

# Analyzing Political Narratives in Books and Blogs

Hamidreza Bahmanyar, LM in Artificial Intelligence, 0001113333

Fatemeh Izadinejad, LM in Artificial Intelligence, 0001118789

Lyudmil Stamenov, LM in Artificial Intelligence, 0001111453

Itrat Zehra, LM in Artificial Intelligence, 0001050816

## 1 Introduction

This project examines the Political Books Network and the Political Blogs Network, two essential datasets for exploring ideological polarization within American politics. The book network captures patterns of joint book buying, while the blog network reveals hyperlink connections during the 2004 presidential election. By analyzing network structures and community formation, this project aims to uncover the mechanisms that drive political divisions and the role of media consumption in reinforcing or bridging ideological divides.

The source code for this project is available [here](#).

## 2 Problem and Motivation

Political polarization, exacerbated by the selective consumption of ideologically aligned content, is a crucial issue influencing public discourse, voter behavior, and social cohesion. The Political Books Network and the Political Blogs Network provide a unique opportunity to analyze how media consumption patterns reinforce or bridge ideological divides. Books are often purchased collectively along ideological lines, while blogs create highly polarized online spaces through selective hyperlinks. Understanding the structure of these networks is crucial to studying the formation of echo chambers and their impact on public opinion.

We decided to compare these two networks to analyze how each type of media impacts the division of political opinions and which one encourages people to explore opposing perspectives. To contribute to this discussion, this project focuses on analyzing political book and blog networks using key measures of structure and centrality. These datasets provide a manageable yet rich framework for assessing polarization and influence in a well-documented historical context. By identifying influential nodes and bridging links, this study aims to uncover mechanisms that drive division or promote cross-group exposure. This provides valuable insights for mitigating polarization in similar media networks and informs future strategies for a more inclusive flow of information.

## 3 Datasets

This project utilizes the Political Books Network and the Political Blogs Network datasets to study ideological polarization and media consumption during the 2004 U.S. presidential election. The Political Books Network comprises 105 nodes representing political books and 441

edges capturing co-purchase relationships, reflecting how readers cluster around ideologically aligned content. The Political Blogs Network consists of 1,224 nodes representing blogs and 19,090 directed edges depicting hyperlinks between them, modeling interactions in polarized online spaces. Both datasets, sourced from the publicly available Networks Repository [2][3], were obtained in digital formats, requiring no additional digitization.

Data handling and analysis were performed using Python libraries, including Pandas [4] and NetworkX [1], which enabled efficient data manipulation and the computation of key network metrics such as centrality, clustering coefficients, and k-core decomposition. The datasets were structured and preprocessed using Pandas for data wrangling and cleaning, while NetworkX provided tools for representing and analyzing the networks. The use of standardized tools and well-documented processes ensured accurate and reproducible results and a comprehensive understanding of the structural dynamics of polarization within both networks.

## **4 Validity and Reliability**

### **4.1 Validity**

The Political Books Network and Political Blogs Network datasets exhibit a high level of validity by accurately modeling real-world dynamics of political polarization and media consumption during the 2004 U.S. presidential election. The books network reflects co-purchase behaviors where readers gravitate toward ideologically aligned content, closely mirroring real consumer tendencies in politically charged contexts. Similarly, the blogs network, built on hyperlink connections between politically aligned blogs, provides a realistic representation of online discourse, where selective linking reinforces ideological cohesion and isolates opposing viewpoints. These networks effectively highlight the formation of echo chambers—an extensively studied phenomenon in political communication. While the datasets do have certain limitations, such as the exclusion of offline purchases and interactions outside blog platforms, they still serve as valid models for understanding ideological divides in media consumption. Their well-structured design, coupled with network metrics like clustering and community detection, makes them highly effective for analyzing polarization patterns within their historical context.

### **4.2 Reliability**

This study ensures a high degree of reliability through its consistent and reproducible approach to data analysis. The datasets were used in their original form without additional preprocessing, as initial inspection showed no issues with missing data or duplicate edges. Standard network metrics were computed using default configurations in NetworkX, ensuring uniform application across repeated studies.

The use of widely accepted network analysis measures—such as degree centrality, modularity, clustering coefficients, and PageRank—further enhances reliability, as these methods have proven reproducible in various network studies. The fact that the datasets represent static snapshots of the 2004 political environment also contributes to their reliability, allowing future studies to replicate and verify the results without temporal variations. Although applying the same methodology to modern datasets may introduce new factors, the well-defined analytical framework ensures consistent and robust reproducibility in similar contexts.

## 5 Measures and Results

### 5.1 Node-Level Measures

#### 5.1.1 Degree Centrality

Degree centrality measures the number of direct connections a node has. It quantifies a node's local influence within the network, with higher degree centrality indicating a node that interacts frequently with others.

