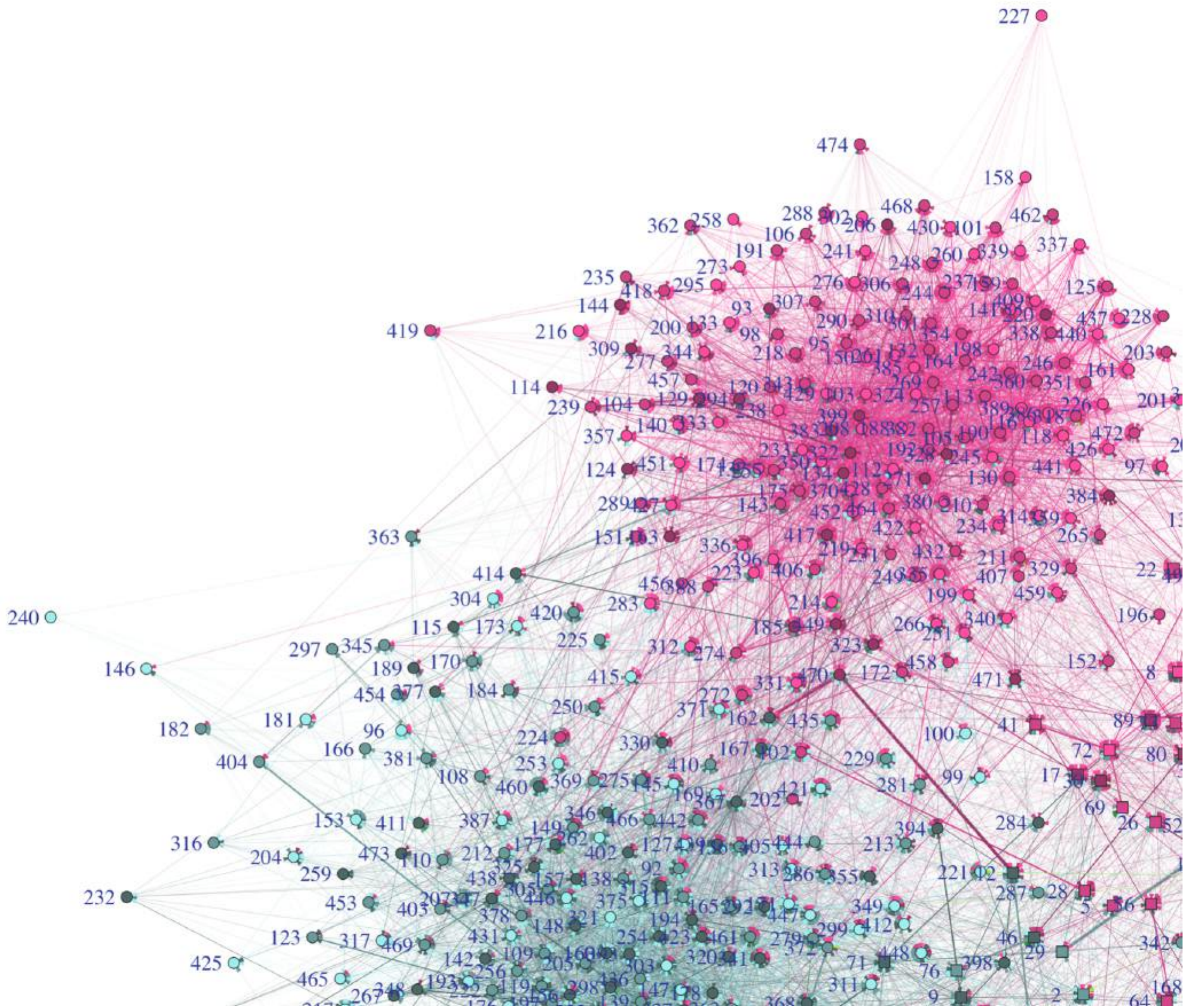
Durham University

Social Network Analysis on Congressional Twitter Network

Anonymous marking code: Z0200253



Contents

1. Abstract

This study analyses the U.S. Congressional Twitter network using descriptive social network analysis (SNA) and Exponential Random Graph Models (ERGMs). By applying a threshold on edge weights, the network is divided into four types: the unweighted all, moderate, active networks, and the original weighted network. This approach simplifies the analysis while preserving the valuable properties of edge weights, allowing both descriptive SNA and ERGMs to thoroughly and effectively explore the network. The study identifies three key politicians—Kevin McCarthy, C. Scott Franklin, and Nancy Pelosi—and their distinct interaction patterns. It underscores the significant impact of party affiliation and legislative chamber on network structure and formation, revealing patterns and dynamics of reciprocity, transitivity, and internal communities. The results also compare and distinguish the network from typical social networks in terms of classical theories such as Preferential Attachment and the Homophily Principle, alongside patterns like positive degree assortativity and high transitivity. This research offers well-rounded insights into the political network, identifies current limitations in theory, data, and methods, and provides valuable perspectives for future studies on political behaviour and social network analysis.

2. Introduction

2.1. Motivation

The Impact of Social Media on Politics

Social media use is enormous, enabling instant interactions and explosive spread of information. As Golbeck et al. (2010) argued, Twitter, a tremendously growing microblogging and social network platform, had approximately 7 million members tweeting and interacting. With the prevalence of Twitter, people believed it would revolutionize the way of communication and information sharing, urging public figures to join the platform (Golbeck et al., 2010). This public, timely, and efficient nature has made it significantly valuable in politics. Increasingly, political topics are being debated by political professionals and citizens, with politicians actively engaging on their social media accounts. Ausserhofer and Maireder (2013) highlighted that under the rapid expansion of the Internet and the World Wide Web, the political sphere is profoundly influenced by the speed and scope of communication, while politicians, journalists, political strategists and citizens now rely on social media as an essential platform for openly discussing political issues.

Additionally, social media has become a crucial tool for politicians to manipulate public opinion and establish their profile in front of the public. Buccoliero et al. (2020) suggested that during the 2016 US presidential election, social media, especially Twitter, surpassed traditional media in terms of influence, becoming the primary channel for shaping public opinion and constructing candidate images, significantly impacting election results by allowing candidates like Donald Trump and Hillary Clinton to express their positions, attack opponents, repost endorsements, encourage voting, provide news updates, even more, allowing candidates to express their perspectives and gauge voter reactions without mainstream media filters. Golbeck et al. (2010) introduced TweetCongress, a grass-roots internet-based campaign, encouraging Congress members to use Twitter to enhance political transparency. However, Golbeck et al. (2010) concluded that while Twitter does indeed support direct communication between Congress members and their constituents, such communication rarely scales up, instead, Congress members primarily use Twitter for outreach activities rather than increasing transparency.

Moreover, social media has become informally involved in political campaigns and events. Enli and Skogerbø (2013) argued that platforms such as Facebook and Twitter are increasingly vital in political communication, with politicians using Facebook for marketing and Twitter for ongoing conversations, thus expanding the political arena to increase personalized campaign activities by shifting the focus from political parties to individual politicians. Therefore, political social networks, formed from the intersection of politics and social media, are a highly researched and sought-after object. As Conway et al. (2015) referenced various scholars, highlighting that social media on political communication patterns is an eager topic in political research, with scholars striving to understand its diverse impacts on agenda setting.

Social Network Analysis and its Application in Politics

With the advent of the Internet, the term “social network” has become a common word in our daily lives, but its applications go beyond social media platforms (Yang et al., 2017). One of the most common approaches to analyzing social networks is social network analysis (SNA). This involves comprehending and researching social relationships and structures using networks and graphs (Otte and Rousseau, 2002). Social network datasets differ significantly from conventional datasets of machine learning because they require not solely individuals and their attributes but also the connections between nodes, which are essential for SNA. Social network datasets consist of nodes and ties. According to Yang et al. (2017), the nodes usually represent actors, including individuals, groups, teams, organizations, political parties, and even countries, such as lobbyists, voters, and parties, while the ties represent various kinds of social interactions, which can be multidimensional, involving different types such as voting for a candidate, retweeting, and commenting on a message. Thus, SNA can be applied to study various social network issues. For example, Bossaert and Meidert (2013) used SNA to study dynamic analysis of the peer support networks in the Harry Potter books. Christakis and Fowler (2008) also applied SNA to study smoking behavior, disclosing crisp clusters of smokers and nonsmokers.

Social Network Analysis is not only employed in various fields such as friendship and health, but it also has significant applications in the political arena, which is the focus of this project. Yang et al. (2017) cited several examples of SNA applications in politics: Nickerson (2008) discovered that college students who live with peers who regularly discuss politics potentially engage in political discussions themselves. The study of Bond et al. (2012) demonstrated that voters are more likely to start a "voting" cascade, persuading their hesitant friends to vote due to the contagion effect of voting. Furthermore, Congress is a particularly intriguing and continuing research subject for the political SNA. Yang et al. (2017) emphasized lively debates on the optimal methods for measuring relevant networks, criticizing co-voting and co-sponsorship mainly reflecting party affiliation. Desmarais et al. (2015) suggested that examining which representatives hold press conferences together offers a better measure of cooperation, as it requires much more interaction and commitment.

2.2. Research Objectives and Methods

This dissertation aims to add a touch to the debates of Congress study by analyzing an up-to-date Twitter interaction network for the US. The research questions and hypotheses will be detailed in the Literature Review section. The Twitter Congress network represents interactions on Twitter among members of the 117th United States Congress where nodes represent the Twitter handles of Congress members, and ties represent their interactions (e.g., replay, mention, retweet or quote). The focus is on analyzing this weighted directed political Twitter interaction network using traditional descriptive methods of social network analysis (SNA) to observe, describe, visualize, and interpret the network itself, and employing Exponential Random Graph Models (ERGMs) to fit the network, identifying terms (e.g., density and homophily) related to network formation.

- Descriptive Methods in SNA: These methods include visualizing network graphs and calculating network metrics (e.g., density, centrality, centralization, cliques, homophily, transitivity) to analyze the patterns of Twitter interaction networks. For instance, using centrality can identify the most influential Congress members (Yang et al., 2017).
- Introduction to ERGMs: Hunter et al. (2008) described the Exponential Random Graph Models as a method for figuring out concisely the local selection forces shaping complex network structures. This method allows for analysis similar to traditional statistical regression but on network datasets with relationship features. For example, ERGM can help determine why certain Congress members frequently interact with each other on Twitter, revealing factors such as shared party affiliation.
- Integration of Methods: The complementary nature of descriptive SNA and ERGMs - SNA providing a macro-level overview and ERGM offering micro-level insights - can reveal the network patterns, formation and evolution of dyadic relationships, and network structures.

The data and methods will be explained in detail in the Methodology section.

2.3. Structure

The following study consists of four main sections. The first section (Literature Review) provides necessary definitions, concludes previous articles and their studies, identifies gaps and lacks in current research, and presents the research questions and hypotheses in detail. The second part (Methodology) details the network dataset and data preprocessing and explains the methods employed in the research, including descriptive methods of SNA and Exponential Random Graph Models (ERGMs). The third section (Analysis, Results and Discussion) analyzes, visualizes, and interprets the results and outcomes of the methods applied to the political Twitter interaction and spreading network. Additionally, it discusses these results and their reflections concerning the literature review. Ultimately, the final section (Conclusion) summarizes the findings and implications of the research, discusses the limitations, and outlines future study plans, which might inspire further research. The references and the appendices can be found at the end of the paper.

3. Literature Review

3.1. Previous Research

The Congressional Twitter network dataset used in this paper is a directed weighted network published by Fink et al. (2023). This dataset, alongside another Higgs Twitter network dataset, was utilized to test a new centrality measure algorithm. Centrality measures are a common measurement in complex networks for identifying the significantly influential nodes in the spread of infectious diseases, rumors, or information. However, traditional centrality measures often fall short in weighted directed networks due to simplified performance and high computational costs. To address this, they proposed a novel centrality measure called Viral Centrality (VC) to excellently approximate the results of Independent Cascade Model (ICM) simulations, a popular but typically requiring computationally intensive Monte Carlo simulations model for simulating information diffusion. The ICM is a generalization of the susceptible-infected-recovered (SIR) epidemiological model raised by Goldenberg et al., where nodes are categorized as susceptible, active (or infected) and inactive (or recovered). To analogous this process, the Viral Centrality algorithm simulates information propagation from a seed node, using edge weights as spreading probabilities and measuring how many nodes are activated to evaluate the node influence. The study of Fink et al. (2023a) showed that Viral Centrality approximates Monte Carlo simulation results with lower computational costs and performs better than conventional centrality measures.

In addition, further research built upon this dataset and the findings by Fink et al. has also focused on proposing and comparing new algorithms:

- Wang et al. (2024) used this Twitter Interaction to test their multi-factor information matrix centrality algorithm, comparing it with nine existing centrality measures. Their method offers methodological and multidimensional support for accurately identifying super-propagators in social networks, which is crucial in leading and regulating public opinions on the Internet, information retrieval, and viral marketing.
- Mao et al. (2024) introduced a novel approach combining Evolutionary Algorithms (EAs) and Large Language Models (LLMs) to generate scoring and ranking functions, which are applied and tested on the Twitter interaction network and other networks to enhance the identification of critical nodes, providing improved efficiency and accuracy.
- According to Leinwand and Lyzinski (2024), the Augmented degree corrected, Community Reticulated Organized Network Yielding Model (ACRONYM), which generates and estimates unweighted networks, was employed on both the Congressional Twitter dataset and a mouse retina dataset, describing a broader class of potential networks than existing models and providing a new perspective on understanding diverse community structures.
- Ganguly et al. (2024) also included this dataset to explore the importance of seed node selection based on community structure for influence maximization, showing the effectiveness of this method through experimental results.

3.2. Gap in Research

Fink et al. (2023) highlight the unique value of this open-source Twitter dataset, noting that it is the only one that measures influence between pairs of users. This makes it especially valuable for researchers studying complex social networks. The dataset is also useful for epidemiological models, as it allows the study of varying transmission probabilities, unlike most studies that focus on centrality measures to identify key influencers, as introduced above. Despite substantial advancements in algorithm performance and efficiency in social network analysis, studies involving the Congressional Twitter dataset have mainly concentrated on algorithmic design and optimization, including proposing or refining new algorithms, comparing them to traditional centrality measures, detecting subgroups, or utilizing this dataset as one of the test objects. However, there has been limited exploration of the specific characteristics and internal structure of the network itself. Moreover, Existing research has often overlooked the impact of the node attributes and their interactive behaviors on network formation.

To address these gaps, it is necessary to explore the influence of member attributes and partisan relationships on the network structure, which involves examining intrinsic properties, dynamic relationships, and complex structures within the dataset. Furthermore, investigating the network itself raises intriguing questions, such as which members of Congress have the most influential power, whether party alliances create distinct Democratic and Republican subgroups with limited cross-party interaction, and what attributes and elements shape the network.

3.3. Research Questions and Hypotheses

This study proposes several hypotheses and employs social network analysis (SNA) to quantitatively verify, analyze and visualize the Congressional Twitter network. In the descriptive analysis, key network metrics such as density, centrality, centralization, prestige, cliques, and subgroups will be the focus. Since Exponential Random Graph Models (ERGMs) cannot directly handle weighted networks, the study will divide the network into three levels: active (high weights only), moderate (high and medium weights), and all connections. These will be converted into non-weighted directed networks to fit the ERGM models. The detailed preprocessing steps will be explained in the Methodology section. By comparing results across networks, the study aims to effectively utilize and simplify the leverage of the weights, gaining deeper insights into the structure and formation of the network.

3.3.1. Research Questions for SNA

- **Network Density:** How connected are the members of the 117th United States Congress on Twitter?
- **Centrality Measures:** Which members of Congress are the most active, influential, and play key roles in information control and dissemination on Twitter?
- **Network Centralization:** To what extent does the Twitter interaction network revolve around a few key individuals? Is there a significant difference in how centralized the network is for Republicans vs Democrats?
- **Prestige Measures:** Which members of Congress receive the most attention and interactions on Twitter? Do senior members receive more attention on Twitter compared to junior members?
- **Cliques and Subgroups:** What are the main groups among Congress members on Twitter? Are there notable communities within Congress based on shared party, state, or seniority?
- **Patterns of Similarity (Homophily):** Do Congress members interact more with others from the same political party? Is there a pattern for members of similar seniority to connect more frequently on Twitter?
- **Mutual Interactions (Reciprocity):** Do legislators engage in reciprocal interactions? Do legislators with vested interests collaborate more to advance their agendas, leading to more give-and-take interactions?
- **Prevalence of Relationships (Transitivity):** Is there a pattern for Congress members to form triangular and tight relationships where all members interact on Twitter?

3.3.2. Hypotheses related to Preferential Attachment and Homophily Principle

According to Maoz (2012), the structural features of many physical, biological, and social networks are shaped by simple principles. Maoz highlighted the two most prominent models of network formation, Preferential Attachment (PA) and Homophily (HO), in the article. Maoz (2012) asserted Preferential Attachment showing that new nodes are more inclined to connect to central nodes for two main reasons: Large alliance has greater capabilities to fend off enemies, and approaching central nodes can provide more indirect support and resources. For instance, security networks frequently form through Preferential Attachment to strengthen the global impact of states or avoid enemy coalitions. Similarly, according to Yang et al. (2017), linking trade relationships with the central cores is the most efficient way to establish short and effective trading routes. Networks formed through Preferential Attachment typically have a star-like shape, with few components, low density and high graph centralization.

On the other hand, nodes with similar attributes tend to connect according to the Homophily principle. The number and order in the definition of similarity is a trick and open question, while most HO studies only account for one attribute at a time (Maoz, 2012). For example, Maoz (2012) described that realism believes that the enemy of the enemy is a friend, with common enemies being the highest priority for similarity, liberalism values sustainable contractual arrangements, with democratic countries seen as more trustworthy, while constructivists prioritize cooperation among countries with the same culture (e.g., Islamic countries), resulting in the order of similarity in cooperation for democratic countries might be 1) democracy, 2) common enemies, 3) similar culture, whereas for nondemocratic countries, it could be 1) common enemies, 2) similar culture, 3) democracy. In contrast to Preferential Attachment, Yang et al. (2017) noted that homophilic networks exhibit many components, such as high transitivity and low group centralization.

However, this research will not delve into the Homophily principle. Instead, it examines the similarity of one attribute at a time, such as party affiliation. By using SNA, the study aims to reveal whether interactions among Congress members follow specific patterns related to two formation principles since ERGMs are more sophisticated tools for modelling and explaining network formation.

- **Preferential Attachment:** This hypothesis posits that junior legislators are more likely to connect with high-centrality legislators to quickly establish their networks, gain exposure, and form alliances for their future careers since Congress members use Twitter to enhance their personal image as mentioned at the beginning. High-centrality legislators are anticipated to accept these connections, resulting in high out-degree and reciprocal interactions, which boost their power and competitiveness. Visualizing these interactions by seniority year should reveal a low-density, high-centralization, and star-like configuration, with senior legislators at the core and junior ones on the periphery connected to the core.
- **Homophily Principle:** This hypothesis examines how party affiliation or other attributes influence network formation. Legislators from the same party tend to form tight-knit clusters with high transitivity, while interactions across party lines might be less frequent. The network is likelier to exhibit high transitivity, low graph centralization, and multiple cores, displaying several groups with members from the same party clustering together.

3.3.3. Hypotheses for ERGMs:

Similar to predicting housing prices based on factors like location or age, network interactions can also be predicted using various variables, including node attributes (e.g., party, seniority), edge attributes, and network statistics. This study models the impact of the following parameters on network interactions, helping to reveal the underlying patterns of connections. Compared to descriptive analysis, ERGMs focus more on the dynamics and trends of connections.

Hypotheses related to Node Covariates:

- **Party, House Affiliation:** Members from the same party or house are more likely to initiate interactions.
- **Seniority Impact:** Less senior members frequently attempt to interact due to better familiarity with social media.

Hypotheses related to Network Statistics:

- **Party Homophily:** Members are more likely to interact with those from the same political party.
- **State Homophily:** Members tend to interact with others from the same state.
- **Seniority Influence:** Members with similar seniority levels are more likely to interact.
- **Nodemix:** Interaction tendencies between members with different attributes can be estimated (e.g., junior members interacting with senior members due to shared political interests).
- **Reciprocity:** Members are likely to form manual connections.
- **Transitivity:** The network is prone to forming clusters of closely connected members.

By addressing these questions, a detailed network profile could be depicted, demonstrating vital influencers, crucial connectors, concentration levels, and diverse communities within the network. While basic metrics hypotheses explain individual properties, the PA and HO hypotheses offer a more comprehensive view by observing and validating combinations of multiple properties. Moreover, understanding network formation can provide social and political scientists with valuable insights into the utility of SNA tools and the dynamics of political network dissemination. This section describes existing research, its shortcomings and deficiencies, as well as the objectives and hypotheses of this study. The next section will discuss the research data and methods used in the project.

4. Methodology

4.1. Political Twitter Network Dataset and Data Preprocessing

Dataset Introduction

This network dataset maps the Twitter interactions of the 117th United States Congress, including the House of Representatives and the Senate, collected by Fink et al. (2023) via Twitter API. This dataset forms a weighted directed network, in which the edge weights stand for the empirical transmission probabilities of impact based on how frequently one member retweeted, quote-tweeted, replied to, or mentioned another. Due to the 3200-tweet limit of the Twitter data use agreement, Fink et al. (2023) collected 179,974 tweets from 525 Congressional Twitter handles from February 9, 2022, to June 9, 2022. February 9th marked the farthest date that can be traced back, as before this, any user had exceeded the 3200 tweets. To ensure the meaning of the weight probabilities, members with fewer than 100 tweets during this period were filtered out, leaving a refined dataset with 475 members and 13,289 interactions. However, these 13,289 interactions do not differentiate between retweets, quote tweets, replies, or mentions, and leave gaps in understanding whether interactions are supportive or argumentative. To further explore the entire tweet contents or to distinguish positive from negative interactions would require a series of more complex research projects, which is beyond the scope of this study. Instead, this dataset was acknowledged to some extent, narrowing the scope by not differentiating tweet content types, making the network less complex and more focused. Therefore, two assumptions are made: (1) only interactions between Congresspersons are considered, and (2) no distinction is made between positive and negative interactions.

