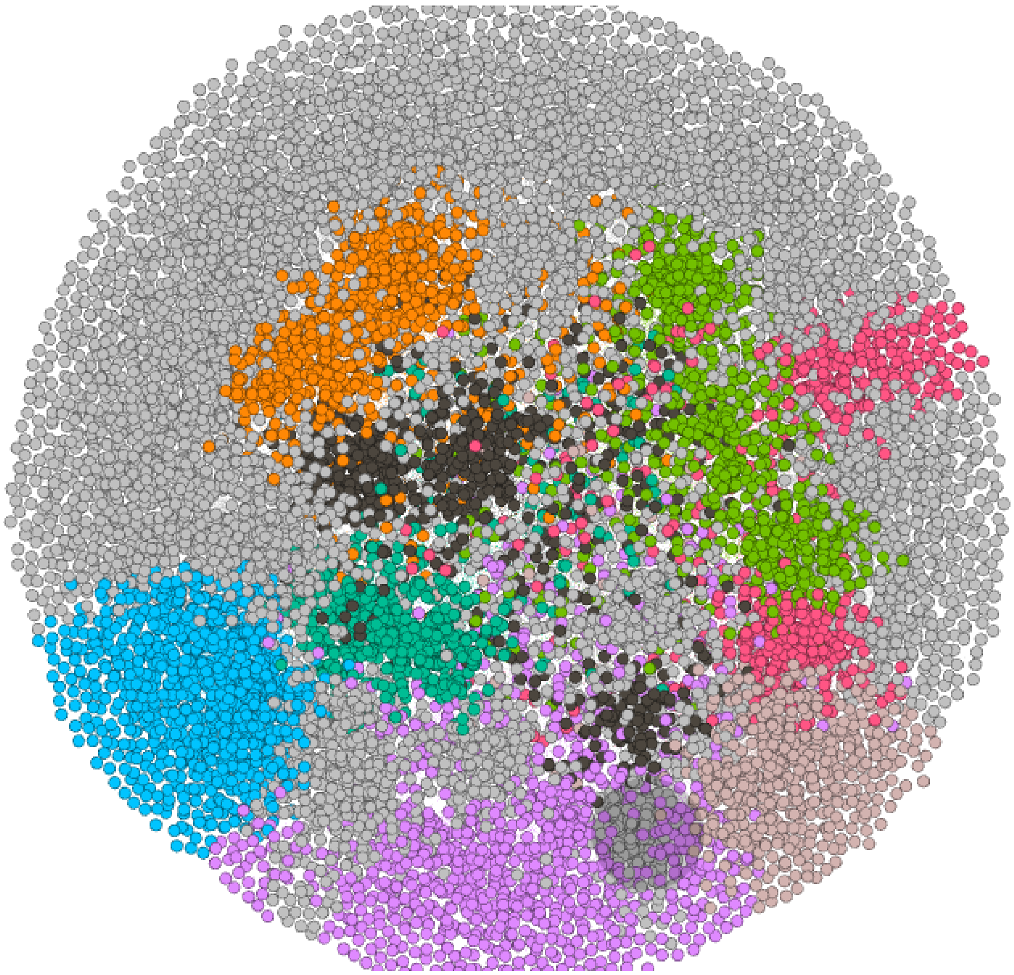


Figure 13: Modularity Groups

In this representation of the network, we focused on the seven largest communities, which contain the most nodes and play a dominant role in the structure of the network. These key communities are distinctly colored, while the smaller communities are grouped in grey, representing a significant portion of the total network but each individually having a lower number of nodes. An interesting observation is that although individually the smaller communities are not as prominent, in total, they comprise a greater number of nodes than the largest communities. This suggests that the network is not solely dominated by large influential groups but also consists of numerous smaller personal networks, likely representing friendships, local communities, and niche interactions. These smaller groups indicate that individual-level interactions still play a role in shaping the network.