**Formula:**

$$C_D(v) = \frac{d(v)}{N-1} \quad (1)$$

where  $C_D(v)$  is the degree centrality of node  $v$ ,  $d(v)$  is the number of edges connected to node  $v$ , and  $N$  is the total number of nodes in the network.

In the political books network, the books with the highest degree centrality are *A National Party No More* and *Off with Their Heads*, both with a degree centrality of 0.240385. Other highly connected books include *Bushwhacked* (0.221154) and *Losing Bin Laden* (0.221154). These books are central hubs within their respective ideological communities, frequently co-purchased with other books that reinforce their ideological viewpoints. The structure suggests strong community formations, where political books tend to be interconnected within ideological boundaries rather than across them. In the political blogs network, the blog with the highest degree centrality is *blogsforbush.com* (0.381848), followed by *dailykos.com* (0.313164) and *instapundit.com* (0.295993). These blogs serve as major hubs in the network, frequently linked by other blogs within their respective ideological spheres. The degree centrality distribution shows that political blogs exhibit a power-law-like structure, where a small number of highly connected blogs dominate the discourse while most blogs have relatively few connections.

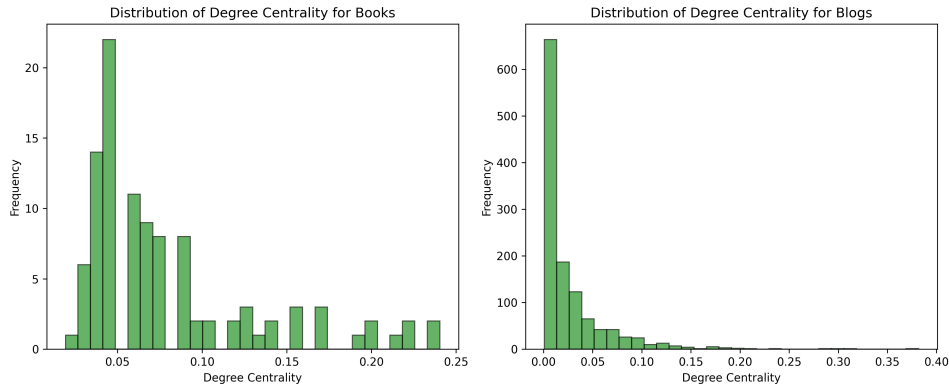


Figure 1: Degree Centrality Distribution for Political Books and Blogs Networks

#### 5.1.2 Betweenness Centrality

Betweenness centrality quantifies the extent to which a node acts as a bridge between other nodes by appearing on the shortest paths between them. It captures the importance of nodes in connecting different communities.

**Formula:**

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2)$$

where  $C_B(v)$  is the betweenness centrality of node  $v$ ,  $\sigma_{st}$  represents the total number of shortest paths between nodes  $s$  and  $t$ , and  $\sigma_{st}(v)$  represents the number of shortest paths passing through node  $v$ .

The highest betweenness centrality books are *Off with Their Heads* (0.095262), *American Dynasty* (0.090934), and *The Price of Loyalty* (0.139478). These books serve as intermediaries between ideological clusters, connecting different ideological audiences. Their structural position suggests that they are more likely to be accessed by readers from diverse political backgrounds. In the political blogs network, *blogsforbush.com* has the highest betweenness centrality (0.146178), followed by *instapundit.com* (0.051034) and *atrios.blogspot.com* (0.060880). These blogs act as key intermediaries between different political discussions, linking various subgroups within the network and shaping the flow of information across ideological divides.

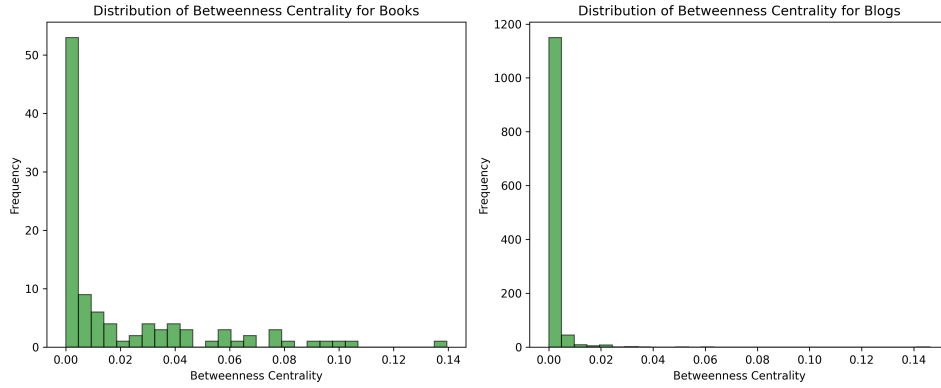


Figure 2: Betweenness Centrality Distribution for Political Books and Blogs Networks

### 5.1.3 Eigenvector Centrality

Eigenvector centrality measures a node's influence by considering not only its own degree but also the importance of the nodes it connects to. A node with high eigenvector centrality is well-connected to other highly influential nodes.