Network Segmentation by Weights

Since ERGMs cannot handle weighted networks and have high computational demands, the analysis will compare different levels of interactions by segmenting interactions based on weight thresholds: active (high-weight interactions), moderate (high and medium-weight interactions), and all interactions. These three segments will be converted to unweighted networks, along with the original weighted network, resulting in four versions of the network: (1) all weighted, (2) all unweighted, (3) moderate unweighted, and (4) active unweighted. Due to some weight-insensitive metrics, the descriptive analysis is limited if only employed on the original weighted network. This shortcoming can be compensated by comparing the original weighted networks with the three unweighted versions, providing a more complete analysis with the leverage of weight. The ERGMs analysis will focus on the three unweighted networks to explore different interaction levels.

Fink et al. (2023a) defined the interaction from member j to member i as:

$$p(w) = \frac{\eta_{ij}^{retweet} + \eta_{ij}^{quote} + \eta_{ij}^{reply} + \eta_{ij}^{mention}}{N_j^{tweets}},$$

Where the numerator represents the number of times member i retweeted, quote tweeted, replied to and mentioned member j, while the denominator is the total number of tweets posted by member j. The result, $p(w)$, denotes the impact probability of member j acting on member i. According to Fink et al. (2023a), edge weights are ratios (ranging from 0 to 1), standardized to avoid interference from varying tweet frequencies among legislators. The weight distribution follows a lognormal distribution (Figure 1.1a). Transforming the weights using log(w) results in an approximate normal distribution, where most values fall within one standard deviation from the mean (Figure 1.1b). Therefore, given mean $\pm$ 1 SD as thresholds, this method allows us to divide the network into three segments: low, medium, and high weights, with medium weights capturing about 68% of the data.

Dividing the network in this way simplifies analysis while still utilizing edge weight information. These segments are converted into binary networks, making it easier to apply SNA tools. Comparing all four networks through descriptive SNA allows us to effectively interpret the results, understand the patterns, and identify any limitations in this approach for future research.

Node Attributes Collection

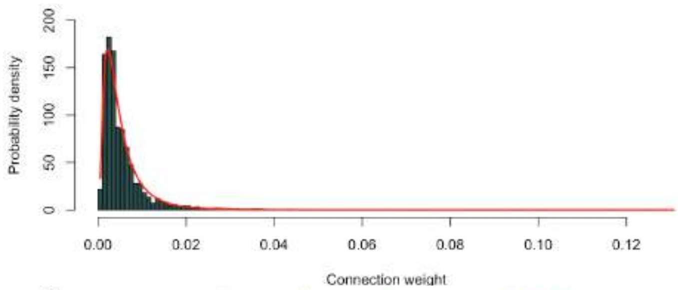
Since the original dataset only contains Twitter usernames, edges and their weights, more information on node attributes needs to be obtained, including handle, last name, first name, state, party affiliation, seniority year, and house affiliation. These attributes were gathered from two main sources: the "Twitter Handles for Members of the 117th Congress" file from the Food Research & Action Center (FRAC), which provided Twitter usernames, last names, first names, states, party affiliations, and chamber information, and "List of United States senators in the 117th Congress" and "List of members of the United States House of Representatives in the 117th Congress by seniority" from Wikipedia, offering seniority details. The data was extracted from tables, cleaned, and organized using Python before being matched with Twitter usernames. However, this process presented challenges, such as inconsistent name formats between the Wikipedia and FRAC tables, and occasional errors in party affiliation, which required manual corrections. For a detailed view of the node attributes, please refer to the Appendix section.

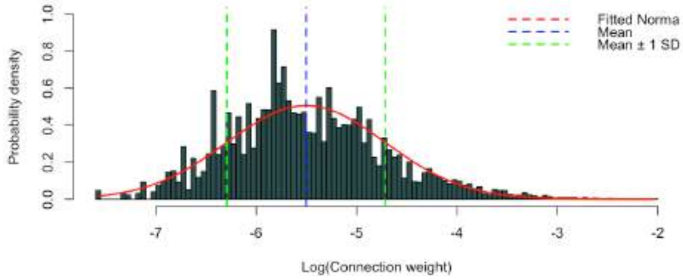
This complement enriches the information of the 117th U.S. Congress members. Basic information like the handle, last name, and first name identifies the members, while party affiliation, state, seniority, and house affiliation provide context to understand their roles and relationships within the network. Party affiliation and constituency significantly affect trust and positions between members, while seniority is a potential factor in forming political bonds. The house to which a member belongs reflects their background, with senators often coming from elite families with long family histories in politics. Therefore, they were chosen as "predictors" for analysing the network.

Initial Data Exploration

Figure 1.2 highlights that the number of members in the two major parties is approximately equal, with 2 Independents in the Senate. The House of Representatives has about four times as many members as the Senate. Figure 1.3 illustrates how U.S. Congress members are distributed across different parties and states. The number of Representatives per state is based on population, meaning states with larger populations have more Representatives, and the party distribution varies in each state. Figure 1.4 displays the distribution of Congress members by state and chamber, with each of the 50 major states in the United States having 2 seats in the Senate, regardless of the population of each state. Therefore, most states have the same number of Senate members. However, due to the network excluding members with insufficient tweets, actual numbers may vary.

Figure 1.5 shows that most current members joined Congress within 10 years. Representatives usually serve two-year terms, resulting in noticeable changes every two years. The chart also discloses that the most senior members are primarily Democrats. Figure 1.6 illustrates the distribution of members by seniority and house. Senators serve six-year terms, with roughly one-third up for re-election every two years, leading to regular changes in the Senate composition. These figures provide a detailed view of Congress members by party, state, chamber, and seniority, which will be further analysed as node attributes along with other network statistics, though the state variable will not be discussed in depth due to the large number of states and varying member counts.





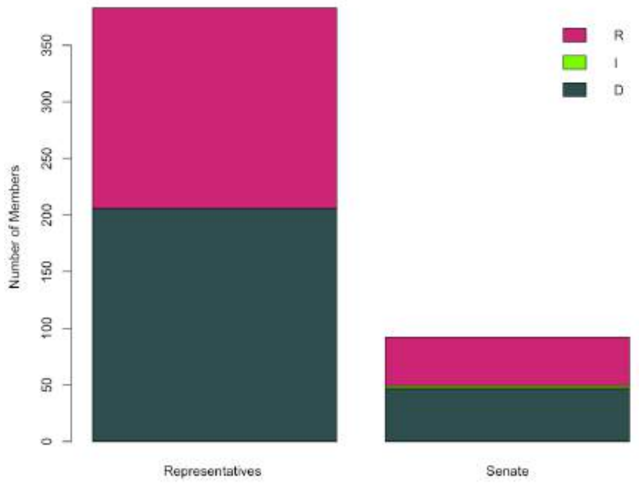


Figure 1.1a. Distribution of Edge Weights in the Network*. Red line represents fit to a lognormal distribution. The figure highlights the right skewness of the edge weight distribution, with most connections concentrated at lower weights. Data were derived from the Congressional Twitter interaction network observed over a four-month period.* ***Figure 1.1b. Distribution of Log-transformed Edge Weights.*** *Red line represents fit to a normal distribution. The figure shows that the transformed weight distribution roughly conforms to a normal distribution. Blue line denotes the mean of log-transformed edge weights. Green lines mark the mean* $\pm$ *1 standard deviation.* ***Figure 1.2. Distribution of Congress Members by House and Party****. House includes both Representatives and Senate. Deep pink block indicates Republican members. Dark green block represents the Democratic members. Light green block means independent members.*

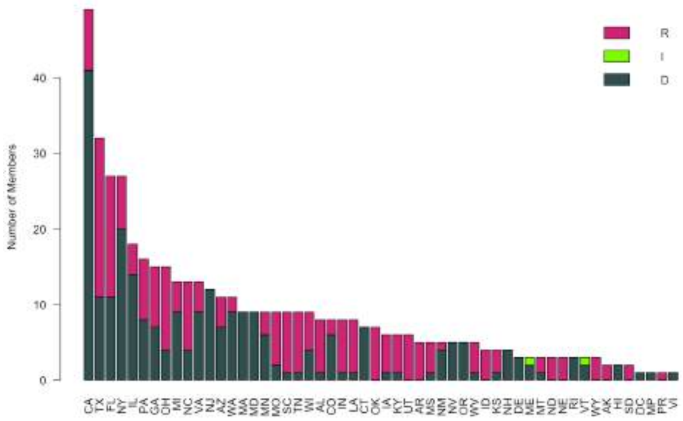


Figure 1.3. Distribution of Congress Members by State and Party. *Each bar represents the number of Congress members of a state, totaling 54 states. Deep pink block indicates Republican members. Dark green block represents the Democratic members. Light green block means independent members.*

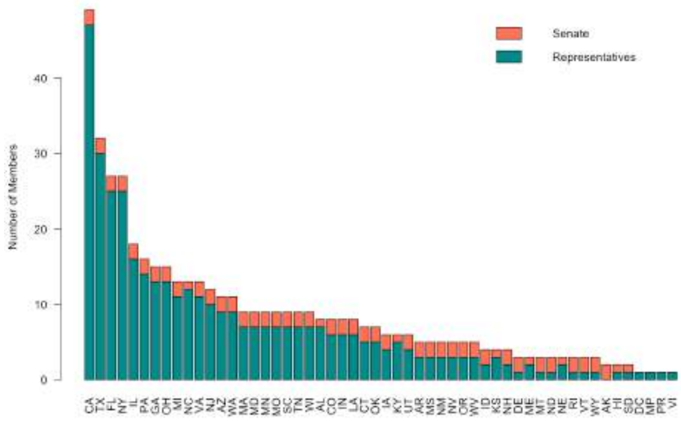


Figure 1.4. Distribution of Congress Members by State and House*. Each bar represents the number of Congress members of a state, totaling 54 states. Orange block represents the chamber of Senate, while deep green block indicates the house of Representatives.*

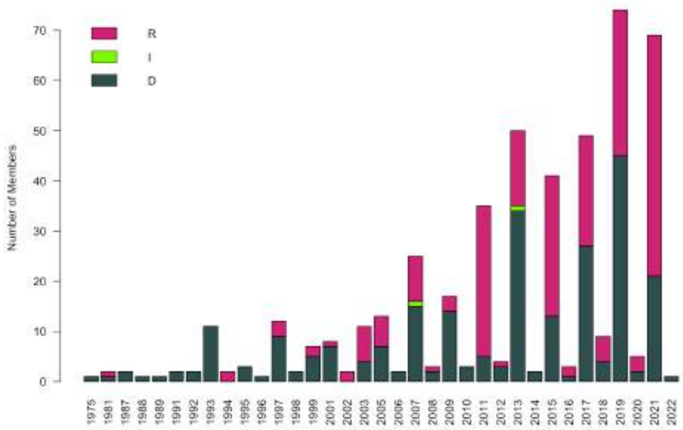


Figure 1.5. Distribution of Congress Members by Seniority and Party. *Each bar represents the year in which a Congress member first joined Congress. Deep pink block indicates Republican members. Dark green block represents the Democratic members. Light green block means independent members.*

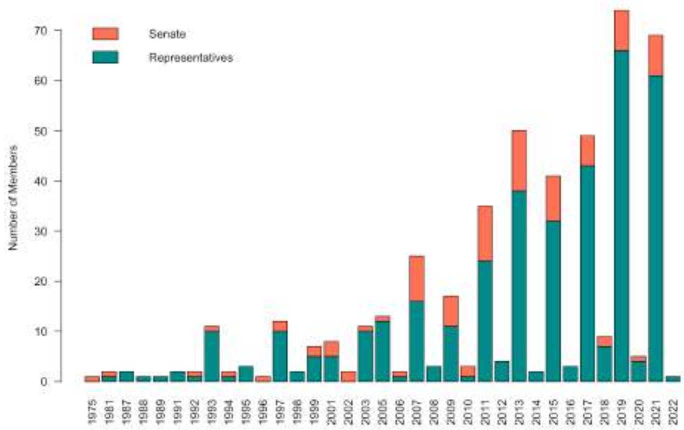


Figure 1.6. Distribution of Congress Members by Seniority and House. *Each bar represents the year in which a Congress member first joined Congress. Orange block represents the chamber of Senate, while deep green block indicates the house of Representatives.*

4.2. Descriptive Methods of Social Network Analysis

Yang et al. (2017) argued that without the use of statistical tools, comprehending even a small network can be a tricky question due to its complex connections and configurations. Descriptive methods are a crucial aspect of social network analysis, measuring various metrics. Some definitions and formulas of key metrics analyzed in the study are explained in detail below, focusing on the directed network:

Density is the ratio of actual connections and the total possible connections in a network, expressing how connected the network is, which is calculated as (Yang et al., 2017):

$$Density = \frac{\sum\sum X_{i,j}}{N*(N-1)}$$

Here, the $\sum\sum X_{i,j}$ is the total number of actual connections, while $N*(N-1)$ represents the maximum possible connections with N nodes. It is notable that the equations of directed and undirected are equivalent since the factor ½ of the numerator and denominator in the undirected graph cancel out.

Diameter measures the longest shortest path between any two individuals within the network (Ognyanova, 2016), exposing how far apart the most distant nodes are, giving insight into overall reachability and the efficiency in spreading information or influence. Similarly, **distances and paths** reveal how connected or closely linked specific nodes are. Yang et al. (2017) exhibited the "six degrees of separation" concept, or the small-world phenomenon, where anyone can be connected to anyone else through a small number of intermediaries—typically five or fewer. By comparing mean distances, the connection efficiency within a network can be assessed, while analysing paths helps identify the neighbourhoods of key members.

Centrality measures assess the importance of individual nodes in a network from different perspectives. The key types of centralities include:

Degree centrality is the simplest centrality measure, counting the number of direct connections a node has, with a higher degree implying a more important node which for example has more friends (Freeman, 1978). In directed networks, degree centrality can be classified into in-degree (incoming connections), out-degree (outgoing connections), and all-degree. Wasserman and Faust (1994) introduced a standardized formula to compare degree centrality across different scaled graphs:

$$C_D'(N_i) = \frac{C_D(N_i)}{g-1} = \frac{\sum_{j=1}^{g} x_{i,j}\ (i \neq j)}{g-1}$$

Where the numerator is the accumulation of connections a node has, while the denominator is the total number of connections it could have with g actors. **Degree centralization** measures the inequality in centrality across nodes (Yang et al., 2017):

$$C_D = \frac{\sum_{i=1}^{N}(C_D(N^*) - C_D(N_i))}{(N-1)(N-2)}$$

Where the numerator adds up the difference in degree between the actor with the highest degree and each of the other actors. The denominator represents the total number of potential connections that could be formed within the network to normalize the centralization. This value ranges from 0 (equal centrality) to 1 (high inequality).

Betweenness centrality defined by Borgatti et al. (2009) captures how often a node lies on the shortest path between other nodes, revealing its potential to control information flow. For instance, the Medici family become a central hub and business broker in communication and political fields in 15th-century Florence due to their high betweenness centrality (Padgett and Ansell, 1993).

Wasserman and Faust (1994) also provided a normalized formula for betweenness centrality in directed networks:

$$C'_B(N_i) = \frac{C_B(N_i)}{(g-1)(g-2)} = \frac{\sum_{j<k} \frac{g_{j,k}(N_i)}{g_{j,k}}}{(g-1)(g-2)}$$

Here, the numerator sums up the frequency with which node i lies on the shortest paths between all other pairs of nodes in a network, while $(g-1)(g-2)$ denotes the possible maximum betweenness centrality of a node i. **Betweenness centralization**, which measures disparity in betweenness centrality across nodes, is given by (Yang et al., 2017):

$$C_B = \frac{\sum_{i=1}^{g}(C_B(N^*) - C_B(N_i))}{(g-1)^2(g-2)}$$

Here, the numerator represents the sum of the differences between the highest betweenness centrality and the centrality of other nodes, whereas the denominator denotes the theoretical maximum value given g nodes, with one node having the maximum betweenness centrality and all others having 0. This ranges from 0 (classless network) to 1 (highly hierarchical).

Eigenvector centrality measures the influence of a node based on its connections to other influential nodes. A node with high eigenvector centrality is well-connected to other high-centrality nodes. Eigenvector centralization assesses this influence across the entire network (Newman, 2008). Moreover, **closeness centrality** reflects how quickly a node can reach all other nodes. The maximum value of 1 means a node is directly connected to all others, while values close to 0 indicate fewer connections. If a node has no edges, its closeness centrality is undefined (Yang et al., 2017). Closeness centralization measures how evenly closeness centrality is distributed across nodes. Only some of the mathematical expressions for centrality and centralization are presented here, not the others as they are similar, but focus on capturing the different properties of centrality, and in-, out-, and all-directionality in specific applications, refer to the reference for a more detailed mathematical explanation. Furthermore, Ognyanova (2016) introduced that the **hubs and authorities** algorithm originally designed by Jon Kleinberg for web page analysis, identifies hubs as nodes with many outgoing links and authorities as nodes receiving numerous links from hubs, indicating their relevance and quality. Hubs and authorities are analogous to out-degree and in-degree centrality, respectively, but emphasise connectivity to important nodes, providing different perspectives of out- and in-connections. These centrality measures and their distributions help understand the roles of nodes in a network, revealing key players and the overall structure.

Reciprocity measures the likelihood of bidirectional connections in a network. Specifically, it indicates the probability that if node A is connected to node B, node B will reciprocate the connection to node A, which is used to analyse interaction patterns exclusively in directed networks. Aric A. Hagberg et al. (2008) defined reciprocity as the proportion of connections that are mutual, where both directions of the link exist.

Homophily of node attributes measures the tendency of individuals to connect with others who share similar attributes (Ognyanova, 2016). Yang et al. (2017) observed that homophily is prevalent and a dyadic level term in social networks, meaning that people with similar characteristics are more likely to establish connections. In addition, Newman and Park, (2003) pointed out a fundamental distinction of degree correlations between social networks and other types of networks, with positive correlations (assortative mixing) in most social networks and negative correlations (disassortative mixing) in most non-social networks. For example, in a social network, a person with many friends (a high-degree node) is more likely to connect with another person who also has many friends. In contrast, in a network like the Internet, a router with many connections is usually linked to routers with fewer connections. Thus, assessing assortativity in node attributes and degrees can help determine if the Congressional Twitter network follows this typical social pattern.

Transitivity (clustering coefficient) measures how often nodes in a network form triangles or tight-knit groups. For instance, if A and B both know C, transitivity assesses the likelihood that A and B also become connected, forming a cluster if all three know each other. Newman and Park, (2003) noted that social networks generally have higher transitivity compared to random networks. This means that in social networks, nodes are more likely to form dense clusters or communities. In contrast, non-social networks, such as food webs or the Internet, often show transitivity similar to random models. Measuring transitivity can investigate if the Congressional Twitter network exhibits the high clustering typical of social networks.

Nevertheless, Newman and Park, (2003) suggested that the distinctions between social networks and other types of networks might be attributed to the strong community structures found in social networks, in which social networks often have both high clustering coefficients and positive degree correlations, unlike non-social networks which lack such community features. To explore and identify these community structures, three main methods are commonly used:

Luce and Perry (1949) defined a **clique** as a group of nodes where every node directly connects to every other node. The largest cliques are rare because all nodes must be fully connected (Wasserman and Faust, 1994). The N-clique relaxes the condition, accepting indirect connections within N steps to generalise the rigid definition. However, the basic clique algorithm used here is only eligible for undirected graphs and focuses on finding the largest cliques.

Seidman (1983) published the **K-core decomposition** as an alternative method to detect subgroups, identifying subgroups where each node connects to at least K other nodes. Higher K-values indicate a more restricted subgroup.

Newman and Girvan (2004) argued that community structure is a common feature of most networks, where connections within groups are dense and sparse between groups. Identifying and interpreting these communities is significantly beneficial to understanding the overall pattern of networks. Here, four algorithms of **community detection** are employed: Edge Betweenness, Propagating Labels, Greedy Optimization of Modularity, and Random Walk.

- Community detection based on edge betweenness developed by Newman and Girvan, iteratively removes edges with high betweenness (edges that frequently occur on the shortest paths between nodes) and recalculates betweenness until the optimal division of the network is found (Ognyanova, 2016).
- Aric A. Hagberg et al. (2008) described that the propagating labels algorithm randomly assigns a unique label to each node and then repeatedly replaces each label with the most common label among its neighbours. This process continues until all nodes share the most frequent label, forming communities.
- Newman and Girvan (2004) defined that modularity measures the quality of a particular partition of a network to answer how good the structure detected is, since there will be meaningless community structure even in a random network, with higher modularity meaning more connections within communities and fewer between them. According to Aric A. Hagberg et al. (2008), the greedy modularity method merges communities until the highest modularity is reached.

- Jayawickrama (2021) suggested that the random walks approach identifies communities by tracking random walks through the network. Walks tend to get "trapped" in densely connected areas, helping to detect and merge communities.

4.3. Exponential Random Graph Models (ERGMs)

Descriptive methods alone are limited in exploring the network formation and dynamics. To test whether observed features are statistically significant, Exponential Random Graph Models (ERGMs) can be used. ERGMs, as described by Jones et al. (2020), are analogous to logistic regression, predicting the likelihood of connection formation between pairs of nodes by comparing the observed network to a randomly generated network which is the "null hypothesis" of the ERGM and referred to as the "Erdős-Rényi" network. The dependency of network relationships runs counter to the basic assumption of traditional regression, and regarding individuals as independent in the social network is pointless. To address these issues, ERGMs can model these special network dyadic and higher dimensional dependencies by simulating network configurations. The general ERGM formula is:

$$P(Y = y|X) = \frac{exp\,\{\theta^T g(y, X)\}}{k(\theta)}$$

- $P(Y = y|X)$ represents the probability that network Y forms a specific structure y, given the condition in the covariate matrix X.
- $\theta^T g(y, X)$ represents the product of a vector of coefficients and a vector of statistics based on the network y and covariate X. The exponential function ensures the probability is positive.
- $k(\theta)$ is a normalizing constant that sums the numerator over all possible networks Y, ensuring all probabilities sum to 1. Since the number of possible networks grows rapidly with the number of nodes, calculating this constant is complex and often requires approximation methods.

Krivitsky et al. (2024) transformed the numerator into a logarithmic form, creating a form similar to traditional linear models:

$$\log(ex\,p\{\theta^T g(y, X)\}) = \theta_1 g_1(y) + \theta_2 g_2(y) + \cdots \theta_p g_p(y)$$

Here, the coefficients θ impact the strength and direction of the covariates $g(y)$, when θs approach zero, the ERGM behaves like a uniform random network, with no specific structure or pattern.