The next visualization bellow highlights the dominant communities within the network, emphasizing the most influential users and structural patterns. The size of the nodes reflects their degree centrality, meaning that larger nodes have significantly more connections and thus represent highly influential users. A hierarchical structure is evident, as many nodes exhibit radial (spoke-like) formations, where influencers act as primary hubs, directly connected to many users. This suggests a top-down flow of information, with certain individuals playing a central role in shaping discourse within their communities. The central red node, which stands out in both size and position, is likely the main account under analysis. Additionally, several clusters maintain their distinctiveness while still being interconnected, reinforcing the modularity findings that strong community structures exist, but certain bridges allow cross-community interactions.

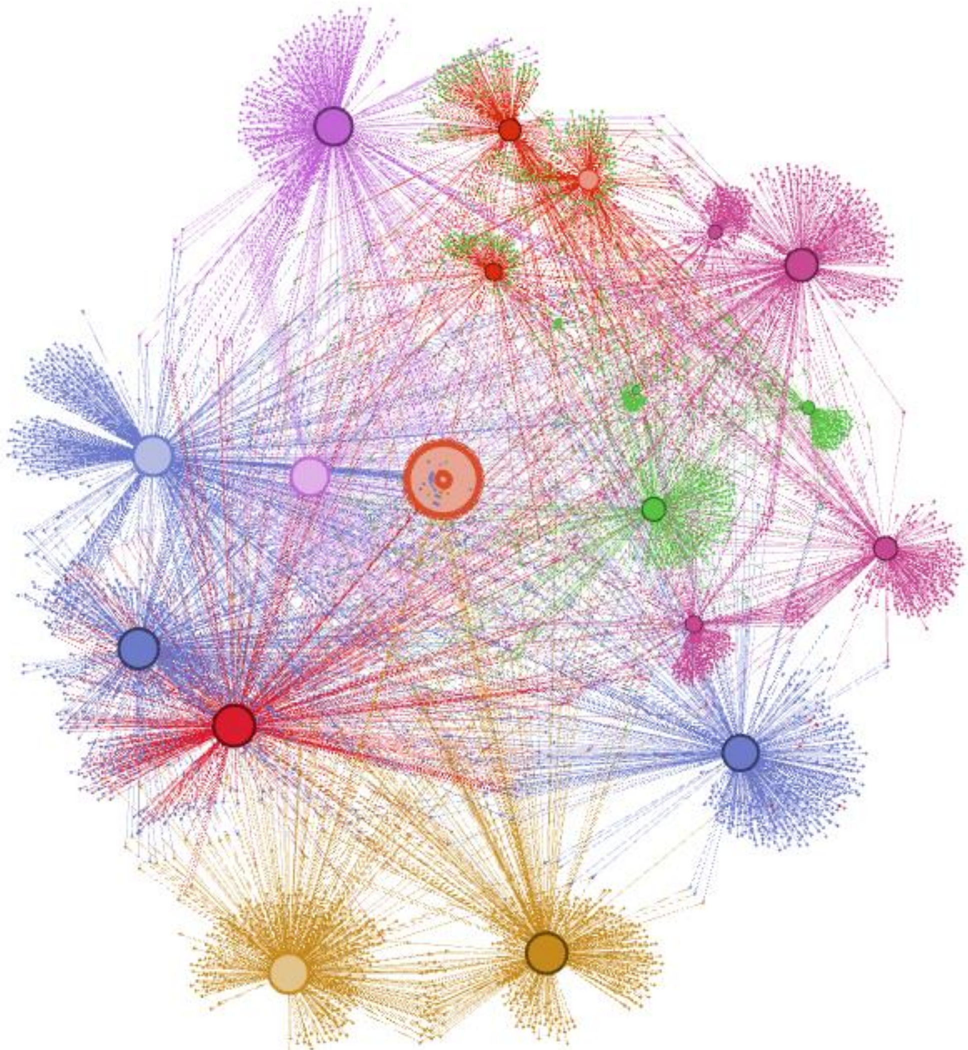


Figure 14: Dominant Communities Graph

12. Page Rank

PageRank is a centrality measure that evaluates node importance based on incoming connections and the influence of linking nodes. Unlike degree centrality, which only counts direct links, PageRank assigns higher scores to nodes that receive connections from already influential users.

The PageRank distribution plot shows that most users have low influence, while a few highly ranked nodes dominate. A small number of key figures attract the majority of attention, while the rest have minimal impact.