**Formula:**

$$C_E(v) = \frac{1}{\lambda} \sum_{u \in N(v)} C_E(u) \quad (3)$$

where  $C_E(v)$  is the eigenvector centrality of node  $v$ ,  $N(v)$  represents its set of neighbors, and  $\lambda$  is the largest eigenvalue of the adjacency matrix.

The books with the highest eigenvector centrality values are *Bushwhacked* (0.296894), *Big Lies* (0.284085), and *The Great Unraveling* (0.254842). These books are embedded within highly influential clusters, frequently appearing in recommendation networks alongside other politically influential books. Their prominence suggests they play a key role in shaping the ideological landscape of their respective communities. In the political blogs network, *dailykos.com* has the highest eigenvector centrality (0.216400), followed by *atrios.blogspot.com* (0.234263) and *talkingpointsmemo.com* (0.210346). These blogs are deeply embedded within influential blog clusters, serving as reference points for other well-connected blogs in their political communities.

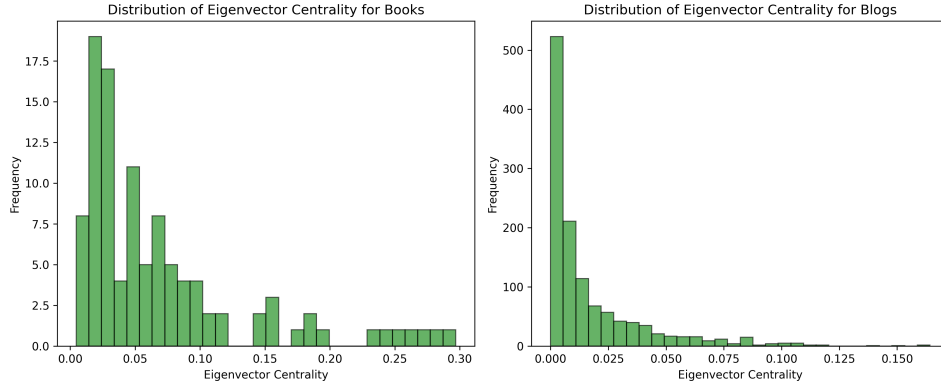


Figure 3: Eigenvector Centrality Distribution for Political Books and Blogs Networks

### 5.1.4 PageRank

PageRank measures a node's influence based on its connections and the importance of the nodes it links to. In political media, high PageRank books and blogs act as key ideological anchors that shape discourse within their communities.

In the political books network, the PageRank distribution is relatively balanced, with a considerable number of nodes having PageRank values above 0.01. The highest PageRank values are 0.025587 for a conservative node and 0.023180 for a liberal node, while the highest for a neutral node is 0.013109. This indicates that neutral nodes do not form as dense clusters. Although the network relies heavily on nodes with high PageRank, such influential nodes are proportionally more abundant than in the political blogs dataset.

In contrast, the political blogs network exhibits a heavily skewed distribution, with most nodes having PageRank values below 0.002. The highest PageRank values are 0.018845 for a liberal blog and 0.013282 for a conservative blog. Notably, while these top values in the blogs network are lower than those in the books network, the blogs network is approximately 12 times larger. Given that PageRank values are normalized, one might expect the values in a larger network to be roughly 12 times smaller; however, this is not the case. This discrepancy indicates that a few nodes in the political blogs network wield even greater relative influence, underscoring the network's heavy reliance on a small number of highly important nodes.