Krivitsky et al. (2024) explained that ERMGs can incorporate covariates into network models, estimate parameters using maximum likelihood methods, evaluate model convergence, test the statistical significance of covariates, interpret coefficients, assess goodness-of-fit, and simulate new networks based on the distributions of the fitted model. According to Krivitsky et al. (2024), the terms $g(y)$ in ERGMs act like covariates in traditional models but represent network features such as node attributes or reciprocity. Notably, some terms capture the frequency of specific configurations at the dyadic level, calculated from the collected data of the observed graph rather than obtained from external sources. Krivitsky et al. (2024) also noted that another notable property of ERGMs is that terms can be categorized as:

- **Dyad-independent**: These, like nodal homophily, depend only on the individual node itself and can be estimated using Maximum Pseudolikelihood Estimation (MPLE).
- **Dyad-dependent**: These involve more complex relationships, such as reciprocity or transitivity, and require Markov Chain Monte Carlo (MCMC) methods for estimation due to their non-linear dependencies, which can sometimes be counterintuitive.

When a model includes dyad-dependent terms, MCMC is used to estimate parameters. Here is a brief overview of how it works: (1) Start with initial parameter estimates, (2) Randomly pick node pairs and flip a coin to decide whether they are connected, (3) Repeat the simulation thousands of times, (4) Calculate key statistics of interest, (5) Compare statistics between the simulated and real networks, adjusting parameters as needed, and (6) Continuing until the simulated network closely matches the real one.

Since this study incorporated dyad-dependent terms, running MCMC diagnostics is crucial to ensure the model has converged before explaining the model results and evaluating the goodness of fit (Krivitsky et al., 2024). According to Jones et al. (2020), key criteria of diagnostics include:

- **Sample statistic auto-correlation** measures the correlation between sample statistics at different points in the MCMC chain. A well-mixed chain should show low autocorrelation (close to 0) after Lag 0 in the results.
- **Burn-in diagnostic (Geweke)** compares the means of sample statistics from different positions within the chain. P-values close to 1 for all individual statistics signify good convergence.
- **MCMC trace plots** show differences between sample statistics and the observed network at each simulation step, ideally demonstrating proper mixing centred around 0.
- **MCMC density plots** should display a normal distribution centred around 0, implying no significant deviation from the observed network.

Krivitsky et al. (2024) cautioned that MCMC chains may diverge from the observed network if the model does not represent it well, leading to model degeneracy. Jones et al. (2020) also noted that models frequently fail to converge, requiring parameter adjustments or additional terms.

To test statistical significance, the p-values need to be checked, similar to traditional statistics. If the p-value is below 0.5%, the term is considered significant. To translate statistic coefficients, for example, a positive coefficient θ_1 suggests that the covariate $g_1(y)$ makes ties more likely than by chance. Once the ERGM coefficients are estimated, the model specifies the probability distribution of all possible networks given the set of nodes. If the model fits the data well, the simulated networks will resemble the observed data, which is also a method to assess the model fit, as described by Krivitsky et al. (2024). Additionally, Krivitsky et al. (2024) explained that ERGMs describe how the overall network structure is formed through local interactions between nodes, which are represented by specific term configurations. To test how well an ERGM fits the data, goodness-of-fit compares certain statistics from the original network with those generated by the model to see if it can reproduce global network properties that were not directly included in the model. It is expected to be observed that real statistics fall within the range of simulated values (Jones et al., 2020). Otherwise, additional terms might be needed to capture the network complexity. Consequently, after this series of ERGM modelling, interpreting and evaluating, the underlying formation and dynamic structure of the network can be revealed.

5. Analysis, Results and Discussion

5.1. Descriptive Network Analysis and Visualization

Density

Network	***Net_all_weighted***	***Net_all***	***Net_moderate***	***Net_active***
If weighted	*Yes*	*No*	*No*	*No*
If directed	*Yes*	*Yes*	*Yes*	*Yes*
Number of Vertices	*475*	*475*	*475*	*456*
Number of Edges	*13289*	*13289*	*11263*	*2008*
Density	*0.059*	*0.059*	*0.050*	*0.010*

Table 1. *Basic Information and Density Comparison of Networks. This table summarizes whether they are weighted and directed graphs, number of nodes, number of edges, and network density. These networks are Net_all-weighted, Net_all, Net_roderate, and Net_active, respectively.*

The variation in the number of vertices, edges, and density highlights the basic characteristics of the networks based on different strengths of connections. Table 1 shows that Net_all_weighted, Net_all, and Net_moderate each have 475 nodes, as they include the same participants. In contrast, Net_active has 456 nodes because some participants did not engage in strong interactions and were excluded. Despite the significant drop in the number of edges from 13,289 to 2,008, the number of nodes only decreased by 19, indicating that most members still maintain strong interaction to some extent. Density is unaffected by weights, simply counting the number of ties, thus Net_all_weighted and Net_all both have a density of 0.059. Among the networks, density decreases significantly from 0.059 to 0.010, suggesting that most nodes do not densely connect to each other with strong connections. The network structure evolves from a tight network coverage to a sparse network dominated.

Diameter

Network	***Net_all_weighted***	***Net_all***	***Net_moderate***	***Net_active***
Diameter	*0.031*	*6*	*6*	*18*
Vertex sequence	*93 -410-109-264-395*	*437-318-62-327-195-264-395*	*102-28-15-9-109-264-395*	*1-51-78-57-12-84-59-263-177-55-90-71-25-61-50-35-5-77-63*
Mean Distance	*0.006*	*2.357*	*2.521*	*6.011*
Normalized Average Distance	*0.208*	*0.393*	*0.420*	*0.334*

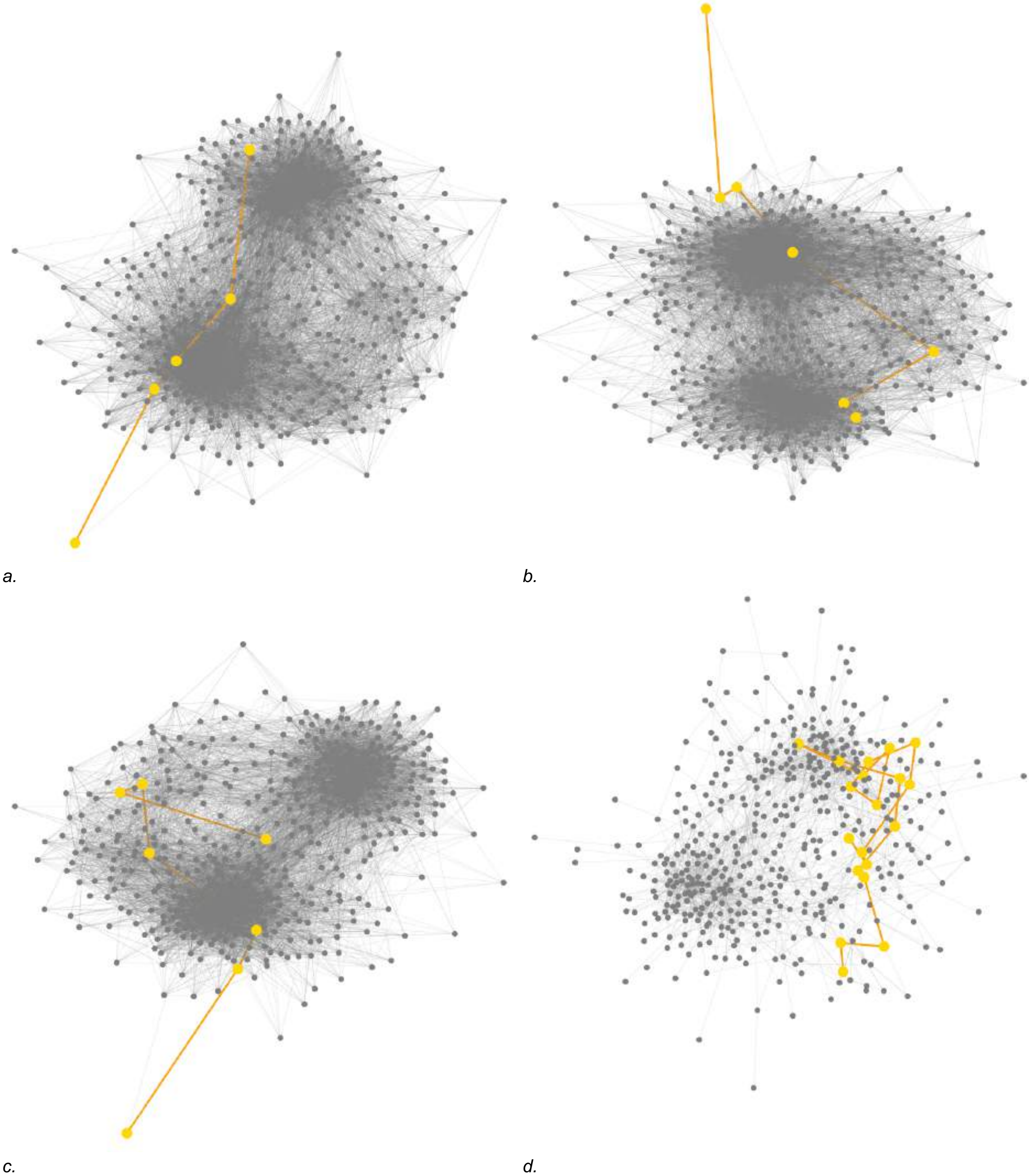
Table 2. *Comparison of Diameters, Vertex Sequences, Average Distances, and Normalized Average Distances of Networks. This table shows that as edges are removed, the network becomes sparser and metrics such as diameter increase accordingly. Normalized Average Distance is obtained by dividing Mean Distance by diameter. If this value is close to 1, it indicates that the difference between the average distance and diameter is small, which means that the distances between nodes in the network are evenly distributed, and vice versa.*

Table 2 shows that the average path length is 2.357 in the Net_all, meaning most nodes are connected in about 2 to 3 steps. In the Net_moderate, the average path length increases to 2.521. With the removal of weak connections, the network becomes slightly looser, but nodes are still connected in about 3 steps. In the Net_active, the average path length rises significantly to 6.011, indicating that the network is much sparser, and it takes more steps to connect nodes. This makes spreading information more challenging in the strong connections.

For Net_all_weighted, where direct comparisons are not possible, the normalized average distance is used. Values of 0.208 and 0.393 reveal significant differences in how paths are distributed, suggesting that the original weighted network is hierarchical and uneven. From Net_all to Net_moderate, the diameter remains at 6, but the normalized average distance increases from

0.393 to 0.42. This increase implies that after removing weak connections, some paths between nodes have become longer, indicating that some members have only weak interactions on Twitter. Although the normalized average distance decreases from 0.420 to 0.334 from Net_moderate to Net_active, the diameter increases from 6 to 18, making it harder to identify a clear pattern.

Figure 2 demonstrates that the longest geodesic distance in Net_all_weighted is shortest, indicating weights may make certain paths more efficient. Despite the removal of weak connections from Net_all to Net_moderate, the longest geodesic distance remains at 6. In contrast, the longest geodesic distance in Net_active increases significantly to 18, passing through multiple sparsely connected nodes, highlighting the looseness of the network in strong connections.



a. b. c. d.

Figure 2. Longest Geodesic Distance in the Networks. These figures display the longest shortest path between any two individuals within the four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Yellow circles represent the nodes along the path, while orange line indicates the longest path. These figures illustrate that in the original weighted network, the longest path is relatively short. However, as weights are filtered and connections are reduced, the longest geodesic distance of the network increases, especially in the Net_active.

Degree Distribution

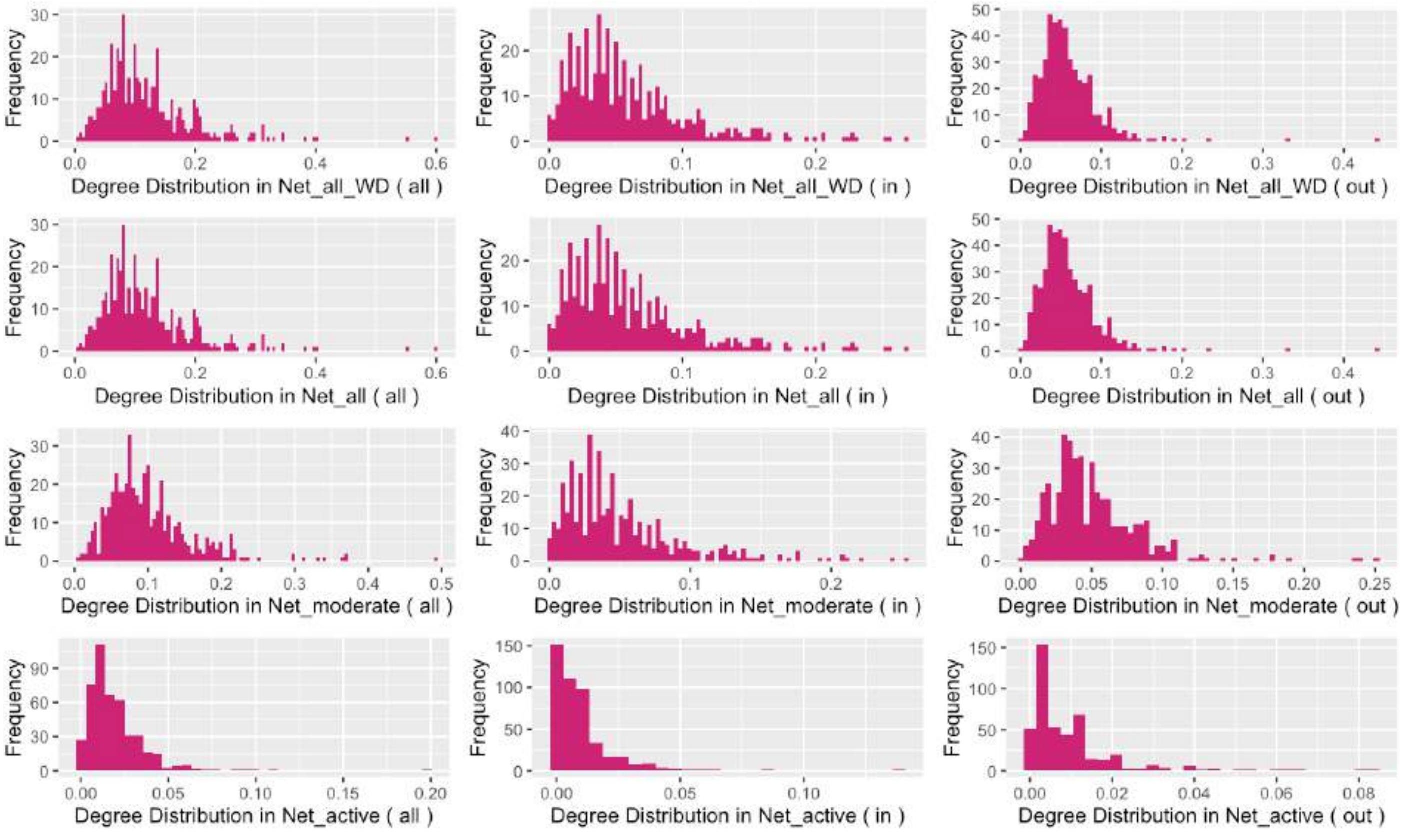


Figure 3. Normalized Degree Distributions of Networks. *The figure illustrates the normalized degree distributions for four networks across different conditions: weighted (Net_all_WD), unweighted (Net_all), moderate and active connections (Net_moderate), and active connections (Net_active). Each row corresponds to one network, with the columns showing the overall degree distribution, in-degree, and out-degree. The distributions for Net_all_weighted and Net_all are the same due to the nature of degree centrality. Net_all_weighted, Net_all and Net_moderate show similar patterns, where most nodes have a moderate number of connections, and only a few have many. Net_active, however, follows a power-law distribution typical of social networks, where most nodes have very few connections.*

Figure 3 displays the degree distributions for each network. The distributions for Net_all_weighted and Net_all are identical because degree centrality measures the presence of connections, rather than their strength. Both networks, along with Net_moderate, display a pattern where most members have a moderate number of connections, with only a few having many connections, implying that while most members are engaged in interactions, only a few are central figures. In contrast, Net_active follows a power-law distribution typical of social networks, showing a long-tail pattern, where most nodes have very few connections, denoting that the network of strong connections is sparse and hieratical. Notably, in Net_all_weighted and Net_all, a few key members have many more outgoing connections than incoming ones, suggesting they are more active in spreading information through weak connections. Conversely, in the Net_active, some key members have many more incoming connections than outgoing ones, indicating they receive more information through strong connections.

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.599**
367	**Nancy Pelosi**	**High**	**0.551**
393	Bobby L. Rush	High	0.401
254	Steny H. Hoyer	High	0.395
208	**C. Scott Franklin**	**Low**	**0.384**
269	Mike Johnson	Medium	0.344
303	Andy Levin	Low	0.344
190	Jeff Duncan	Medium	0.331
399	Steve Scalise	High	0.321
111	Don Beyer	Medium	0.312

All-degree Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.599**
367	**Nancy Pelosi**	**High**	**0.551**
393	Bobby L. Rush	High	0.401
254	Steny H. Hoyer	High	0.395
208	**C. Scott Franklin**	**Low**	**0.384**
269	Mike Johnson	Medium	0.344
303	Andy Levin	Low	0.344
190	Jeff Duncan	Medium	0.331
399	Steve Scalise	High	0.321
111	Don Beyer	Medium	0.312

All-degree Centrality in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.492**
208	**C. Scott Franklin**	**Low**	**0.371**
393	Bobby L. Rush	High	0.367
254	Steny H. Hoyer	High	0.365
367	**Nancy Pelosi**	**High**	**0.342**
269	Mike Johnson	Medium	0.333
399	Steve Scalise	High	0.314
190	Jeff Duncan	Medium	0.3
192	Tom Emmer	Medium	0.297
385	John W. Rose	Low	0.251

All-degree Centrality in Net_moderate

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.2**
399	Steve Scalise	High	0.11
208	**C. Scott Franklin**	**Low**	**0.099**
269	Mike Johnson	Medium	0.092
470	Joe Wilson	High	0.09
393	Bobby L. Rush	High	0.079
389	Chip Roy	Low	0.07
226	Bob Good	Low	0.068
367	**Nancy Pelosi**	**High**	**0.066**
147	Sean Casten	Low	0.062

All-degree Centrality in Net_active

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.268**
208	**C. Scott Franklin**	**Low**	**0.255**
190	Jeff Duncan	Medium	0.253
111	Don Beyer	Medium	0.23
254	Steny H. Hoyer	High	0.228
385	John W. Rose	Low	0.228
269	Mike Johnson	Medium	0.224
192	Tom Emmer	Medium	0.222
147	Sean Casten	Low	0.205
303	Andy Levin	Low	0.205

In-degree Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.268**
208	**C. Scott Franklin**	**Low**	**0.255**
190	Jeff Duncan	Medium	0.253
111	Don Beyer	Medium	0.23
254	Steny H. Hoyer	High	0.228
385	John W. Rose	Low	0.228
269	Mike Johnson	Medium	0.224
192	Tom Emmer	Medium	0.222
147	Sean Casten	Low	0.205
303	Andy Levin	Low	0.205

In-degree Centrality in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.253**
208	**C. Scott Franklin**	**Low**	**0.243**
190	Jeff Duncan	Medium	0.222
269	Mike Johnson	Medium	0.213
385	John W. Rose	Low	0.209
192	Tom Emmer	Medium	0.207
254	Steny H. Hoyer	High	0.198
111	Don Beyer	Medium	0.192
92	Alma Adams	Medium	0.177
303	Andy Levin	Low	0.177

In-degree Centrality in Net_moderate

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.138**
269	Mike Johnson	Medium	0.084
389	Chip Roy	Low	0.064
147	Sean Casten	Low	0.057
215	Jesús "Chuy" García	Low	0.053
208	**C. Scott Franklin**	**Low**	**0.048**
318	Thomas Massie	Medium	0.048
335	Mariannette Miller-Meeks	Low	0.048
113	Andy Biggs	Medium	0.044
401	Jan Schakowsky	High	0.044

In-degree Centrality in Net_active

ID	Name	Seniority	Centrality
367	**Nancy Pelosi**	**High**	**0.443**
322	**Kevin McCarthy**	**High**	**0.331**
393	Bobby L. Rush	High	0.234
71	Charles E. Schumer	High	0.205
399	Steve Scalise	High	0.188
436	Mark Takano	Medium	0.179
179	Rosa L. DeLauro	High	0.177
254	Steny H. Hoyer	High	0.167
105	Jim Banks	Medium	0.158
87	Elizabeth Warren	Medium	0.15

Out-degree Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
367	**Nancy Pelosi**	**High**	**0.443**
322	**Kevin McCarthy**	**High**	**0.331**
393	Bobby L. Rush	High	0.234
71	Charles E. Schumer	High	0.205
399	Steve Scalise	High	0.188
436	Mark Takano	Medium	0.179
179	Rosa L. DeLauro	High	0.177
254	Steny H. Hoyer	High	0.167
105	Jim Banks	Medium	0.158
87	Elizabeth Warren	Medium	0.15

Out-degree Centrality in Net_all

ID	Name	Seniority	Centrality
367	**Nancy Pelosi**	**High**	**0.251**
322	**Kevin McCarthy**	**High**	**0.238**
393	Bobby L. Rush	High	0.234
399	Steve Scalise	High	0.188
436	Mark Takano	Medium	0.179
179	Rosa L. DeLauro	High	0.177
254	Steny H. Hoyer	High	0.167
87	Elizabeth Warren	Medium	0.15
298	Barbara Lee	High	0.143
422	Victoria Spartz	Low	0.131

Out-degree Centrality in Net_moderate

ID	Name	Seniority	Centrality
470	Joe Wilson	High	0.084
399	Steve Scalise	High	0.081
226	Bob Good	Low	0.066
393	Bobby L. Rush	High	0.064
322	**Kevin McCarthy**	**High**	**0.062**
436	Mark Takano	Medium	0.059
367	**Nancy Pelosi**	**High**	**0.055**
456	Ann Wagner	Medium	0.055
208	**C. Scott Franklin**	**Low**	**0.051**
328	Patrick T. McHenry	High	0.046

Out-degree Centrality in Net_active

ID	Name	Seniority	Centrality
111	Don Beyer	Medium	0.085
322	**Kevin McCarthy**	**High**	**0.071**
17	John Cornyn	High	0.068
428	Elise Stefanik	Medium	0.062
147	Sean Casten	Low	0.05
367	**Nancy Pelosi**	**High**	**0.045**
303	Andy Levin	Low	0.044
254	Steny H. Hoyer	High	0.041
113	Andy Biggs	Medium	0.04
215	Jesús "Chuy" García	Low	0.038