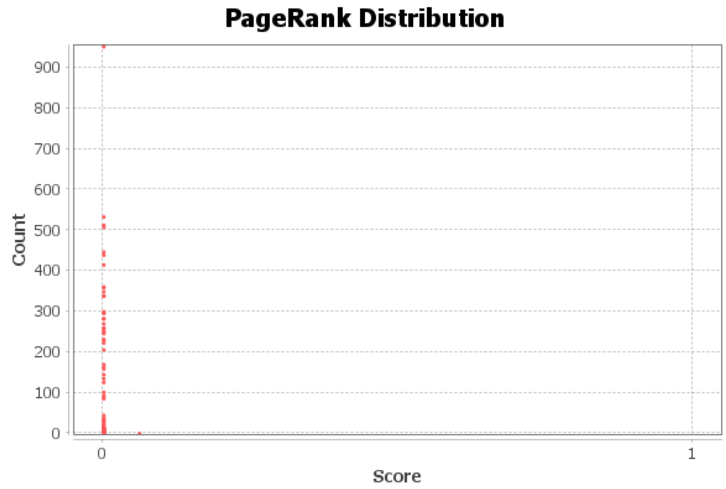


Figure 15: PageRank Distribution

13. Discussion

While this study provided valuable insights into polarization, homophily, community structures, and influence dynamics within a political network, several areas remain open for further exploration. A deeper look into temporal dynamics—how connections and influence evolve over time—could reveal shifts in political discourse and network fragmentation. Additionally, incorporating sentiment analysis on user interactions could help distinguish between supportive, neutral, or antagonistic engagements, offering a more nuanced understanding of ideological divisions.

Furthermore, our current approach relies on profile descriptions for classification, which, while useful, may not fully capture engagement behaviors. A more precise method would be to analyze actual interactions, such as comments, replies, reposts, or quote tweets, to understand how users communicate and reinforce their ideological positions. This would allow us to distinguish between passive followers and active participants, giving a clearer picture of how polarization manifests in discussions and engagement patterns.

Lastly, analyzing misinformation pathways and echo chambers would be crucial in understanding how political narratives reinforce themselves and contribute to further polarization in American politics.

14. Conclusion

This study provided a detailed analysis of political polarization, homophily, influence structures, and community dynamics within a politically engaged social media network. By examining the followers of a key progressive figure, we observed that ideological clustering is strong, with users forming well-defined communities rather than engaging broadly across different political groups. The assortativity and homophily analyses revealed that self-identified women and non-binary users tend to interact more with each other, reinforcing the idea of solidarity-driven engagement in political activism. At the same time, the Anti-Trump community exhibited stronger homophily compared to general Democrats, suggesting that opposition-based political engagement often leads to more insular discussions.

The structural analysis of the network highlighted a hierarchical, hub-and-spoke model, where a few highly connected individuals act as central figures, directing most of the information flow. PageRank and degree centrality metrics confirmed that influence is concentrated within a small group of users, while the majority remain passive participants with limited reach. Despite this concentration, bridging nodes and cross-community interactions still exist, demonstrating that not all discourse is fully isolated. The modularity findings also showed that while large ideological communities dominate, smaller, personal networks remain an important part of the ecosystem, representing localized discussions among friends, relatives, and niche groups.

Ultimately, this analysis underscores the complexity of political engagement online, where both ideological reinforcement and cross-community interactions coexist. While polarization is evident, it is not absolute, and understanding these digital spaces is crucial for addressing the growing divide in American political discourse. Further studies focusing on direct interactions rather than profile descriptions, comparative analyses of opposing political networks, and misinformation flow would provide even deeper insights into how political narratives form, spread, and solidify in online communities.

15. References

1. Barberá, P. (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis, 23(1), 76-91.*
2. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). "The Spreading of Misinformation Online."
3. "Homophily and Polarization in Twitter Political Networks" (2022)
4. "Fueling the Fire: How Social Media Intensifies U.S. Political Polarization—And What Can Be Done About It" (2021)
5. "The Polarizing Impact of Political Disinformation and Hate Speech" (2023)