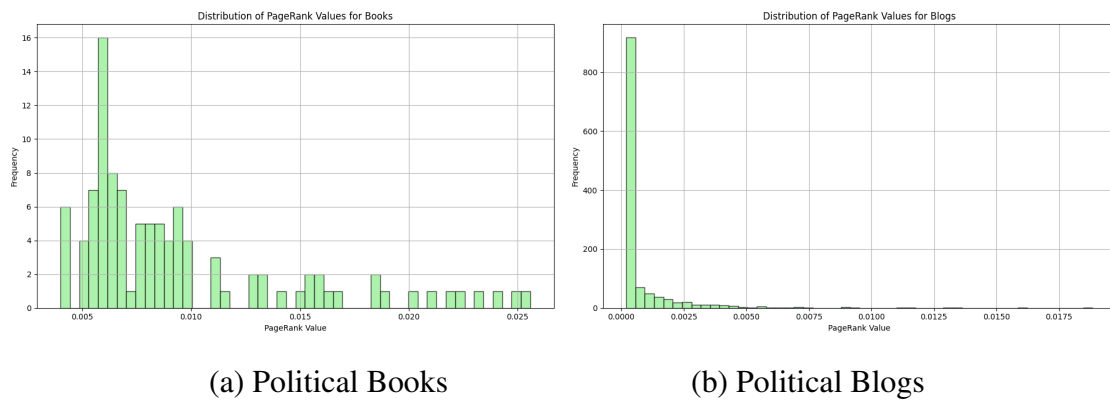


Figure 4: PageRank Distributions for Political Media

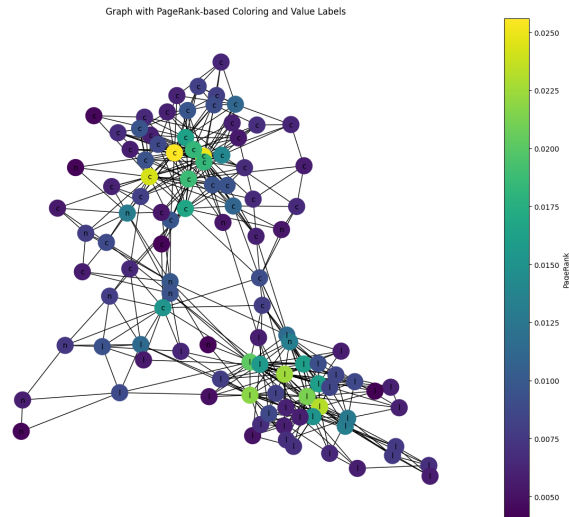


Figure 5: Political Books Network with PageRank-based coloring

## 5.2 Network-Level Structure & Cohesion Measures

### 5.2.1 Density

Density measures how connected the network is, based on the ratio of actual to possible edges, where the total possible edges in a network with  $N$  nodes is  $N(N-1)$  for directed networks and  $N(N-1)/2$  for undirected ones. On calculating the density for Political books network and Political blogs network, the values gained are 0.0808 and 0.0127, which indicates a significantly higher density for the books network. It suggests that people who purchase political books often buy multiple related books, creating a denser co-purchase network. In contrast, blogs network formed links selectively forming a more fragmented and less interconnected network. Blogs' lower density suggests linking primarily like-minded sources and indicate polarization. The higher density of books indicates a diverse readership, which also suggests that book buyers are open to exploring multiple points of view than blog readers, who may stay within their ideological bubble.

### 5.2.2 Average Path Length

Average Path Length (APL) measures the average number of steps required to travel between any two connected nodes in a network. It reflects how efficiently information or influence can spread. For books' network, the APL is 3.079 and APL for blogs' network is 3.189 which exists for its largest strongly connected component. Shorter path between books indicate that readers tend to buy books that are highly interrelated, forming tightly connected clusters, and slightly longer paths among blogs' suggest more fragmentation and possible ideological separation. Since the value is computed for the Largest Strongly Connected Component (SCC), this means that even within the most connected part of the blog network, it takes more steps to reach different blogs. This could be due to political echo chambers, where blogs primarily link within their own ideological groups rather than across groups. The blogs network is likely more polarized, with longer paths making it harder for information to bridge different communities. And the books network is more navigable, suggesting that readers may explore a wider range of political books, even across ideological lines.

### 5.2.3 Clustering Coefficient

The clustering coefficient measures the tendency of nodes to form tightly-knit groups (triangles). It is computed as the ratio of number of closed triplets (triangles) and total triplets. For the books network, the value is 0.488 and for the blogs, it is 0.218. A higher clustering coefficient among books suggests a more locally interconnected network meaning co-purchased books often share strong topic/ideology ties, while a lower value in the blogs network indicates fewer tightly connected groups, possibly due to ideological divisions. This aligns with the polarization hypothesis that political blogs might be part of separate, weakly connected ideological groups. It also suggests that book readers may be more willing to explore multiple sources, leading to overlapping clusters, while political discussions in blogs might be more segregated, forming distinct ideological clusters rather than a well-mixed discussion space.

### 5.2.4 Assortativity Coefficient

Assortativity measures the tendency of nodes in a network to connect with similar nodes based on a specific characteristic. It helps analyze homophily (similar nodes connecting) or heterophily (dissimilar nodes connecting). Mathematically, assortativity is computed using the Pearson correlation coefficient. We'll calculate two types of assortativity here:

**Degree Assortativity (Structural Assortativity)** measures how nodes with similar degree (number of connections) tend to be connected. It ranges from -1 to 1. For both of our networks, we got a negative value which means high-degree nodes connect to low-degree nodes (hub-and-spoke structure). The values obtained are -0.1279 for political books network and -0.2308 for the political blogs. As blogs network is more disassortative, it means that influential blogs spread information to smaller, niche blogs rather than reinforcing a dominant elite group. The weaker disassortativity of books network indicates that well-known books are not as dominant as blogs and may be more interconnected with medium-sized and niche books. This could support more diverse exploration of ideas compared to blogs. The plot in the Figure 6(a) shows a slightly downward trend which suggests a mix of popular and niche books being bought together, though there is no strong division. And the steep downward trend in Figure 6(b) shows that the network is highly centralized around a few key influential blogs.