Betweenness Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.073**
367	**Nancy Pelosi**	**High**	**0.041**
254	Steny H. Hoyer	High	0.031
393	Bobby L. Rush	High	0.027
335	Mariannette Miller-Meeks	Low	0.018
208	**C. Scott Franklin**	**Low**	**0.017**
17	John Cornyn	High	0.016
303	Andy Levin	Low	0.016
461	Bonnie Watson Coleman	Medium	0.014
113	Andy Biggs	Medium	0.013

Betweenness Centrality in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.058**
254	Steny H. Hoyer	High	0.044
393	Bobby L. Rush	High	0.032
367	**Nancy Pelosi**	**High**	**0.027**
208	**C. Scott Franklin**	**Low**	**0.026**
399	Steve Scalise	High	0.019
192	Tom Emmer	Medium	0.018
269	Mike Johnson	Medium	0.017
87	Elizabeth Warren	Medium	0.016
143	Buddy Carter	Medium	0.016

Betweenness Centrality in Net_moderate

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.188**
208	**C. Scott Franklin**	**Low**	**0.093**
470	Joe Wilson	High	0.086
215	Jesús "Chuy" García	Low	0.076
171	Jason Crow	Low	0.072
236	Raul Grijalva	High	0.064
367	**Nancy Pelosi**	**High**	**0.059**
375	Katie Porter	Low	0.055
298	Barbara Lee	High	0.053
331	Peter Meijer	Low	0.052

Betweenness Centrality in Net_active

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
389	Chip Roy	Low	0.711
269	Mike Johnson	Medium	0.677
246	Jody Hice	Medium	0.569
113	Andy Biggs	Medium	0.511
335	Mariannette Miller-Meeks	Low	0.446
188	Byron Donalds	Low	0.418
208	**C. Scott Franklin**	**Low**	**0.416**
161	Andrew S. Clyde	Low	0.391
116	Dan Bishop	Low	0.385

Eigenvector Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
190	Jeff Duncan	Medium	0.971
208	**C. Scott Franklin**	**Low**	**0.919**
269	Mike Johnson	Medium	0.902
385	John W. Rose	Low	0.861
192	Tom Emmer	Medium	0.836
389	Chip Roy	Low	0.772
188	Byron Donalds	Low	0.758
324	Lisa C. McClain	Low	0.727
113	Andy Biggs	Medium	0.722

Eigenvector Centrality in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
269	Mike Johnson	Medium	0.934
190	Jeff Duncan	Medium	0.933
208	**C. Scott Franklin**	**Low**	**0.887**
385	John W. Rose	Low	0.807
192	Tom Emmer	Medium	0.806
113	Andy Biggs	Medium	0.732
188	Byron Donalds	Low	0.717
324	Lisa C. McClain	Low	0.704
244	Kevin Hern	Medium	0.699

Eigenvector Centrality in Net_moderate

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
269	Mike Johnson	Medium	0.903
113	Andy Biggs	Medium	0.685
389	Chip Roy	Low	0.67
246	Jody Hice	Medium	0.549
271	Jim Jordan	High	0.533
116	Dan Bishop	Low	0.435
318	Thomas Massie	Medium	0.405
428	Elise Stefanik	Medium	0.38
245	Yvette Herrell	Low	0.375

Eigenvector Centrality in Net_active

ID	Name	Seniority	Centrality
17	John Cornyn	High	0.956
111	Don Beyer	Medium	0.907
147	Sean Casten	Low	0.889
126	Anthony Brown	Medium	0.859
322	**Kevin McCarthy**	**High**	**0.855**
54	Bob Menendez	High	0.85
389	Chip Roy	Low	0.843
149	Joaquin Castro	Medium	0.829
193	Veronica Escobar	Low	0.822
89	Roger F. Wicker	High	0.81

All-closeness Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
367	**Nancy Pelosi**	**High**	**0.646**
322	**Kevin McCarthy**	**High**	**0.617**
254	Steny H. Hoyer	High	0.585
111	Don Beyer	Medium	0.571
208	**C. Scott Franklin**	**Low**	**0.57**
393	Bobby L. Rush	High	0.57
461	Bonnie Watson Coleman	Medium	0.562
92	Alma Adams	Medium	0.558
190	Jeff Duncan	Medium	0.558
71	Charles E. Schumer	High	0.556

All-closeness Centrality in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.59**
367	**Nancy Pelosi**	**High**	**0.577**
254	Steny H. Hoyer	High	0.576
208	**C. Scott Franklin**	**Low**	**0.56**
393	Bobby L. Rush	High	0.557
269	Mike Johnson	Medium	0.547
111	Don Beyer	Medium	0.545
192	Tom Emmer	Medium	0.545
335	Mariannette Miller-Meeks	Low	0.538
92	Alma Adams	Medium	0.537

All-closeness Centrality in Net_moderate

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.416**
470	Joe Wilson	High	0.391
111	Don Beyer	Medium	0.38
208	**C. Scott Franklin**	**Low**	**0.377**
399	Steve Scalise	High	0.375
456	Ann Wagner	Medium	0.375
367	**Nancy Pelosi**	**High**	**0.373**
269	Mike Johnson	Medium	0.369
393	Bobby L. Rush	High	0.369
335	Mariannette Miller-Meeks	Low	0.366

All-closeness Centrality in Net_active

ID	Name	Seniority	Centrality
393	Bobby L. Rush	High	0.403
242	Vicky Hartzler	Medium	0.399
350	Troy E. Nehls	Low	0.397
126	Anthony Brown	Medium	0.397
370	August Pfluger	Low	0.396
111	Don Beyer	Medium	0.396
147	Sean Casten	Low	0.396
442	Dina Titus	Medium	0.395
461	Bonnie Watson Coleman	Medium	0.395
421	Abigail Spanberger	Low	0.392

In-closeness Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.562**
111	Don Beyer	Medium	0.554
208	**C. Scott Franklin**	**Low**	**0.547**
335	Mariannette Miller-Meeks	Low	0.542
269	Mike Johnson	Medium	0.536
92	Alma Adams	Medium	0.534
188	Byron Donalds	Low	0.531
461	Bonnie Watson Coleman	Medium	0.531
192	Tom Emmer	Medium	0.531
389	Chip Roy	Low	0.528

In-closeness Centrality in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**0.531**
111	Don Beyer	Medium	0.528
208	**C. Scott Franklin**	**Low**	**0.524**
335	Mariannette Miller-Meeks	Low	0.513
92	Alma Adams	Medium	0.512
269	Mike Johnson	Medium	0.511
461	Bonnie Watson Coleman	Medium	0.506
192	Tom Emmer	Medium	0.502
102	Don Bacon	Medium	0.496
350	Troy E. Nehls	Low	0.496

In-closeness Centrality in Net_moderate

ID	Name	Seniority	Centrality
16	Christopher Coons	Medium	1
20	Kevin Cramer	Low	1
22	Ted Cruz	Medium	1
38	John Hoeven	Medium	1
41	Ron Johnson	Medium	1
91	Todd Young	Medium	1
94	Pete Aguilar	Medium	1
124	Kevin Brady	High	1
133	Tim Burchett	Low	1
180	Suzan DelBene	Medium	1

In-closeness Centrality in Net_active

ID	Name	Seniority	Centrality
17	John Cornyn	High	0.888
3	Marsha Blackburn	Low	0.738
88	Sheldon Whitehouse	High	0.735
149	Joaquin Castro	Medium	0.727
32	Bill Hagerty	Low	0.706
367	**Nancy Pelosi**	**High**	**0.689**
22	Ted Cruz	Medium	0.683
428	Elise Stefanik	Medium	0.678
263	Pramila Jayapal	Medium	0.676
111	Don Beyer	Medium	0.673

Out-closeness Centrality in Net_all_weighted

ID	Name	Seniority	Centrality
367	**Nancy Pelosi**	**High**	**0.639**
322	**Kevin McCarthy**	**High**	**0.577**
71	Charles E. Schumer	High	0.537
393	Bobby L. Rush	High	0.525
254	Steny H. Hoyer	High	0.521
87	Elizabeth Warren	Medium	0.502
17	John Cornyn	High	0.5
315	Carolyn Maloney	High	0.498
179	Rosa L. DeLauro	High	0.493
436	Mark Takano	Medium	0.493

Out-closeness Centrality in Net_all

ID	Name	Seniority	Centrality
367	**Nancy Pelosi**	**High**	**0.544**
254	Steny H. Hoyer	High	0.509
393	Bobby L. Rush	High	0.505
322	**Kevin McCarthy**	**High**	**0.502**
87	Elizabeth Warren	Medium	0.478
179	Rosa L. DeLauro	High	0.477
436	Mark Takano	Medium	0.474
399	Steve Scalise	High	0.471
422	Victoria Spartz	Low	0.466
27	Dianne Feinstein	High	0.465

Out-closeness Centrality in Net_moderate

ID	Name	Seniority	Centrality
22	Ted Cruz	Medium	1
69	Marco Rubio	Medium	1
88	Sheldon Whitehouse	High	1
143	Buddy Carter	Medium	1
144	John Carter	High	1
165	Gerald E. "Gerry" Connolly	Medium	1
176	Madeleine Dean	Low	1
180	Suzan DelBene	Medium	1
194	Anna A. Eshoo	High	1
254	Steny H. Hoyer	High	1

Out-closeness Centrality in Net_active

ID	Name	Seniority	Centrality
226	Bob Good	Low	1
159	Michael Cloud	Medium	0.859
399	Steve Scalise	High	0.842
164	James Comer	Medium	0.699
129	Vern Buchanan	High	0.638
220	Louie Gohmert	High	0.611
118	Lauren Boebert	Low	0.592
440	Thomas P. Tiffany	Low	0.568
354	Ralph Norman	Medium	0.555
322	**Kevin McCarthy**	**High**	**0.546**

Hubs in Net_all_weighted

ID	Name	Seniority	Centrality
367	**Nancy Pelosi**	**High**	**1**
322	**Kevin McCarthy**	**High**	**0.706**
393	Bobby L. Rush	High	0.508
399	Steve Scalise	High	0.473
105	Jim Banks	Medium	0.446
113	Andy Biggs	Medium	0.421
428	Elise Stefanik	Medium	0.421
436	Mark Takano	Medium	0.419
254	Steny H. Hoyer	High	0.418
71	Charles E. Schumer	High	0.409

Hubs in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
367	**Nancy Pelosi**	**High**	**0.955**
399	Steve Scalise	High	0.87
393	Bobby L. Rush	High	0.698
208	**C. Scott Franklin**	**Low**	**0.631**
354	Ralph Norman	Medium	0.63
269	Mike Johnson	Medium	0.627
175	Rodney Davis	Medium	0.626
254	Steny H. Hoyer	High	0.617
436	Mark Takano	Medium	0.581

Hubs in Net_moderate

ID	Name	Seniority	Centrality
226	Bob Good	Low	1
399	Steve Scalise	High	0.88
208	**C. Scott Franklin**	**Low**	**0.659**
118	Lauren Boebert	Low	0.652
440	Thomas P. Tiffany	Low	0.64
164	James Comer	Medium	0.562
159	Michael Cloud	Medium	0.548
322	**Kevin McCarthy**	**High**	**0.547**
337	Mary E. Miller	Low	0.542
388	David Rouzer	Medium	0.523

Hubs in Net_active

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
389	Chip Roy	Low	0.825
269	Mike Johnson	Medium	0.627
113	Andy Biggs	Medium	0.503
246	Jody Hice	Medium	0.447
208	**C. Scott Franklin**	**Low**	**0.426**
188	Byron Donalds	Low	0.409
318	Thomas Massie	Medium	0.343
192	Tom Emmer	Medium	0.34
335	Mariannette Miller-Meeks	Low	0.331

Authorities in Net_all_weighted

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
208	**C. Scott Franklin**	**Low**	**0.942**
190	Jeff Duncan	Medium	0.919
269	Mike Johnson	Medium	0.896
385	John W. Rose	Low	0.802
111	Don Beyer	Medium	0.8
192	Tom Emmer	Medium	0.786
389	Chip Roy	Low	0.773
188	Byron Donalds	Low	0.744
335	Mariannette Miller-Meeks	Low	0.713

Authorities in Net_all

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
208	**C. Scott Franklin**	**Low**	**0.919**
269	Mike Johnson	Medium	0.894
190	Jeff Duncan	Medium	0.873
385	John W. Rose	Low	0.798
192	Tom Emmer	Medium	0.792
188	Byron Donalds	Low	0.671
113	Andy Biggs	Medium	0.661
389	Chip Roy	Low	0.639
335	Mariannette Miller-Meeks	Low	0.637

Authorities in Net_moderate

ID	Name	Seniority	Centrality
322	**Kevin McCarthy**	**High**	**1**
269	Mike Johnson	Medium	0.934
389	Chip Roy	Low	0.704
246	Jody Hice	Medium	0.524
113	Andy Biggs	Medium	0.519
318	Thomas Massie	Medium	0.477
208	**C. Scott Franklin**	**Low**	**0.477**
188	Byron Donalds	Low	0.472
324	Lisa C. McClain	Low	0.439
190	Jeff Duncan	Medium	0.423

Authorities in Net_active

Table 3. Top 10 Members of Each Centrality. *The tables list the top 10 members based on different centrality measures, along with their ID, name, party, house, seniority, and centrality scores. Deep blue represents the Democratic Party in the House of Representatives, light blue represents the Democratic Party in the Senate, orange represents the Republican Party in the House of Representatives, and light orange represents the Republican Party in the Senate. Kevin McCarthy, C. Scott Franklin, and Nancy Pelosi frequently appear among the top 10. Additionally, it's notable that almost all of the top 10 in eigenvector centrality, hubs, and authorities are dominated by House Republicans, and seniority does not affect rankings.*

Centrality & Centralization, Hubs & Authorities

Centrality Measure Correlations:

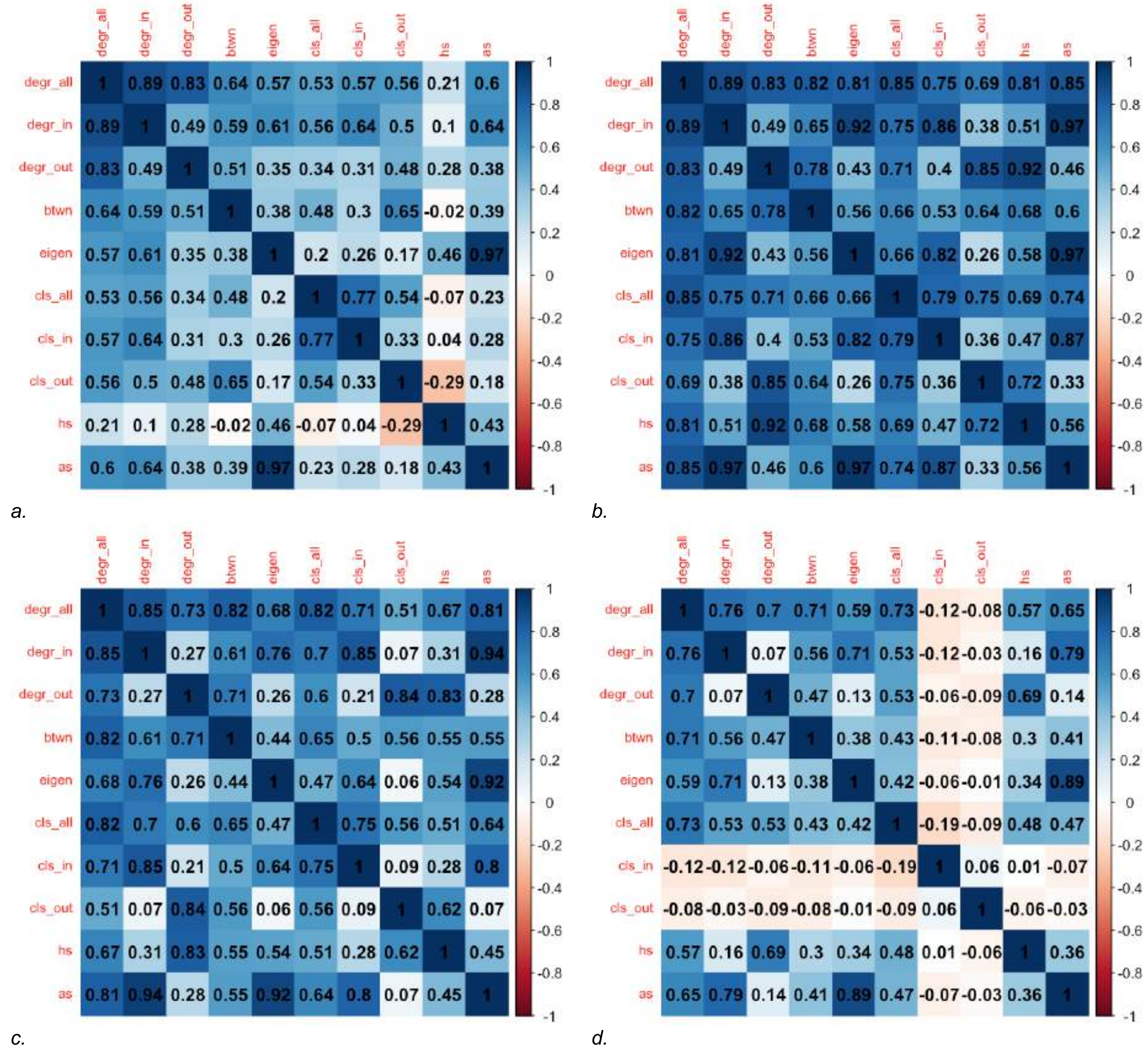


Figure 4. Correlation between Centrality Measures in Networks. *Each matrix represents a network: (a) weighted (Net_all_weighted), (b) unweighted (Net_all), (c) moderate and active connections (Net_moderate), and (d) active connections (Net_active). From left to right and from top to bottom, the centrality measures are All-degree Centrality, In-degree Centrality, Out-degree Centrality, Betweenness Centrality, Eigenvector Centrality, All-closeness Centrality, In-closeness Centrality, Out-closeness Centrality, Hubs and Authorities. The colour intensity indicates the strength of the correlation, with darker blue showing stronger positive correlations and darker red showing stronger negative correlations.*

- In all four networks, all-degree centrality is highly correlated with almost all other centrality measures. This suggests that members with many connections not only have high connectivity but also control information flow and are closer to other members.
- However, as connection strength increases and weaker ties are removed, the correlation between all-degree centrality and other measures decreases. This indicates that relying solely on all-degree centrality might not fully capture the influence of connection weights or the importance of nodes.
- The low correlation between in-degree and out-degree centrality suggests that the patterns of receiving and sending information are different. As the strength of connections increases, this difference becomes more pronounced, with correlation changes from 0.49 to 0.27, and then to 0.07 indicating a growing disparity between incoming and outgoing links of members.

- In-degree centrality is strongly related to authority centrality, and out-degree centrality aligns with hubs centrality. These metrics correspond to incoming and outgoing connections, respectively. Notably, as weaker connections are filtered out, the correlation between these pairs weakens. Additionally, the correlation of hubs and authorities changes from 0.56 to 0.45, and then to 0.36, indicating that the difference between incoming and outgoing links becomes more pronounced as ties are removed. This suggests that while weak connections typically show that important members have many connections with important members, strong connections reveal the opposite: nodes with many strong connections tend to be linked to general members rather than other influential ones.
- Eigenvector centrality is highly correlated with authority centrality, indicating that nodes with high eigenvector centrality often get links from important nodes, which are also recognized as authority nodes. The lower correlation between eigenvector and hubs centrality might be due to key members generally receiving more connections from others than they send out, especially in the context of strong connections.

Next, specific observations will be made to verify the relationships reflected by the correlation matrix. Specifically, observing the connection patterns of network members under different attributes (such as party, seniority, and house) at different strengths of connections can help verify and understand the relationships between these centrality measures.

Key Members:

Kevin McCarthy, C. Scott Franklin, and Nancy Pelosi are identified as the most influential legislators in the Congressional Twitter network, as highlighted in Table 3. They all exhibit high all-degree and betweenness, and closeness centralities across all networks, meaning that they have numerous connections, act as intermediaries on many paths, and can quickly connect with others. Kevin McCarthy stands out as the most significant figure, consistently ranking high across almost all centralities. This aligns with his role as the 55th Speaker of the U.S. House of Representatives from January to October 2023.

C. Scott Franklin, also a Republican, is notable for his high in-degree centrality, indicating he receives many connections and serves as a major authority figure, attracting links from highly central members. His high hub score in stronger networks also suggests he sends many strong outgoing connections.

In addition, Table 4 further shows that Kevin McCarthy and C. Scott Franklin predominantly connect with Republican Representatives, primarily those with low to medium seniority. McCarthy receives more strong incoming links, while Franklin primarily receives weaker ones.

In comparison, Nancy Pelosi, a Democrat and the 52nd Speaker of the House, is marked by her high out-degree centrality, making her a major hub, especially in weak and moderate connections. This suggests she plays a key role in initiating interactions. Moreover, Table 4 also shows that Pelosi has a more diverse network, frequently connecting across different parties and houses, particularly in sending outgoing links.