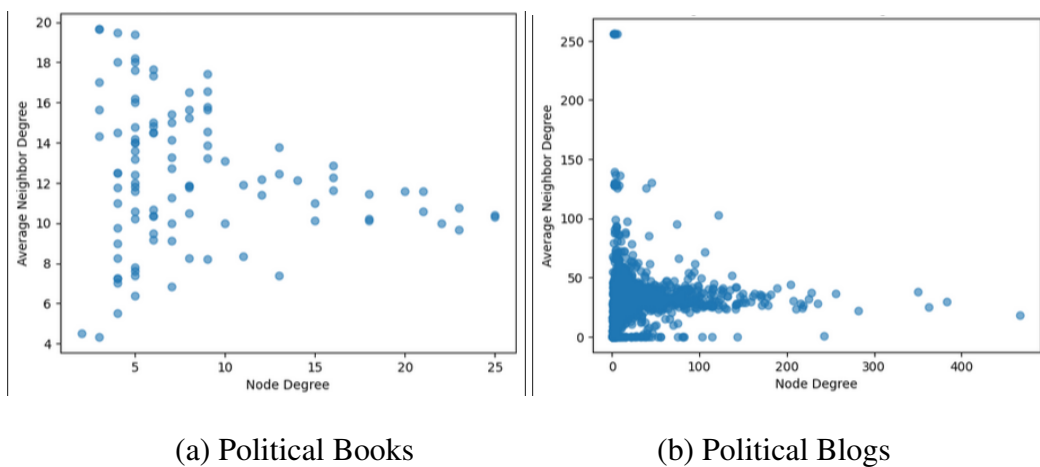


Figure 6: Degree Assortativity for Political Media

**Attribute Assortativity** measures the tendency of nodes to connect with similar attributes which in this case is political affiliation. It can confirm whether books or blogs mainly link within their

political affiliation forming echo chambers. Political attribute assortativity for blogs is 0.8229 and for books, it is 0.7233, which means that both networks are highly politically segregated, but blogs are even more polarized than books. Although books network shows slightly more cross-political connections, possibly due to books being referenced across groups, and blogs forms strong echo chambers, where information stays within political factions.

### 5.2.5 Modularity

Modularity measures how well a network is divided into distinct communities by comparing the density of edges within groups to the density of edges between groups.

**Formula:**

$$Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (4)$$

where  $Q$  is the modularity score,  $A_{ij}$  represents the adjacency matrix,  $k_i$  and  $k_j$  are the degrees of nodes  $i$  and  $j$ ,  $m$  is the total number of edges, and  $\delta(c_i, c_j)$  is 1 if nodes  $i$  and  $j$  belong to the same community and 0 otherwise.

The modularity score for the political books network is 0.526966, indicating strong ideological clustering. This means books are primarily co-purchased within ideological communities, reinforcing distinct political divisions. The political blogs network has a lower modularity score of 0.427098, suggesting that while ideological clustering still exists, there is slightly more cross-community interaction compared to books. However, both networks display characteristics of ideological echo chambers.

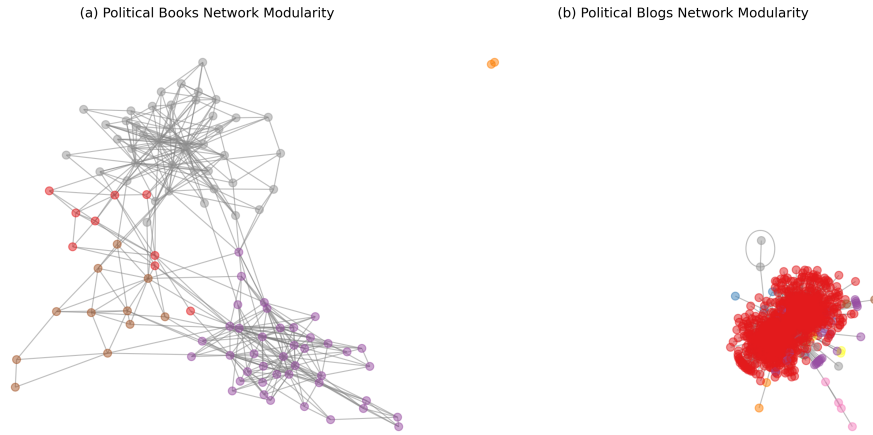


Figure 7: Modularity Clustering of Political Books and Blogs Networks

### 5.2.6 Scale Freeness

Scale-freeness is a network property characterized by a degree distribution that follows a power-law, whereby a small number of nodes (hubs) possess exceptionally high degrees while the majority of nodes exhibit low degrees. This property is typically evaluated by fitting a power-law to the observed degree distribution and computing the corresponding coefficient.

Our analysis indicates that neither network adheres to a power-law distribution. As illustrated in Figure 8, most nodes have low degrees, with only a few nodes attaining high degree values. This observation is corroborated by Figure 9, which presents the cumulative degree distribution.



Although the Political Books dataset exhibits a cumulative degree distribution that is somewhat more linear—suggesting a closer approximation to scale-freeness—the predominance of low-degree nodes causes a significant deviation from linearity. This deviation is even more pronounced in the Political Blogs dataset, where the proportion of low-degree nodes further distances the distribution from a power-law behavior.

The scarcity of nodes with high degrees suggests that both networks rely heavily on a very small number of highly connected nodes. In political media, this concentration of connectivity implies that a limited number of influential entities play a critical role in information dissemination and control. Consequently, while the networks do not strictly follow a scale-free model, their functionality may nonetheless depend substantially on these few hubs.

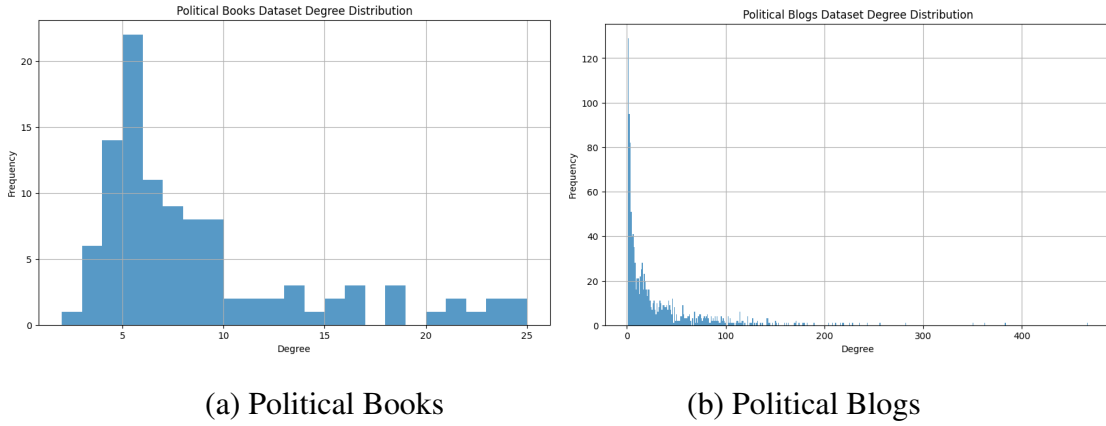


Figure 8: Nodes Degree Distributions for Political Media

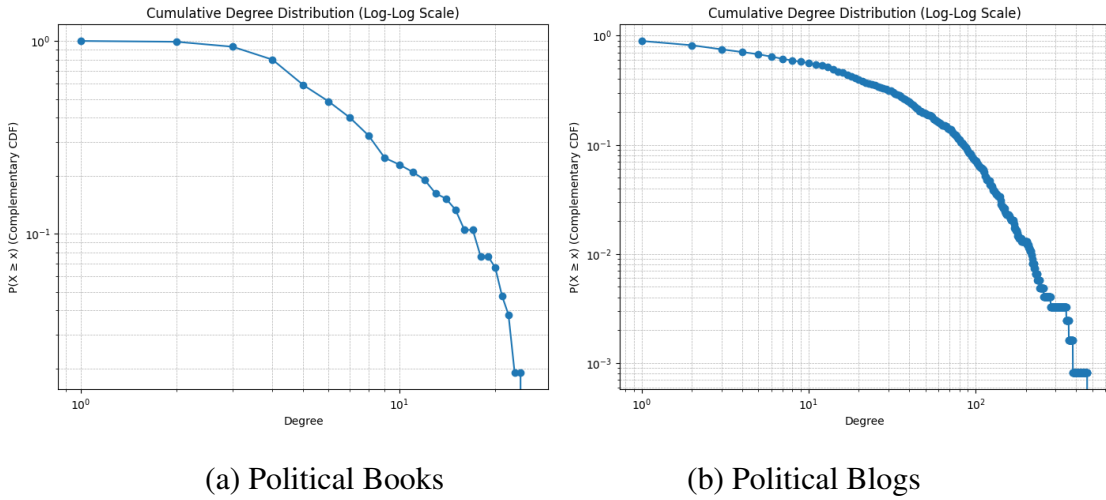


Figure 9: Cumulative Degree Distributions for Political Media

## 5.3 Substructure & Motif-Based Analysis

### 5.3.1 k-Core Decomposition

The k-core decomposition is a graph-based metric that identifies the core structure of a network by iteratively removing nodes with degree less than a specified threshold, eventually revealing

core subgraphs. Each node in a  $k$ -core has at least  $k$  connections, and this decomposition helps analyze the resilience and community structure of a network.