Kevin McCarthy - Number of Neighbors: 284

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	4	123	0	21	50	56	126	1
out	18	139	0	29	67	61	152	5
all	22	262	0	50	117	117	278	6

C. Scott Franklin - Number of Neighbors: 182

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	12	109	0	13	59	49	121	0
out	1	60	0	8	28	25	61	0
all	13	169	0	21	87	74	182	0

Nancy Pelosi - Number of Neighbors: 261

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	50	1	0	26	16	9	43	8
out	137	73	0	62	84	64	191	19
all	187	74	0	88	100	73	234	27

Net_all_weighted

Kevin McCarthy - Number of Neighbors: 284

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	4	123	0	21	50	56	126	1
out	18	139	0	29	67	61	152	5
all	22	262	0	50	117	117	278	6

C. Scott Franklin - Number of Neighbors: 182

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	12	109	0	13	59	49	121	0
out	1	60	0	8	28	25	61	0
all	13	169	0	21	87	74	182	0

Nancy Pelosi - Number of Neighbors: 261

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	50	1	0	26	16	9	43	8
out	137	73	0	62	84	64	191	19
all	187	74	0	88	100	73	234	27

Net_all

Kevin McCarthy - Number of Neighbors: 233

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	4	116	0	20	48	52	120	0
out	6	107	0	21	49	43	112	1
all	10	223	0	41	97	95	232	1

C. Scott Franklin - Number of Neighbors: 176

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	11	104	0	12	56	47	115	0
out	1	60	0	8	28	25	61	0
all	12	164	0	20	84	72	176	0

Nancy Pelosi - Number of Neighbors: 162

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	42	1	0	23	13	7	37	6
out	84	35	0	34	48	37	115	4
all	126	36	0	57	61	44	152	10

Net_moderate

Kevin McCarthy - Number of Neighbors: 91

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	0	63	0	9	21	33	63	0
out	2	26	0	4	12	12	28	0
all	2	89	0	13	33	45	91	0

C. Scott Franklin - Number of Neighbors: 45

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	0	22	0	3	9	10	22	0
out	0	23	0	3	10	10	23	0
all	0	45	0	6	19	20	45	0

Nancy Pelosi - Number of Neighbors: 30

DIRECTION	D	R	I	HIGH	MEDIUM	LOW	REPRESENTATIVES	SENATE
in	5	0	0	2	3	0	5	0
out	19	6	0	7	10	8	25	0
all	24	6	0	9	13	8	30	0

Net_active

Table 4. Neighbours of Key Members. *This table compares the neighbours of Kevin McCarthy, C. Scott Franklin, and Nancy Pelosi across different network configurations. It details the direction of connections, party affiliation, seniority, and house representation, showing how these proportion change as weaker connections are filtered out. Net_all_weighted and Net_all have the same results as weight is not accounted in these statistics. Kevin McCarthy and C. Scott Franklin predominantly connect with members of the same party and house (Republican in the Representatives), primarily those with low to medium seniority. In contrast, Nancy Pelosi has a more diverse network, connecting across different parties and houses, particularly in sending information (outgoing links), and frequently interacting with members of high to medium seniority. However, as weaker connections are removed, all three members are more concentrated within their own affiliations.*

Centrality Differences among Different Parties, Seniorities, and Houses:

Net_all_weighted

PARTY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
D	0.117	0.058	0.059	0.004	**0.016**	0.619	0.339	0.398	**0.027**	**0.018**
R	0.120	0.060	0.060	0.004	**0.112**	0.603	0.330	0.352	**0.188**	**0.096**

Net_all

PARTY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
D	0.117	0.058	0.059	0.003	**0.158**	0.488	0.429	0.429	**0.136**	**0.208**
R	0.120	0.060	0.060	0.003	**0.232**	0.487	0.430	0.424	**0.174**	**0.238**

Net_moderate

PARTY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
D	0.097	0.048	0.049	0.003	**0.056**	0.470	0.400	0.400	**0.161**	**0.134**
R	0.104	0.052	0.052	0.003	**0.204**	0.468	0.404	0.399	**0.282**	**0.202**

Net_active

PARTY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
D	0.016	0.008	0.008	0.006	**0.005**	0.291	0.240	0.253	**0.013**	**0.015**
R	0.023	0.011	0.011	0.008	**0.088**	0.307	0.251	0.225	**0.142**	**0.097**

***Table 5. Normalised Mean Centrality Measures by Party in Each Network**. The table compares average centrality scores, including degree, betweenness, eigenvector, closeness, hubs, and authorities, between Democratic (D) and Republican (R) members within four network contexts. These measures show how centrality shifts as weaker connections are filtered out, revealing differences in how connected members of each party are within the Congressional Twitter network. From the highlighted values, Republicans generally show higher eigenvector, hub and authority scores compared to Democrats, indicating they are more active in engaging with the tweets of other influential members, especially through stronger connections. There are only two members of the Independence Party, so they were not included in the statistics.*

After analysing individual centralities, the focus shifts to the overall patterns based on different attributes. Table 3 displays the top 10 members in eigenvector centrality, hubs and authorities are predominantly Republican Representatives, especially in networks with stronger connections. Table 5 reveals that while the average centrality scores of both parties are generally similar, Republicans score higher in eigenvector centrality, hub, and authority measures as connections strengthen. These suggest Republicans tend to engage more with influential members, especially strong interactions.

Net_all_weighted

SENIORITY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
High	0.129	0.061	0.068	0.005	0.044	0.628	0.337	0.393	0.077	0.044
Medium	0.112	0.057	0.055	0.004	0.057	0.606	0.332	0.366	0.099	0.050
Low	0.118	0.061	0.057	0.004	0.078	0.609	0.335	0.378	0.124	0.067

Net_all

SENIORITY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
High	0.129	0.061	0.068	0.004	0.169	0.492	0.429	0.437	0.160	0.217
Medium	0.112	0.057	0.055	0.002	0.186	0.484	0.426	0.423	0.145	0.214
Low	0.118	0.061	0.057	0.003	0.216	0.487	0.434	0.424	0.161	0.237

Net_moderate

SENIORITY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
High	0.107	0.051	0.056	0.004	0.089	0.473	0.401	0.406	0.206	0.151
Medium	0.095	0.048	0.047	0.003	0.123	0.466	0.398	0.395	0.209	0.159
Low	0.102	0.052	0.049	0.003	0.154	0.471	0.408	0.400	0.238	0.183

Net_active

SENIORITY	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
High	0.021	0.010	0.012	0.009	0.039	0.299	0.254	0.254	0.061	0.040
Medium	0.018	0.009	0.009	0.004	0.040	0.297	0.251	0.223	0.065	0.051
Low	0.020	0.011	0.009	0.008	0.056	0.302	0.228	0.250	0.100	0.069

***Table 6. Normalised Mean Centrality Measures by Seniority in Each Network**. The table compares average centrality scores, including degree, betweenness, eigenvector, closeness, hubs, and authorities, among members with high, medium, and low seniority across four different networks. These measures illustrate how centrality changes as weaker connections are removed, showing no significant differences in how connected members of varying seniority are within the Congressional Twitter network.*

Regarding seniority, Table 3 does not provide many clear patterns, and the values of different seniority in Table 6 are also quite average. The only notable exceptions are in the Net_active, where members with medium seniority have a relatively low average betweenness centrality (0.004), while those with low seniority have a relatively high average hub score (0.1). Overall, seniority does not significantly impact connectivity within the network.

Net_all_weighted

HOUSE	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
Representatives	0.120	0.062	0.058	0.004	**0.072**	0.607	0.338	0.359	**0.118**	**0.063**
Senate	0.109	0.047	0.062	0.005	**0.012**	0.633	0.318	0.449	**0.035**	**0.016**

Net_all

HOUSE	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
Representatives	0.120	0.062	0.058	0.003	**0.219**	0.489	0.436	0.424	**0.164**	**0.245**
Senate	0.109	0.047	0.062	0.003	**0.079**	0.480	0.401	0.440	**0.110**	**0.123**

Net_moderate

HOUSE	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
Representatives	0.104	0.053	0.051	0.003	**0.146**	0.472	0.409	0.399	**0.241**	**0.187**
Senate	0.085	0.038	0.048	0.003	**0.034**	0.457	0.375	0.402	**0.119**	**0.072**

Net_active

HOUSE	AVG_DEGREE	AVG_INDEGREE	AVG_OUTDEGREE	AVG_BETWEENNESS	AVG_EIGENVECTOR	AVG_CLOSENESS	AVG_IN_CLOSENESS	AVG_OUT_CLOSENESS	AVG_HUB_SCORE	AVG_AUTHORITY_SCORE
Representatives	0.020	0.010	0.010	0.007	**0.054**	0.303	0.241	0.249	**0.089**	**0.064**
Senate	0.016	0.007	0.009	0.006	**0.004**	0.282	0.259	0.197	**0.016**	**0.010**

***Table 7. Normalised Mean Centrality Measures by House in Each Network**. The table compares average centrality scores, including degree, betweenness, eigenvector, closeness, hubs, and authorities, between Representatives and Senate members across four network contexts. These measures illustrate how centrality changes as weaker connections are removed, uncovering differences in connectivity between the two houses in the Congressional Twitter network. The highlighted values show that Representatives generally have higher eigenvector, hub, and authority scores compared to Senate members, demonstrating they are more active in engaging with the tweets of other important members, especially in networks with stronger connections.*

Similar patterns as party, Table 3 shows that nearly all the top 10 members are from the House of Representatives. Table 7 confirms that Representatives show higher eigenvector, hub, and authority centrality scores, indicating greater activity. They outperform Senate members, especially as connections strengthen, denoting that Representatives are more active on Twitter, interact more closely with key members, and play central and authoritative roles.

Patterns Observed from Degree Centrality Visualization:

Figures 5 to 9 reveal three distinct clusters: House Republicans, House Democrats, and a loosely connected group of Senators from both parties. The visualizations of all-degree, in-degree, and out-degree centralities in the original weighted network show that high centrality members are mainly concentrated among Representatives, with certain Democrats having higher out-degrees. When looking at different connection strengths, Republican Representatives dominate in stronger interactions, as seen in Figures 6d and 7d, while Democratic Representatives have higher out-degrees in weaker connections, as seen in Figure 7b. This confirms that within the House of Representatives, Republicans engage in more intense and interactive connections, whereas Democrats tend to broadcast information more widely.

Figures 8 and 9 further investigate the patterns. Republican Representatives are more likely to engage with influential members, with frequent interactions. This is supported by Figures 8d and 9d, where Republican Representatives show significantly higher hub and authority scores. Oppositely, Democratic Representatives engage in broader, more dispersed interactions. Figure 8b shows that some Democratic Representatives connect with key members, though primarily through weak connections. These observed phenomena are consistent with the previous analysis from the table, supplementing the visualization and providing a more intuitive representation.

a. b. c. d.

Figure 5. All-degree Centrality Visualization in Different Networks. *This figure shows the distribution of all-degree centrality in four different networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Node size reflects centrality, with larger nodes indicating more connections. Pink nodes represent Republicans, deep green nodes represent Democrats, and light green nodes represent Independents. White-framed nodes are Representatives, while black-framed nodes are Senators. The edge colour indicates the party origin of the connection. High all-degree centrality members are mostly concentrated among Representatives of both parties, while there is a significant concentration of Republican Representatives in the strong connection network.*

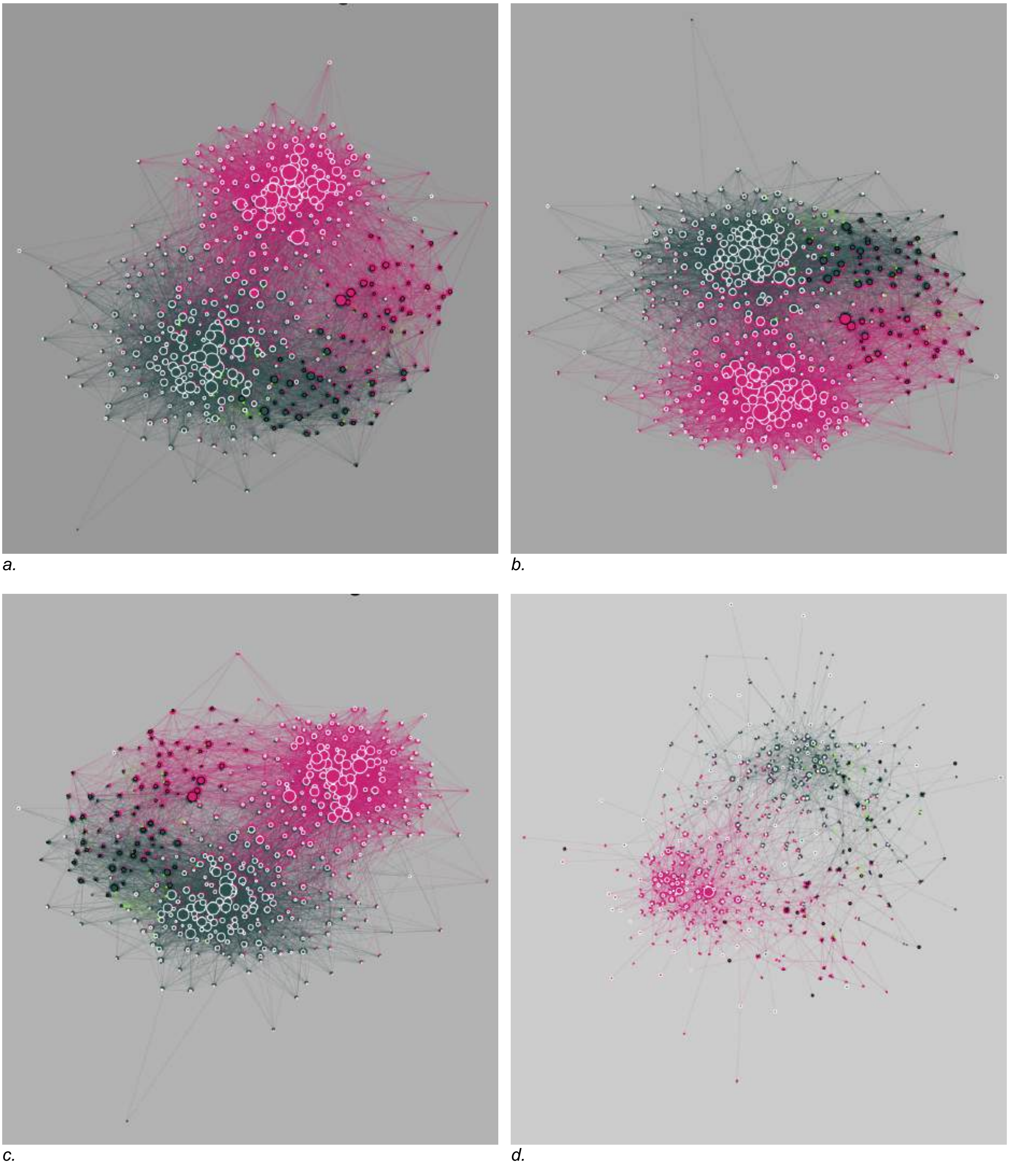


Figure 6. In-degree Centrality Visualization in Different Networks. *This figure shows the distribution of in-degree centrality in four different networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Node size reflects centrality, with larger nodes indicating more connections. Pink nodes represent Republicans, deep green nodes represent Democrats, and light green nodes represent Independents. White-framed nodes are Representatives, while black-framed nodes are Senators. The edge colour indicates the party origin of the connection. High in-degree centrality members are mostly concentrated among Representatives of both parties, while there is a significant concentration of Republican Representatives in the strong connection network.*

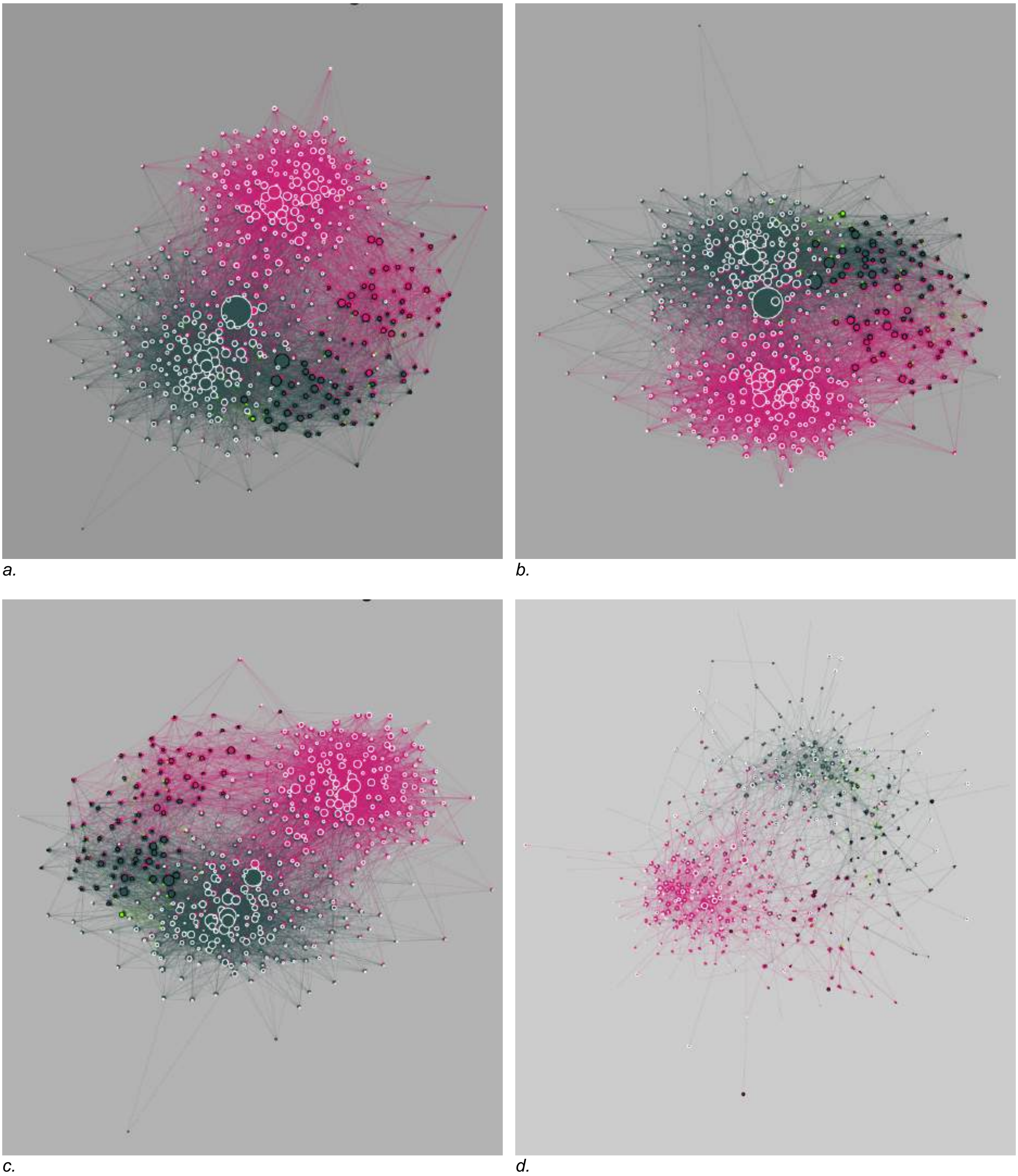
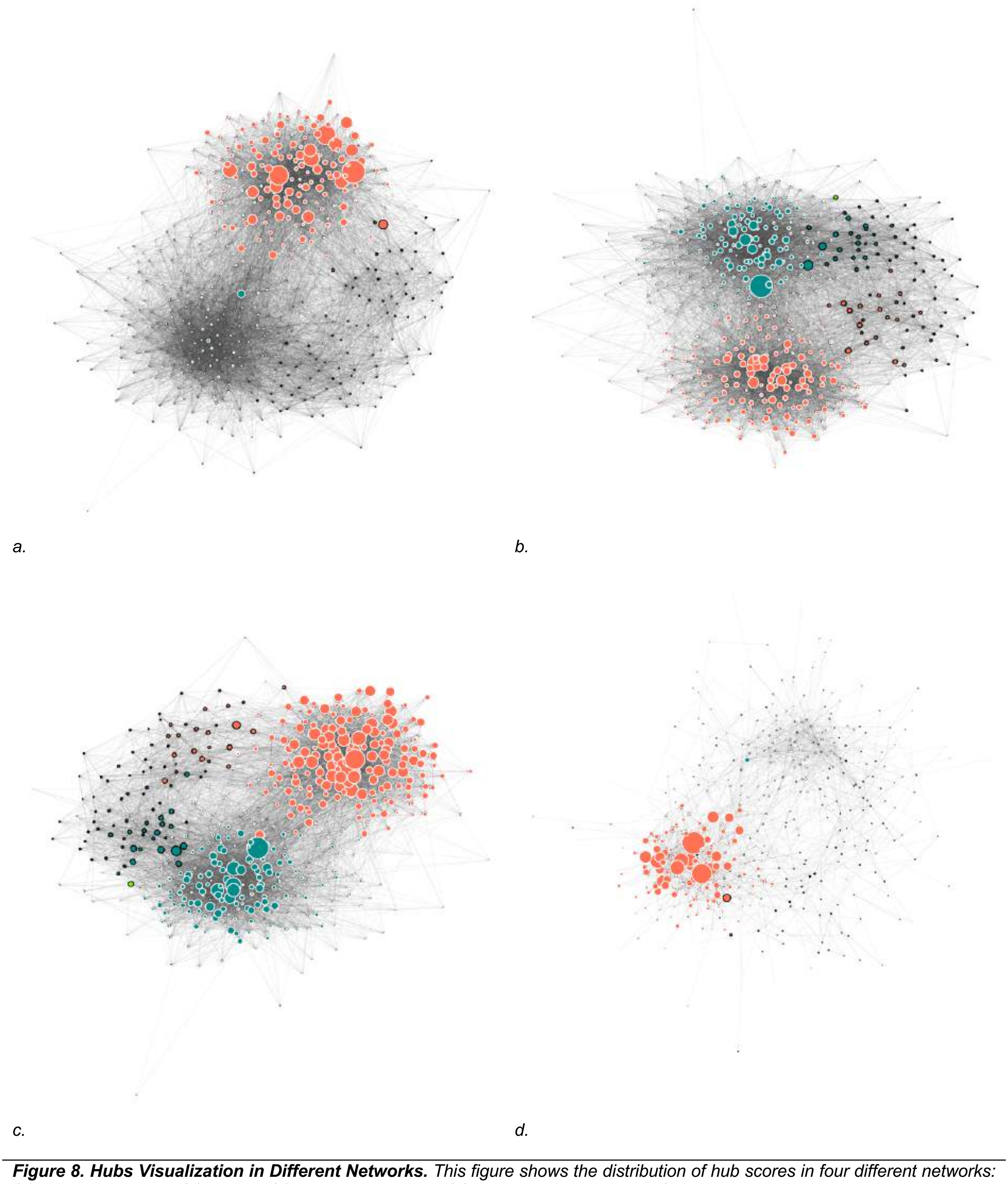
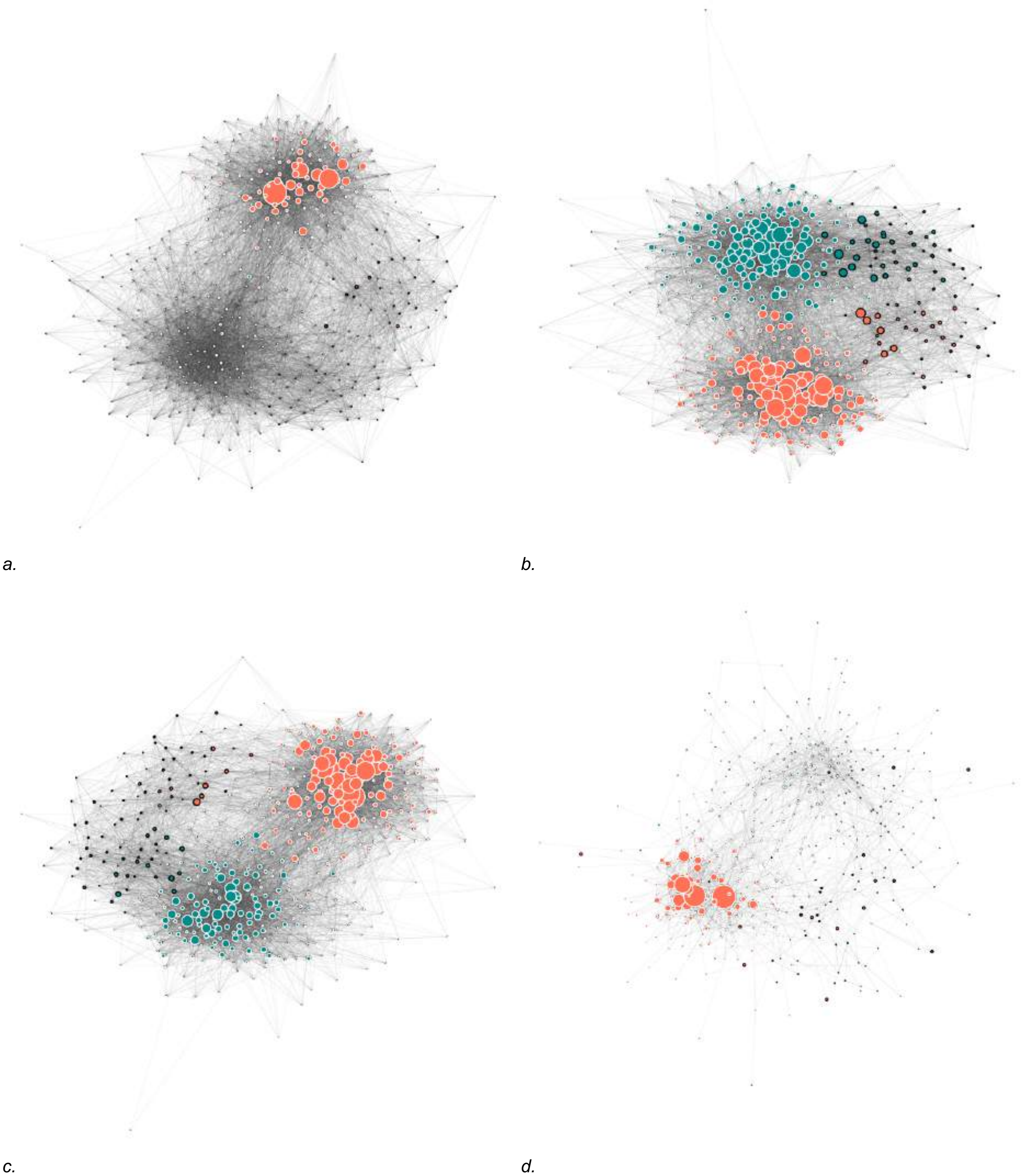