For the political books network, the  $k$ -core decomposition reveals that conservative and liberal nodes are primarily located in the higher cores (4, 5, and 6), while neutral nodes are distributed more evenly across the cores. This suggests that conservative and liberal ideologies tend to form distinct clusters, while neutral nodes are more spread out throughout the network.

In contrast, the political blogs dataset shows a different pattern. The majority of the nodes are found in the lower cores (1, 2, and 3) and higher cores (37 and 43). Conservative nodes are notably absent from cores higher than 37, but they compensate for this by having a significant presence in core 37, which remains a key point of connection. Meanwhile, a large number of liberal nodes are concentrated in core 43. This suggests that, while liberal blogs form a strong, well-defined cluster in core 43, conservative blogs maintain a substantial presence through their concentration in core 37. The larger size of the higher cores in the blogs network, especially core 43, indicates a greater degree of ideological clustering, making the political blogs network more fragmented compared to the books network, where ideological groups are more integrated and balanced. This highlights how political blogs, particularly liberal ones, have more concentrated and localized communities, whereas the political books network displays more dispersed ideological affiliations.

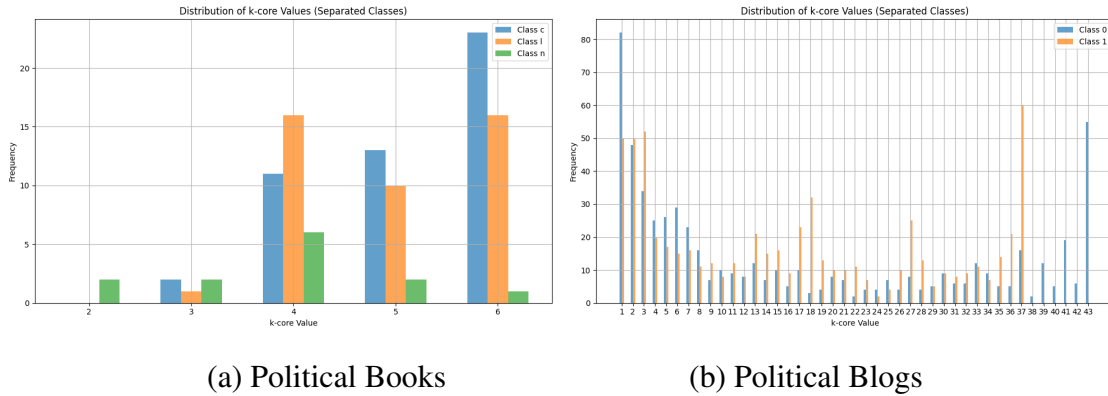


Figure 10:  $k$ -Core Values Distributions for Political Media

### 5.3.2 Triad Census

The triad census is a metric used in social network analysis to classify all possible triads (subgraphs of three nodes) based on their unique configurations of directed edges. Since the original graphs for both the books and blogs datasets are undirected, we converted them to directed graphs by treating each undirected edge as bidirectional, enabling the analysis of structural patterns within triads.

The results show notable differences in triad distributions between the two networks. In the books network, the most frequent triad is 003 (78.04%), representing three disconnected nodes, followed by 102 (19.98%), indicating a single bidirectional edge with the third node disconnected. The presence of 201 (1.68%) and 300 (0.30%) triads suggests minimal clustering and reciprocity.

In the blogs network, the most common triad is also 003, but its prevalence is even more pronounced at 93.71%, with a total count of 285,693,976 triads. The next most frequent triad in the blogs network is 012 (5.11%), representing a single directed edge with two disconnected nodes. Triad 102 (0.81%) appears next, followed by 021D (0.05%), 021U (0.16%), and 021C

(0.04%), all indicating weak connections and a lack of strong clustering. The blogs network also exhibits a small proportion of more complex triads, such as 111D (0.04%) and 111U (0.03%), which further highlights the sparsity of connections in the network.

The high prevalence of 003 triads in both networks reflects their sparse structure, with limited clustering, particularly in the blogs network. The higher proportion of disconnected triads in the blogs network, especially with a larger number of 003 triads, is consistent with its larger scale and lower density compared to the books network. This suggests that the blogs network is more fragmented, with a higher degree of disconnectedness, while the books network, despite being sparse, shows slightly more connectedness and clustering within its triadic structure.