Figure 7. Out-degree Centrality Visualization in Different Networks. *This figure shows the distribution of out-degree centrality in four different networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Node size reflects centrality, with larger nodes indicating more connections. Pink nodes represent Republicans, deep green nodes represent Democrats, and light green nodes represent Independents. White-framed nodes are Representatives, while black-framed nodes are Senators. The edge colour indicates the party origin of the connection. High out-degree centrality members are mostly concentrated among Representatives of both parties. There is a significant concentration of Republican Representatives in the strong connection network, while some Democrat Representatives have significant high out-degree centrality in weak connections.*



a. *b.* *c.* *d.*

Figure 8. Hubs Visualization in Different Networks. *This figure shows the distribution of hub scores in four different networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Node size reflects hub score, with larger nodes linking to more authoritative nodes. Orange nodes represent Republicans, deep green nodes represent Democrats, and light green nodes represent Independents. Nodes with white frames are Representatives, while those with black frames are Senators. In the original weighted network, high hub scores are mainly concentrated among Republican Representatives. For weaker connections, Democratic Representatives show higher hub scores. In moderate connections, high hub scores are distributed between both parties in the House of Representatives. In the strong connection network, high hub scores are predominantly found among Republican Representatives.*



a. b. c. d.

Figure 9. Authorities Visualization in Different Networks. *This figure displays the distribution of authority scores in four different networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Node size represents authority score, where larger nodes are frequently referenced by hubs. Orange nodes represent Republicans, deep green nodes represent Democrats, and light green nodes represent Independents. Nodes with white frames indicate Representatives, while those with black frames are Senators. In the original weighted network, high hub scores are mainly concentrated among Republican Representatives. As connection strength increases, Republican Representatives continue to dominate in authority scores.*

Centralisation

Network	***Net_all_weighted***	***Net_all***	***Net_moderate***	***Net_active***
All-degree centralization	*0.241*	*0.241*	*0.196*	*0.091*
In-degree centralization	*0.209*	*0.209*	*0.203*	*0.129*
Out-degree centralization	*0.384*	*0.384*	*0.201*	*0.074*
Betweenness centralization	*0.070*	*0.070*	*0.055*	*0.182*
Eigenvector centralization	*0.810*	*0.810*	*0.877*	*0.957*
All-closeness centralization	*0.318*	*0.318*	*0.242*	*0.234*
In-closeness centralization	*NaN*	*NaN*	*NaN*	*NaN*
Out-closeness centralization	*0.214*	*0.214*	*NaN*	*NaN*

Table 8. Comparison of Centralization Across Four Networks. *This table presents the centralization scores for various networks, covering all-degree, in-degree, out-degree, betweenness, eigenvector, and closeness centralizations across four different networks: Net_all_weighted, Net_all, Net_moderate, and Net_active. The scores measure how centralized or equal the networks are, with higher values showing the network is more concentration or hierarchical, on the contrary, it is more equal. The identical scores for Net_all_weighted and Net_all are due to the centralization algorithm in "igraph" not supporting weighted networks. The NaN values for in-closeness and out-closeness centralizations are due to the presence of directionally isolated nodes in the networks.*

Due to the inability of the algorithm to account for network weights, comparing the centralization of networks with different connection strengths helps address this limitation and reveals the impact of weights on centralization.

- Table 8 lists that across all three centralization measures (all-degree, in-degree, and out-degree), a consistent trend of decreasing suggests that the network becomes less centralized as the weaker connections are filtered out. Notably, the slight drop in in-degree centralization from 0.209 to 0.203 suggests weak incoming links are evenly distributed. In contrast, the sharp decline in out-degree centralization from 0.384 to 0.074 highlights that as weaker connections are eliminated, outgoing connections become more evenly distributed, implying that these weaker connections were previously more centralized.
- Betweenness centralization decreases from 0.07 to 0.055, suggesting as weaker connections are filtered out, the importance of bridge nodes is somewhat reduced, possibly because the network becomes more decentralized with multiple paths between nodes, implying the weak connections might frequently connect to key members. However, the sharp rise to 0.182 in the Net_active network shows that as the network becomes more focused on strong connections, it increasingly relies on a few central nodes to maintain its connectivity.
- Eigenvector centralization increases steadily from 0.810 to 0.957, indicating that as the network filters out weaker connections, central nodes become more interconnected with other central nodes, enhancing their influence.
- In-closeness and out-closeness centralization show NA values due to some nodes being directionally isolated (e.g., a node might send connections but not receive any). Therefore, only all-closeness centralization is discussed here. The decrease in all-closeness centralization as the network shifts from all to moderate and then strong connections suggests that the network becomes less tightly organized around a single central node, leading to a more distributed and less hierarchical structure.

To wrap up, central members generally have weak interactions with others, particularly in terms of outgoing connections and serving as information bridges. Overall, weak incoming links are evenly distributed, and weak outgoing connections are more hierarchical. As stronger connections dominate, the network becomes more evenly distributed and less centralized, but with a focus on influential members.

Assortativity & Homophily

Network	***Net_all_weighted***	***Net_all***	***Net_moderate***	***Net_active***
Homophily of Party	*0.711*	*0.711*	*0.727*	*0.744*
Homophily of State	*0.103*	*0.103*	*0.113*	*0.201*
Homophily of Seniority	*0.057*	*0.057*	*0.060*	*0.053*
Homophily of House	*0.588*	*0.588*	*0.609*	*0.627*
Degree assortativity	*-0.096*	*-0.096*	*-0.087*	*-0.093*

Table 9. Homophily of Node Attributes and Degree Assortativity across Four Networks. *This table compares the assortativity and homophily scores for various attributes—Party, State, Seniority, and House—across four different networks: Net_all_weighted, Net_all, Net_moderate, and Net_active. Homophily measures how likely nodes with similar attributes are to connect, while assortativity reflects the tendency of nodes to connect based on the similarity of their degrees. The scores indicate that Party and House are the most significant factors driving homophily, with values increasing as weaker connections are filtered out. Conversely, the degree assortativity values remain negative and close to 0, suggesting no tendency for nodes with similar degrees to connect.*

Since homophily and degree assortativity measures do not account for weights, the values for Net_all_weighted and Net_all are identical. The connection tendency of homophily is revealed by comparing the three separated networks.

- **Party homophily** is high across all networks, with values over 0.7, meaning members of the same party are more likely to connect. As connections strengthen, party homophily rises slightly from 0.711 to 0.744, suggesting that party similarity becomes pronounced in stronger networks. Figure 10 reflects this, with two distinct groups forming within each party.
- **State homophily** is relatively low, suggesting that members from the same state are not strongly inclined to connect. However, it does increase slightly to around 0.2 as connections intensify, indicating a modest influence. Figure 11 shows a mixed pattern, reflecting this low homophily.
- **Seniority homophily** remains low (below 0.06) in all networks, representing that seniority has minimal impact on connection tendency. Members do not tend to connect based on similar levels of seniority. Figure 12 also shows a mixed, scattered pattern similar to that of state homophily.
- **House homophily** is relatively high, around 0.6 in all networks, demonstrating that members from the same chamber are more likely to connect. As weaker connections are removed, this homophily becomes more prominent, with clear divisions between the two houses, as seen in Figure 13.
- **Degree assortativity** ranges from -1 to 1, with values near 1 indicating that high-degree nodes tend to connect with other high-degree nodes, and vice versa. In all networks, the values are slightly negative, indicating a subtle tendency for high-degree nodes to connect with low-degree nodes.

This analysis highlights that party and house are strong predictors of connections, while state and seniority have minimal influence. Degree assortativity indicates that influential members are not likely to connect with other influential members.

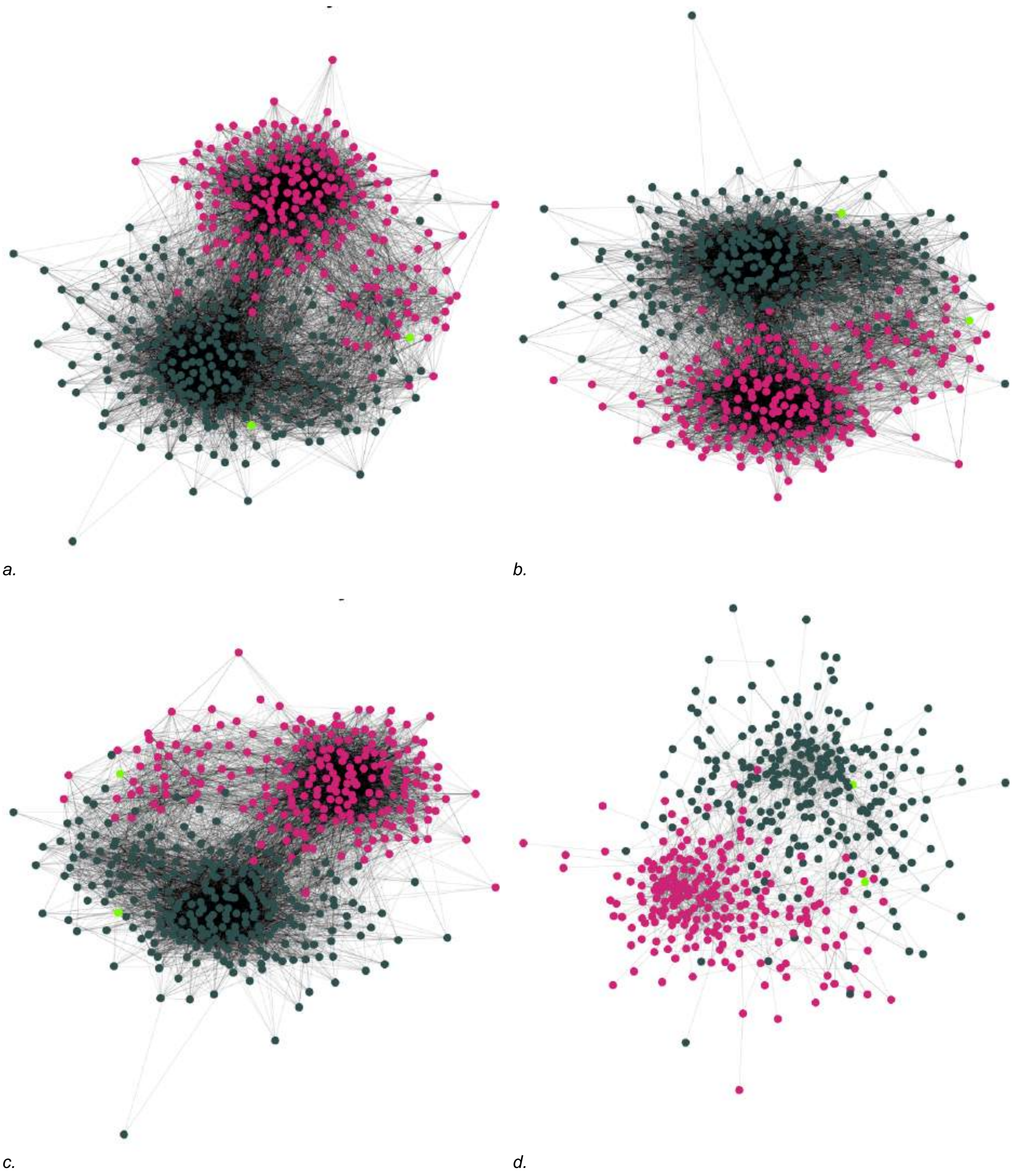
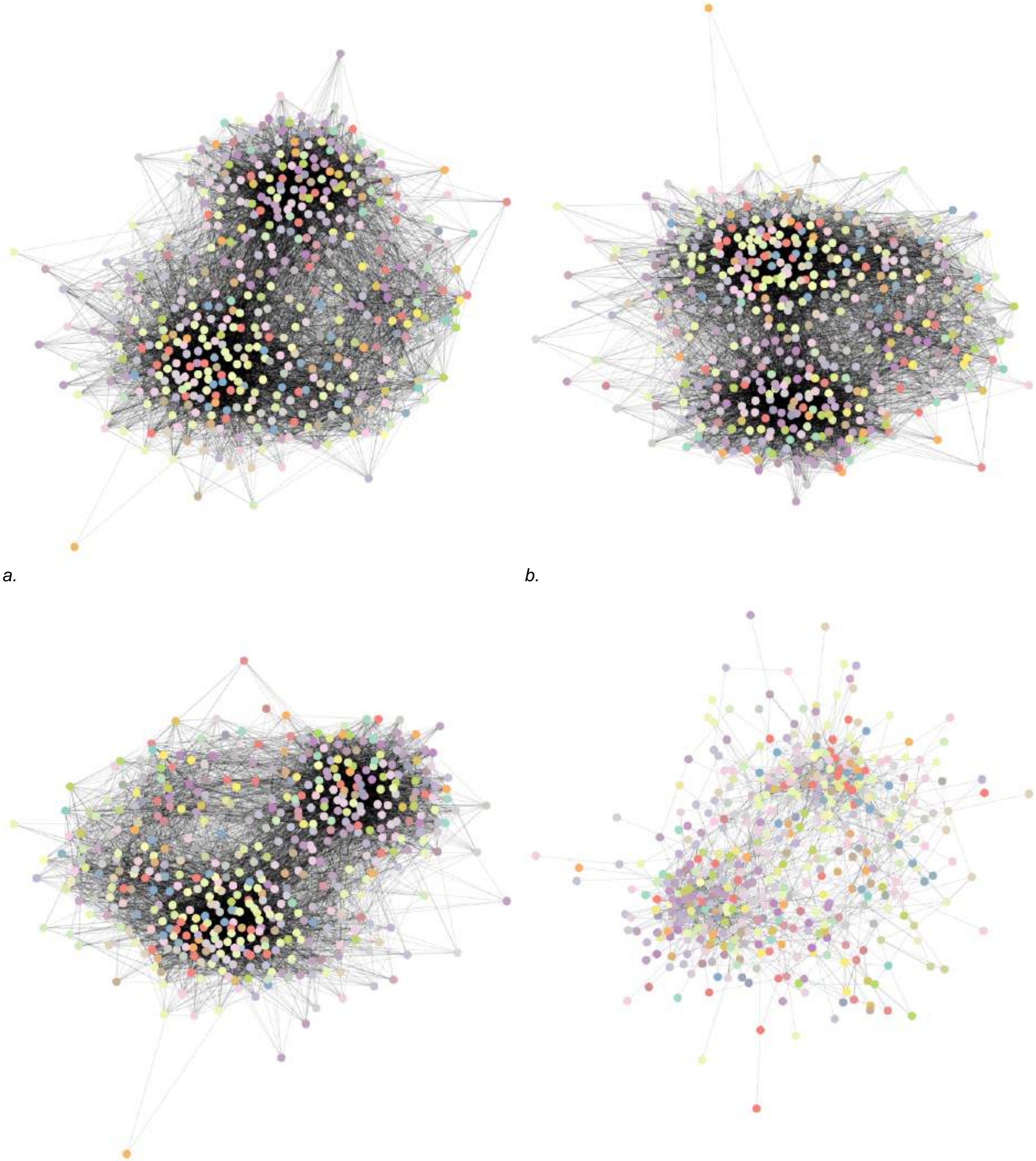


***Figure 10. Party Homophily in Different Networks**. This figure shows whether Congress members tend to connect with others from the same party in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Republicans are shown in deep pink, Democrats in deep green, and Independents in light green. In all networks, two distinct groups are visible, indicating that members are more likely to interact with others from their own party.*



a.   b.   c.   d.

***Figure 11. State Homophily in Different Networks**. This figure shows whether Congress members tend to connect with others from the same state in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active Each colour represents a different state. The scattered and mixed appearance of nodes across all networks suggests that members do not particularly prefer to interact with others from their own state.*

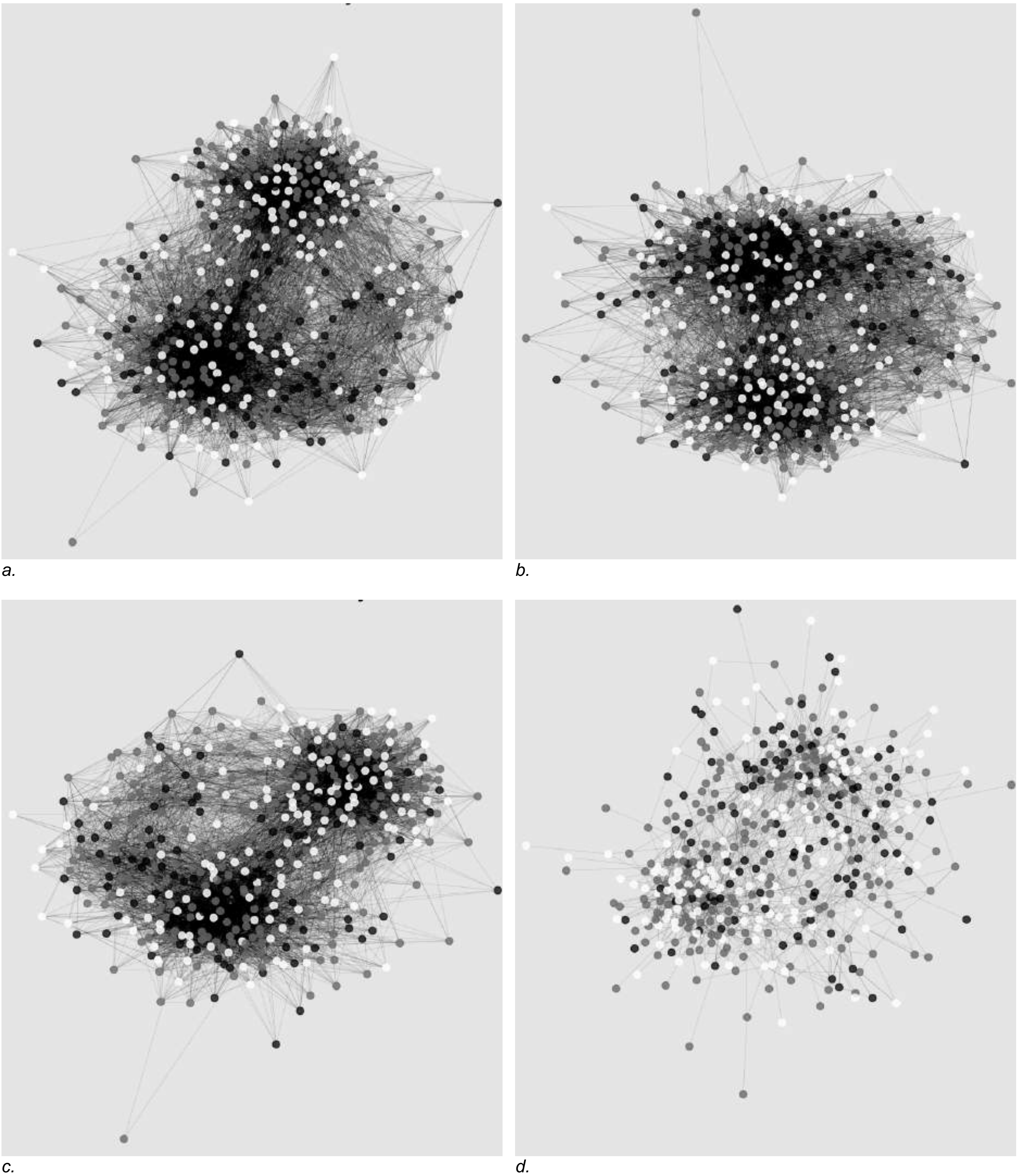
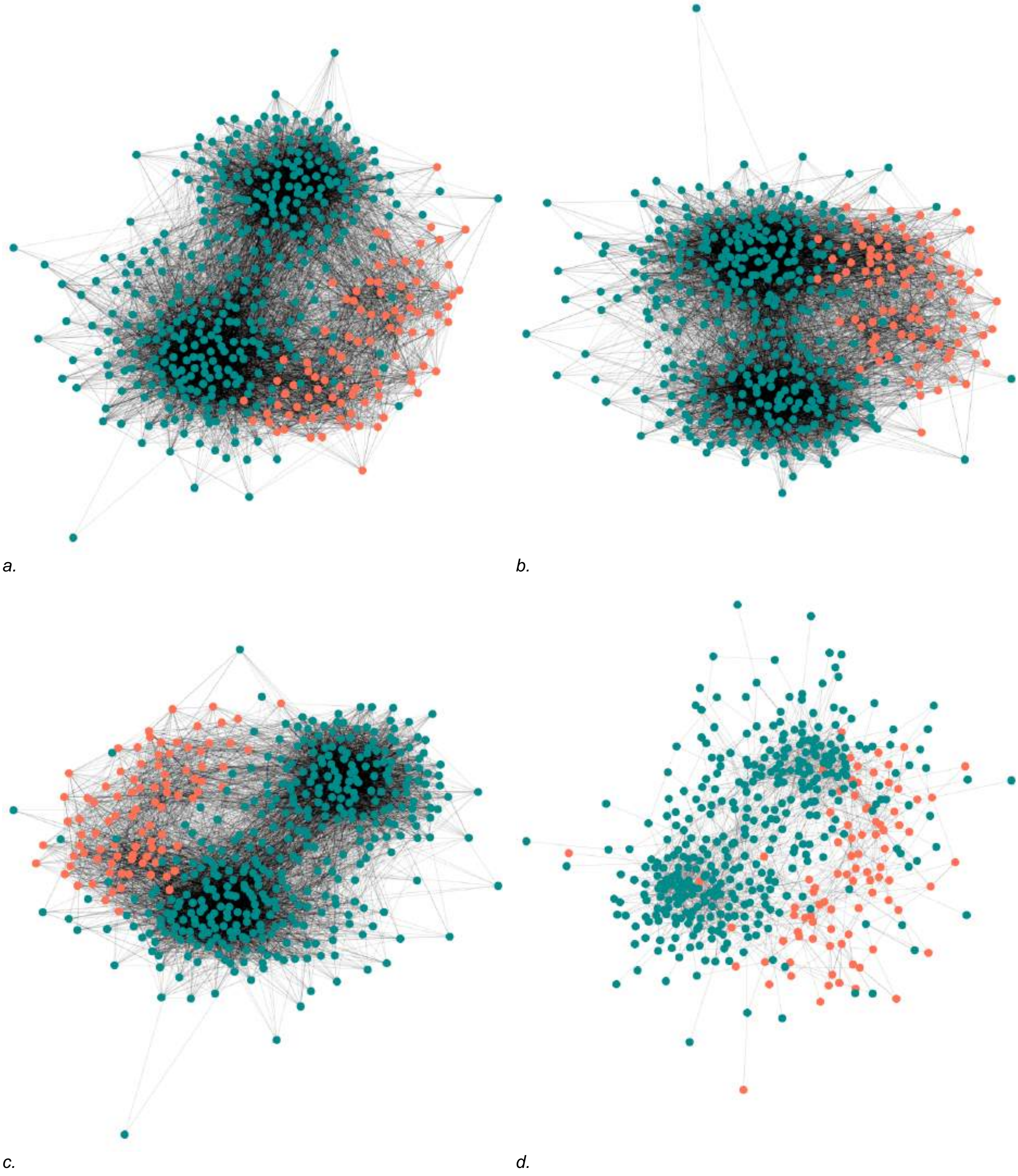


Figure 12. Seniority Homophily in Different Networks*. This figure examines whether Congress members tend to connect with others of similar seniority in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. White nodes represent members with low seniority, grey nodes represent those with medium seniority, and black nodes represent those with high seniority. The dispersed and mixed pattern of nodes in all networks indicates that members do not show a strong preference for interacting with others of similar seniority.*



a. b. c. d.

Figure 13. House Homophily in Different Networks. *This figure examines whether Congress members tend to connect with others from the same house in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Deep green nodes represent members of the House of Representatives, while orange nodes represent Senators. In all networks, two distinct groups emerge, indicating that members are more likely to interact with others from their own house.*

Reciprocity

Network	Net_all_weighted	Net_all	Net_moderate	Net_active
Reciprocity	0.462	0.462	0.393	0.212

Table 10. Reciprocity in Different Networks. *This table presents reciprocity values across four networks. Reciprocity measures the likelihood of mutual connections between nodes. The values decrease from 0.462 to 0.212, suggesting that fewer mutual patterns persist in stronger connections. Net_all_weighted and Net_all has the same value since this metric only detects whether nodes are connected, not the strength of those connections.*

Reciprocity is relatively high, with values of 0.462, 0.393, and 0.212, indicating that bidirectional connections are common in these networks. However, as weaker connections are filtered out, fewer mutual connections remain. This suggests that while weak interactions are widespread, stronger connections are often one-way.

Transitivity

Network	Net_all_weighted	Net_all	Net_moderate	Net_active
Transitivity	0.270	0.270	0.251	0.135

Table 11. Transitivity in Different Networks. *Transitivity measures the likelihood that two nodes connected to a common third node are also connected to each other, forming a triangle. The transitivity value decreases from 0.270 to 0.135 as weaker connections are filtered out, indicating that triangles become less frequent in stronger connections, reflecting a change towards more linear and less clustered structures. This metric also cannot identify the weights of connections,*

Again, all transitivity-relevant measures here are not suitable for edge weights. As connections strengthen, triangular relationships become weaker, as seen in the decrease from 0.27 to 0.135 in Table 11, implying fewer group structures in active relationships. Figure 14 displays the distribution of local transitivity, which refers to the probability that neighbors of a node are interconnected, i.e., whether the neighbors form a tightly knit group. The distributions in Net_all and Net_moderate are roughly similar, with a bell-shaped pattern concentrated between 0.2 and 0.4, meaning that most members have relatively well-connected neighbours, which supports frequent interactions. In contrast, the distribution in Net_active shows a concentration in the low transitivity range (0 to 0.25), with a few nodes reaching a transitivity of 1. This suggests that while most neighbours are not tightly connected in the active network, some small, tightly knit groups exist, consistent with the observed loose connections and closely clustered Republican Representatives in the strong connections.

Moreover, since the network is directed, triangular structures can take multiple forms. Triad types detail the specific 16 types of connections, where triads can be classified based on the presence and direction of edges between the three nodes, such as completely unconnected triads, partially connected triads, and fully closed triads. The bar charts of Figure 15 illustrate the proportions of different triad types in the networks. In all connections, "003" (empty triad) is the most common at 75.7%, followed by "012" (A→B, C) at 15.1%, and "102" (A↔B, C) at 6.4%, indicating that unconnected structures, simple one-way and two-way connections are dominated. Net_moderate is similar to Net_all, with "003" still being the most prevalent at 78.2%. However, in Net_active, the proportion of "012" and "102" triads significantly drops, displaying fewer one-way and two-way connections in active interactions. This aligns with the observed sparser connections and reduced reciprocity of a strong connection network.

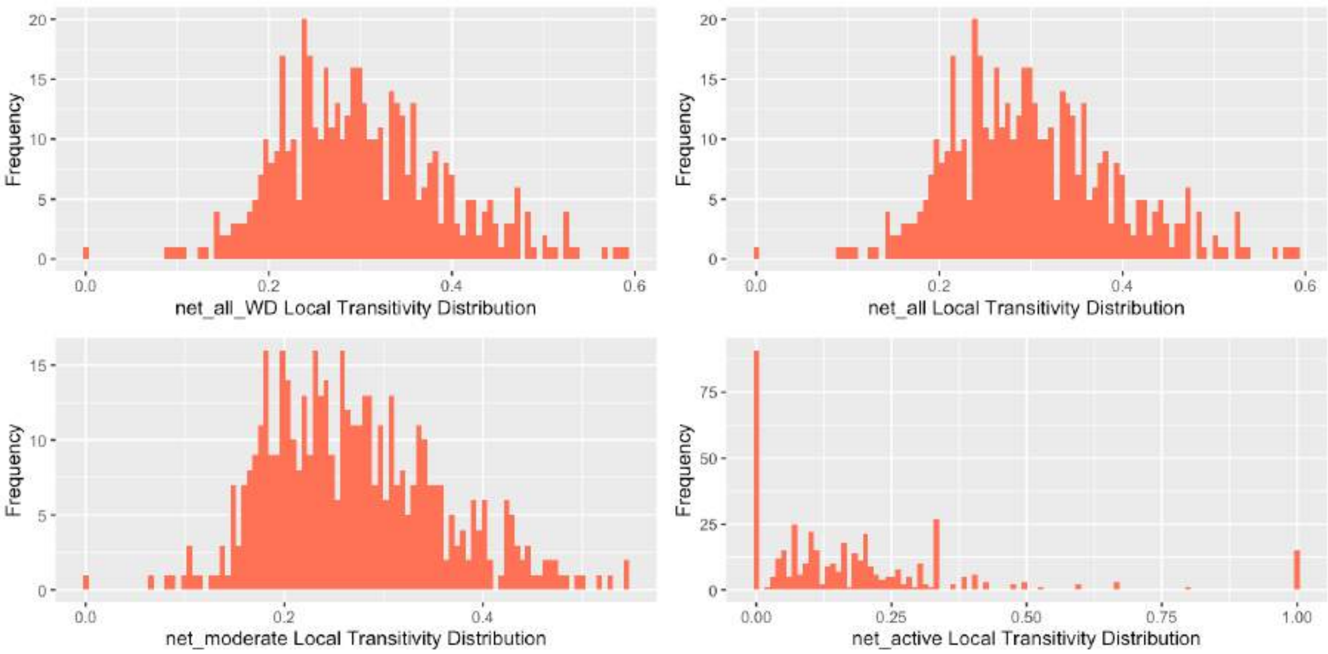


Figure 14. Transitivity Distribution in Different Networks. *Local transitivity reflects how likely neighbours of a node are also connected to each other. The distributions indicate that as connections become stronger, overall clustering decreases. More nodes have very low transitivity, while a few exhibit very high transitivity, suggesting more polarized structures with only a few tightly connected groups. Since this measure does not consider edge weights, the distributions in Net_all_weight and Net_all are identical.*

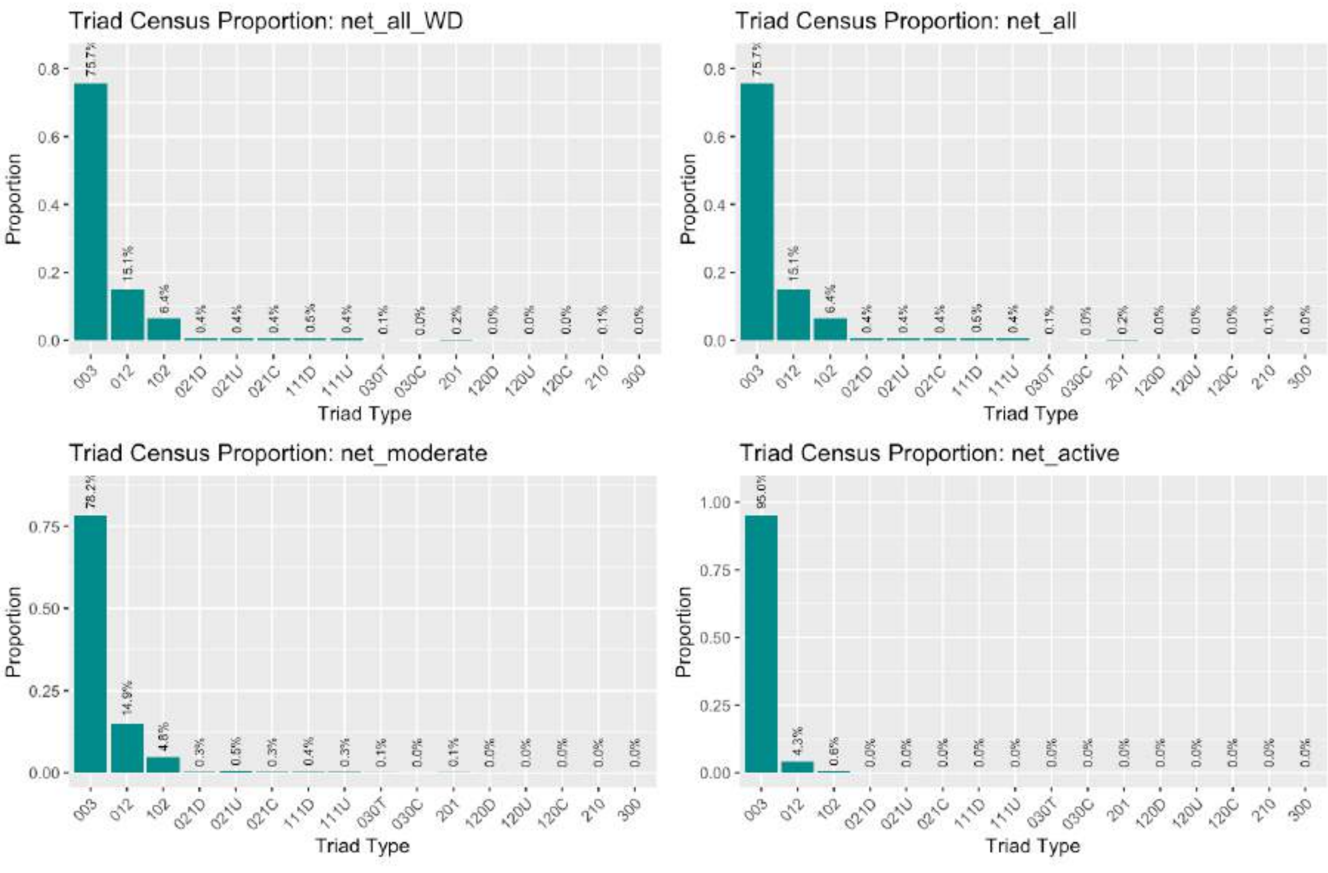


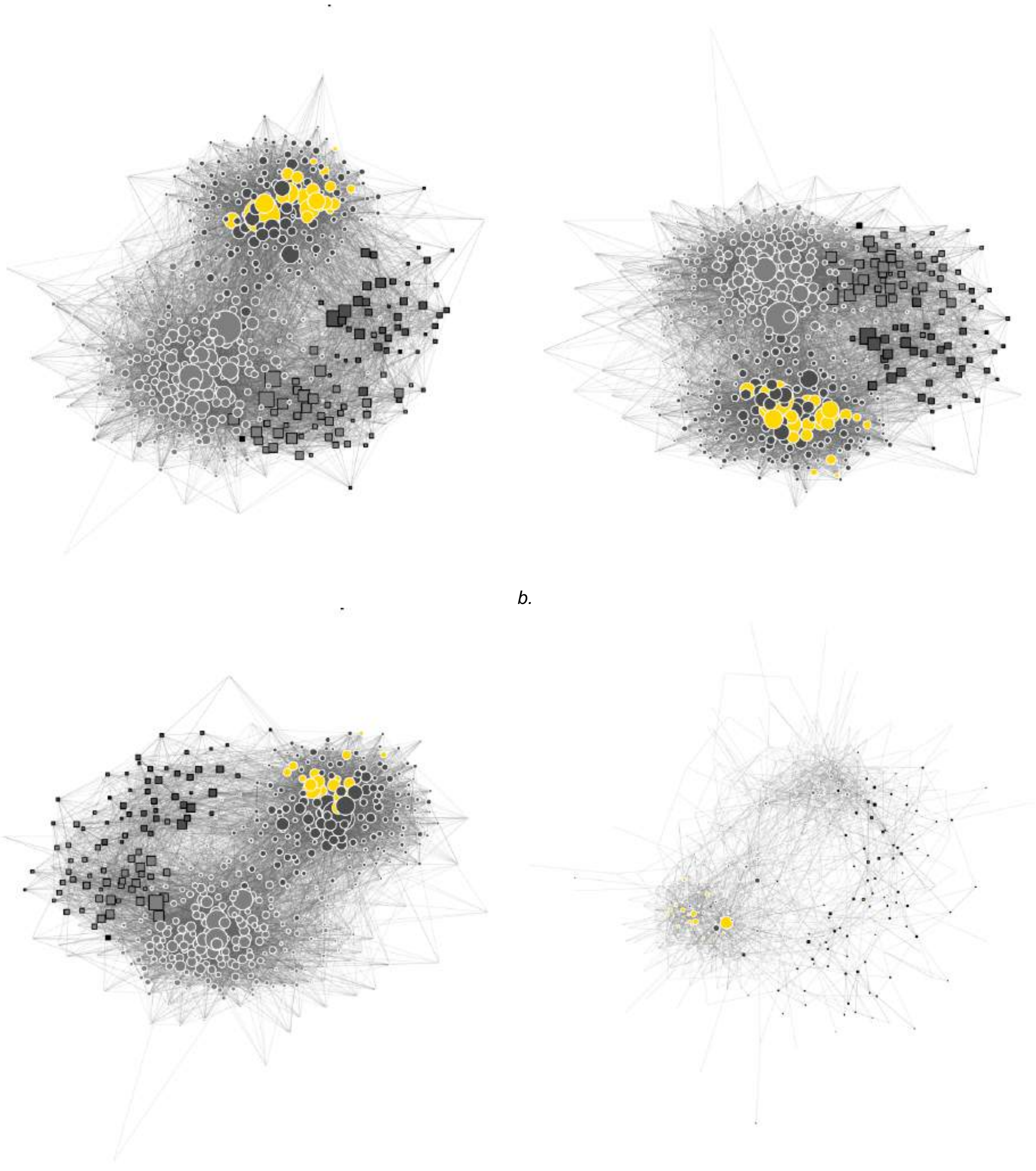
Figure 15. Proportions of Different Triad Types in Four Networks. *Each bar shows the percentage of different triad types. The most common types are "003", "012", "102", which represent an empty triad (A, B, C), a directed connection between two nodes (A→B, C), and a mutual connection (A↔B, C), respectively. Stronger connections lead to fewer triangular relationships. Like transitivity, this measure also does not involve edge weights, so the ratios in the Net_ all weight and Net_ all are the same.*

Subgroups & Communities

Cliques:

Network	*Net_all_weighted*	*Net_all*	*Net_moderate*	*Net_active*
Max number of nodes	*13*	*13*	*12*	*5*

Table 12. Maximum Size of Cliques in Different Networks. *This table shows the largest number of connected members in various network groups. Clique algorithm is not eligible for direction and weights, both Net_all_weighted and Net_all have the same results.*

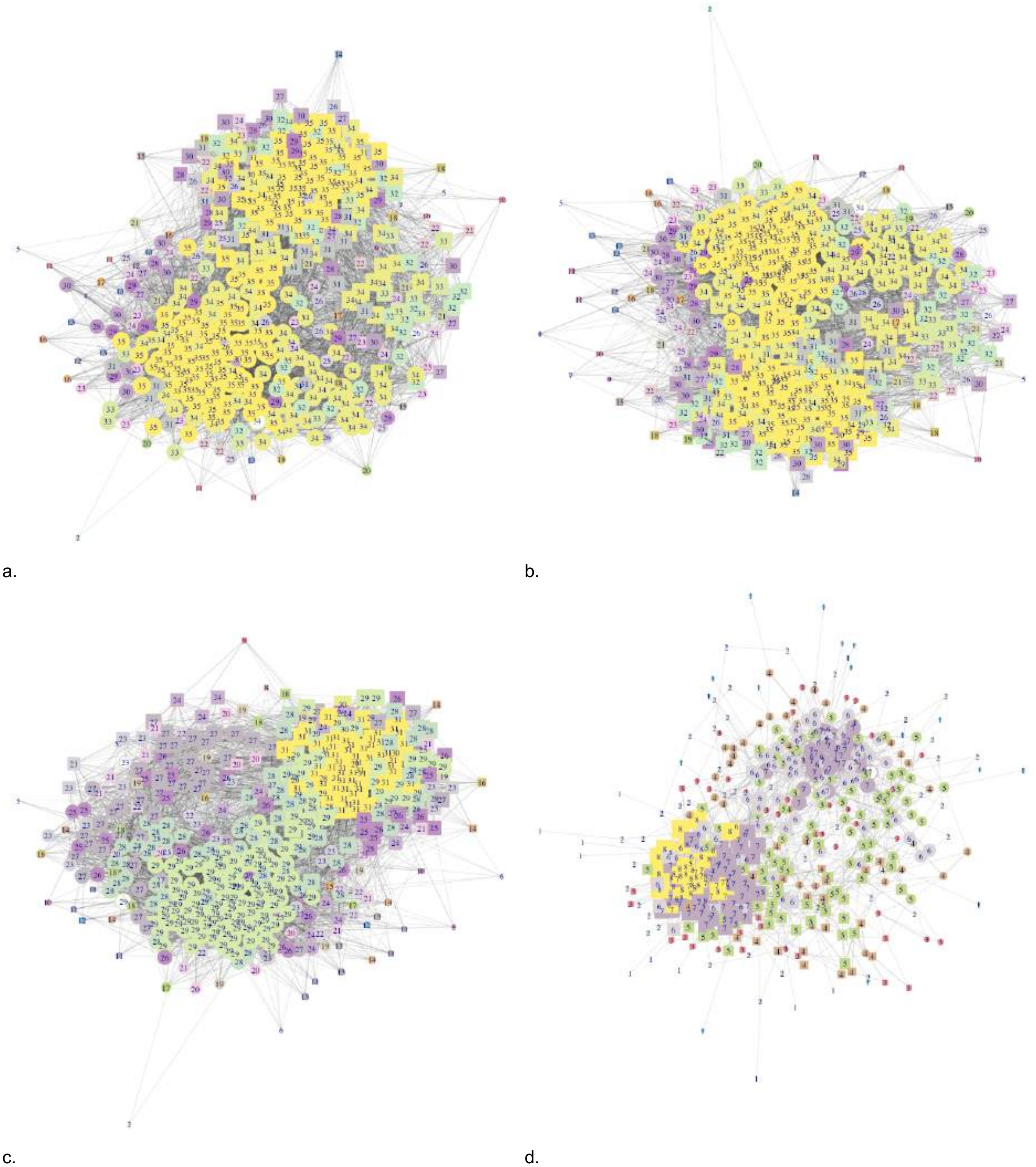


a. *b.* *c.* *d.*

Figure 16. Maximum Size of Cliques in Different Networks. *This figure displays biggest cliques in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. Deep grey nodes represent Republicans, light grey nodes represent Democrats, and black nodes represent Independents. White-framed nodes are Representatives, while black-framed nodes are Senators. The largest cliques appear among Republican Representatives in all networks.*

The clique algorithm applied here is designed for undirected networks, limiting its ability to detect groups in directed networks. However, it still offers valuable preliminary insights. Figure 16 explores that the largest cliques are found among Republican Representatives. As weaker connections are filtered out, the network becomes sparser and more dispersed, which is shown by the decrease in the maximum number of clique nodes from 13 to 5, as indicated in Table 12.