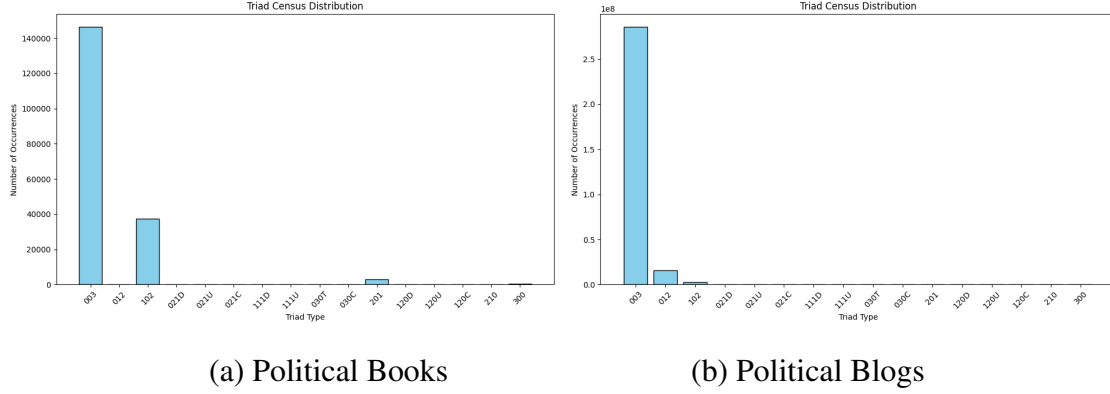


Figure 11: Triad Census Distributions for Political Media

## 6 Conclusion

This report provides a comparative analysis of the Political Books Network and the Political Blogs Network, examining how different media platforms shape ideological polarization and information flow. Using network analysis techniques, it analyzes influence, connectivity, and clustering patterns across both platforms. The findings reveal distinct structural characteristics, highlighting varying degrees of ideological segregation and influence distribution within these networks.

The Political Books Network exhibits a moderately interconnected structure, where degree centrality, eigenvector centrality, and PageRank highlight influential books that shape political discourse. Despite some cross-ideological connections, high betweenness centrality values suggest that a few key books serve as bridges between ideological groups, making the network dependent on specific nodes for balanced discourse. The density and clustering coefficient indicate that book readers are more likely to engage with multiple related books, creating interconnected but ideologically distinct communities. Encouraging cross-ideological reading habits through diverse recommendation strategies could enhance ideological exposure and reduce polarization in book consumption.

Conversely, the Political Blogs Network demonstrates a fragmented and polarized structure, with high modularity and assortativity scores confirming strong ideological clustering. The clustering coefficient and attribute assortativity analysis reveal that blogs tend to link within their ideological group, reinforcing echo chambers. Additionally, scale-freeness analysis suggests that a small number of highly influential blogs dominate political discourse, concentrating information flow around a few central nodes. This structure amplifies polarized narratives, making it harder for users to encounter diverse perspectives. Addressing this issue requires al-

gorithmic modifications in content recommendations, fact-checking mechanisms, and platform policies that promote ideological diversity to mitigate extreme polarization.

The k-core decomposition and triad census further illustrate structural differences between the networks. The books network features a more distributed core, indicating a relatively balanced ideological reach, while the blogs network has high-core groups dominated by ideologically homogeneous clusters, reinforcing stronger echo chambers. The average path length also suggests that information circulates more efficiently in the books network, whereas blogs require longer paths for cross-ideological connections, further restricting interaction between opposing viewpoints.

These findings emphasize the impact of media structure on polarization, with books allowing for some ideological overlap, while blogs function as reinforcement hubs for political segregation. Future research could explore temporal analyses to track polarization trends, sentiment analysis to assess ideological intensity, and alternative community detection methods to identify hidden ideological subgroups.

By understanding how different media reinforce or mitigate polarization, policymakers, educators, and media platforms can implement strategies to promote balanced political engagement and reduce ideological isolation. Encouraging diverse media consumption and fostering algorithmic transparency in digital platforms are crucial steps toward addressing the growing challenges of political polarization and information silos.

## 7 Critique

The study effectively analyzes ideological polarization by comparing the Political Books Network and Political Blogs Network, highlighting key differences in influence, clustering, and connectivity. While it provides valuable insights into how books allow for some ideological overlap while blogs reinforce echo chambers, it does not fully solve the problem.

The research primarily focuses on structural network properties but lacks content-based analysis, such as sentiment or topic modeling, to understand the nature of discourse within these networks. Additionally, the reliance on static datasets from 2004 limits the study's applicability to modern political interactions. A temporal analysis or the inclusion of recent datasets would enhance its relevance.

While centrality measures identify influential nodes, further exploration of bridging nodes—using edge betweenness or structural hole analysis—could offer deeper insights into cross-ideological interactions. Similarly, alternative community detection methods, such as Infomap or Leiden algorithms, could provide different perspectives on ideological clustering.

Despite these limitations, the study lays a strong foundation for understanding media-driven polarization. Expanding the analysis with content-based methods, temporal studies, and alternative network models would further strengthen its findings and practical applications.

## References

- [1] *NetworkX: Exploring network structure, dynamics, and function*. <https://networkx.org>.
- [2] *Netzschleuder*. <https://networks.skewed.de/net/polbooks>.
- [3] *Netzschleuder*. <https://networks.skewed.de/net/polblogs>.
- [4] *Pandas: Data structures for statistical computing in Python*. <https://pandas.pydata.org>.