K-core Decomposition:



a. b. c. d.

Figure 17. K-core Decomposition in Different Networks. *This figure shows the core structure of various networks, where nodes are coloured based on their core level (more connected nodes are in yellow). Panels represent different network configurations: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active, illustrating how the tightly connected core areas change, as the weaker connections are removed. K-core decomposition does not account for edge weights.*

K-core decomposition provides a layered view of a network, from highly connected cores to less connected ones, offering a better understanding of the graph organization and levels. Figure 17 shows that members of the House of Representatives generally have more connections. As weak connections are removed, the difference between Representatives and Senators becomes more noticeable. Additionally, there is a gap between the two parties, suggesting that Democratic Representatives and Senators have more weak connections. When moderate connections are further removed, this difference becomes even clearer, indicating that Republican Representatives have more strong and active connections and play more central roles in the network.

Community Detection:

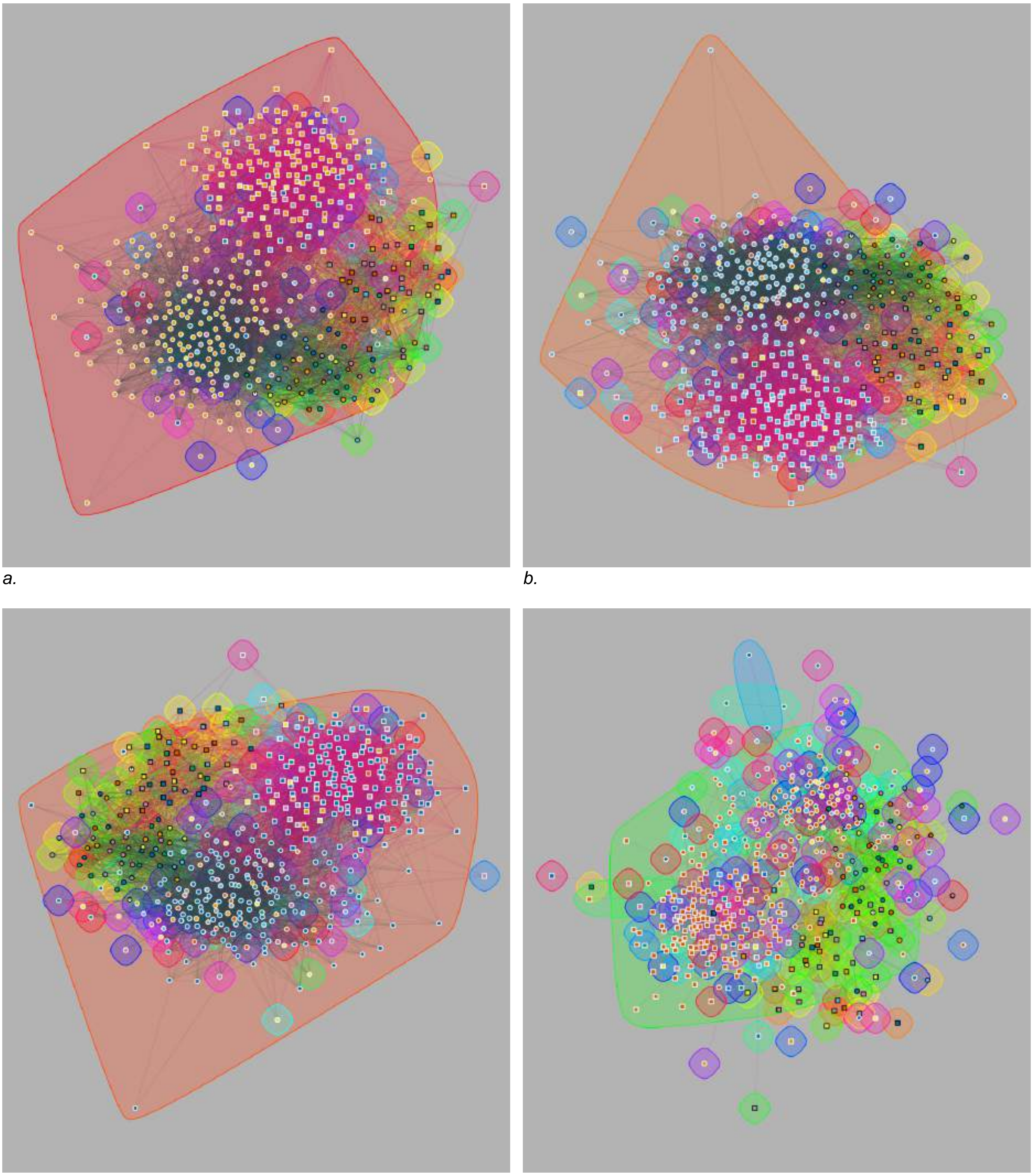
Next, four community detection algorithms were used to detect the communities within the network. The communities represent tightly connected groups with dense internal links and fewer external connections. Modularity indicates the density within communities, with higher modularity values indicating stronger and more well-defined communities. The number of edges across communities shows how many connections span different communities. Table 13 provides quantitative results, while Figures 18 to 21 offer visual insights to verify whether communities are well detected. Here is a summary of the results:

- **Based on Edge Betweenness:** This method results in many small, scattered communities across all networks, as shown in Figure 18. It identifies a high number of communities (all over 200), but these divisions are less meaningful and cohesive. The low modularity and high number of edges crossing between communities suggest that the detected communities are not well-defined.
- **Based on Propagating Labels:** This algorithm achieves relatively high modularity with fewer communities and identifies the Democratic and Republican communities (Figure 19). However, especially in the Net_active, it detects 79 communities, leading to overly detailed and less practical community divisions.
- **Based on Greedy Optimization:** This approach provides high modularity (all over 0.4) with the fewest number of communities (4, 3, 3, 8), indicating high-quality community divisions and tighter communities. It effectively balances the number of communities with the quality of internal connections, making it one of the most effective algorithms for detecting well-defined groups.
- **Based on Random Walk:** This method shows nearly the highest modularity in all networks. Notably, it identifies two distinct communities in the Senate, as shown in Figure 21a, which other algorithms did not detect in the original weighted network. However, in Net_active, it identifies many more communities, suggesting that it may over-detect and get trapped due to the loose strong connections.

Overall, The Greedy Optimization and Random Walk methods perform best, each excelling in different contexts—Greedy Optimization in loose networks (Figure 20d) and Random Walk in weighted networks (Figure 21a). Weak and moderate connections mainly form three distinct communities: Republican Representatives, Democratic Representatives, and the Senate. In contrast, strong connections have not formed clear and distinct groups except for two main communities in the House of Representatives, due to loose connections. In the original weighted network, four communities are detected, representing the two parties in the House of Representatives and the Senate, showing that political affiliation and house membership are key factors in shaping these communities in the Congressional Twitter network.

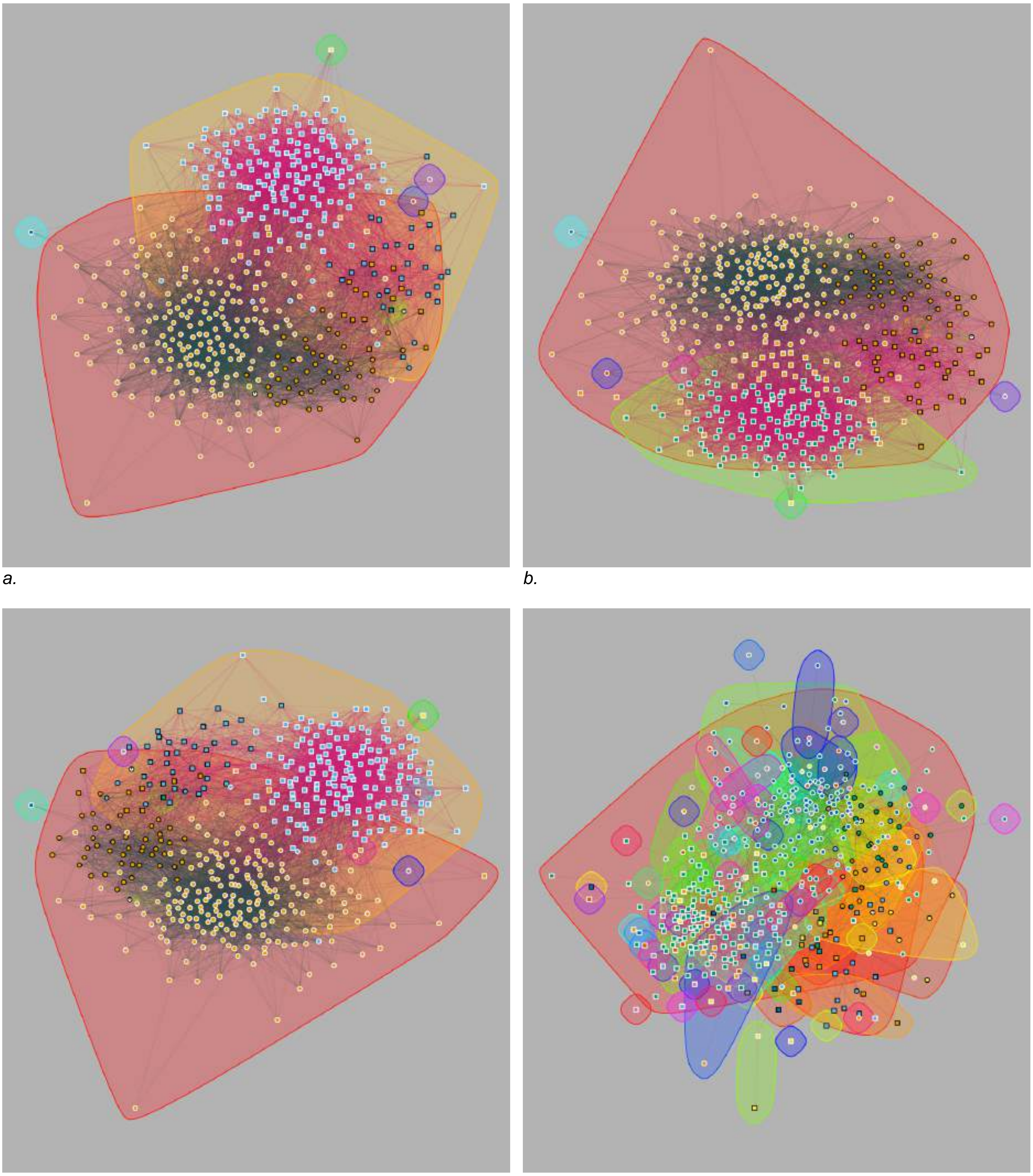
Network	***Net_all_weighted***	***Net_all***	***Net_moderate***	***Net_active***
Based on edge betweenness				
Modularity	*0.039*	*0.048*	*0.040*	*0.071*
Number of communities	*203*	*245*	*233*	*239*
Edges across communities	*8338//13289*	*8540/13289*	*7295/11263*	*1267/2008*
Based on propagating labels				
Modularity	*0.370*	*0.297*	*0.367*	*0.266*
Number of communities	*8*	*8*	*9*	*79*
Edges across communities	*2068/13289*	*1949/13289*	*1541/11263*	*830/2008*
Based on greedy optimization of modularity				
Modularity	*0.438*	*0.400*	*0.416*	*0.480*
Number of communities	*4*	*3*	*3*	*8*
Edges across communities	*3487/13289*	*2631/13289*	*2076/11263*	*415/2008*
Based on random walk				
Modularity	*0.433*	*0.414*	*0.431*	*0.492*
Number of communities	*4*	*3*	*3*	*31*
Edges across communities	*3522/13289*	*2627/13289*	*2045/11263*	*531/2008*

Table 13. Community Detection in Different Networks. *This table compares the modularity, number of communities, and the number of edges between communities in four networks using four different algorithms: edge betweenness, label propagation, greedy optimization of modularity, and random walk. Higher modularity values indicate stronger community structure, while the number of communities and edges across them show how the community are divided.*



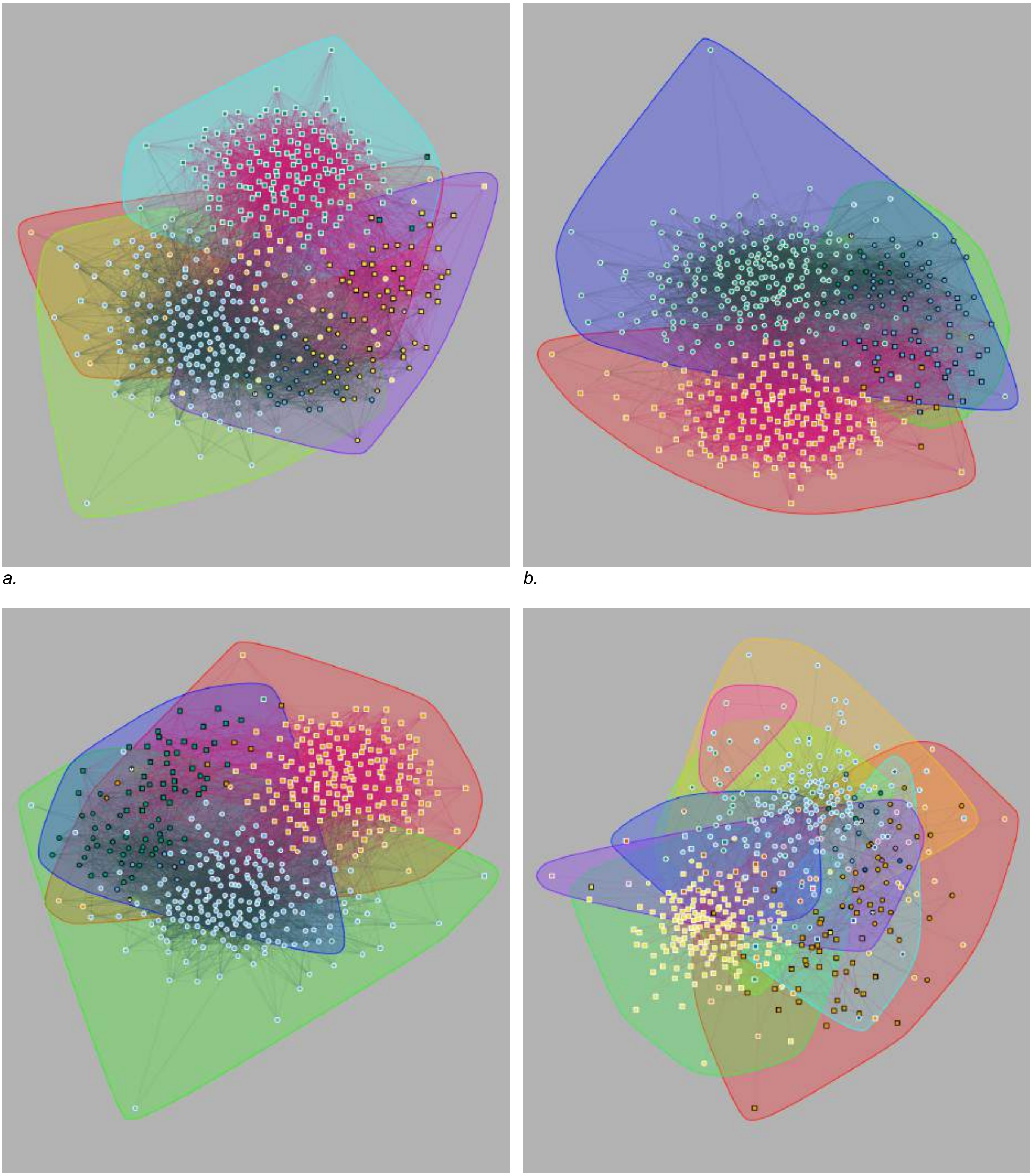
a. b. c. d.

Figure 18. Community Detection based on Edge Betweenness in Different Networks. *These diagrams illustrate how the edge betweenness method detects communities in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. White-framed squares represent Republican Representatives, white-framed circles represent Democratic Representatives, black-framed squares represent Republican Senators, and black-framed circles represent Democratic Senators. Each coloured circle represents a different community. The diagrams show one large community in each network, with many small groups consisting of only a single member.*



a. b. c. d.

Figure 19. Community Detection based on Propagating Labels in Different Networks. *These diagrams show how communities are identified using the propagating labels method in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. White-framed squares represent Republican Representatives, white-framed circles represent Democratic Representatives, black-framed squares represent Republican Senators, and black-framed circles represent Democratic Senators. Each coloured circle represents a different community. In the first three networks, the algorithm clearly distinguishes between Democratic and Republican groups, along with a few isolated points. In the Net_active network, however, the algorithm detects many overlapping and unclear communities.*



a. b. c. d.

Figure 20. Community Detection based on Greedy Optimization of Modularity in Different Networks. *These diagrams show how communities are identified using the greedy optimization of modularity algorithm in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. White-framed squares represent Republican Representatives, white-framed circles represent Democratic Representatives, black-framed squares represent Republican Senators, and black-framed circles represent Democratic Senators. Each colored circle indicates a separate community. In the Net_all and Net_moderate networks, three main communities are identified: Democratic Representatives, Republican Representatives, and Senators. Some borderline points are placed into new communities in the original weighted network, and smaller, less significant groups are found in the Net_active.*

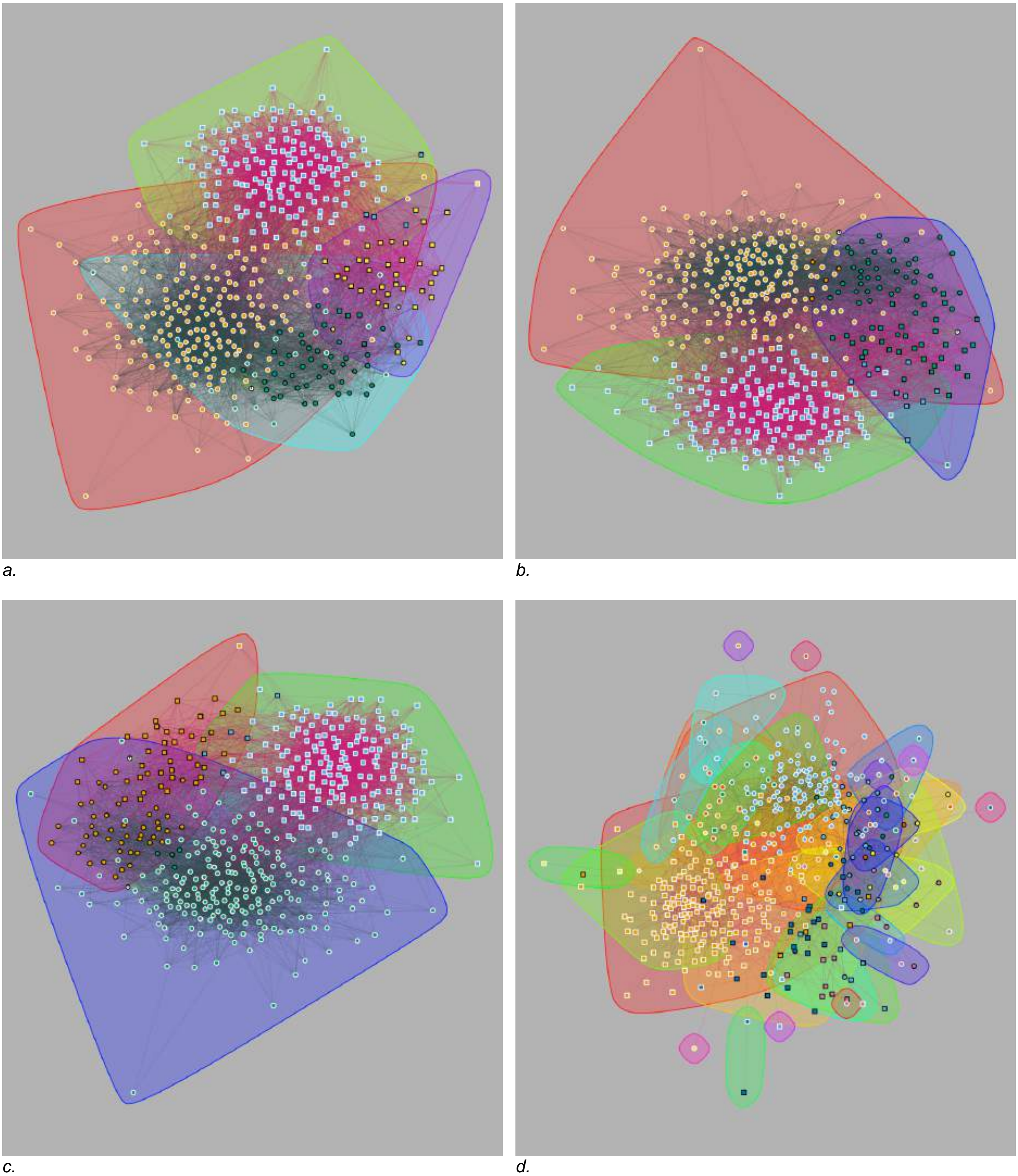


Figure 21. Community Detection based on Random Walk in Different Networks. *These diagrams show the use of the random walk algorithm to detect communities in four networks: (a) Net_all_weighted, (b) Net_all, (c) Net_moderate, and (d) Net_active. White-framed squares represent Republican Representatives, white-framed circles represent Democratic Representatives, black-framed squares represent Republican Senators, and black-framed circles represent Democratic Senators. Each colored circle indicates a separate community. The algorithm identifies three main communities in the Net_all and Net_moderate: Democratic Representatives, Republican Representatives, and Senators. In the Net_active, additional fragmented communities are detected alongside the main Democratic Representatives and Republican Representatives groups. In the weighted network, the algorithm finds four communities, roughly corresponding to Democratic Representatives, Republican Representatives, Democratic Senators, and Republican Senators.*

5.2. Fitting, Diagnosing, and Comparing Networks modeled by ERGMs

Term Summary

	TERM.NET_ALL	TERM.NET_MODERATE	TERM.NET_ACTIVE
edges	13289	11263	2008
mutual	3067	2212	213
nodefactor.Party.I	68	62	24
nodefactor.Party.R	12477	10843	2224
nodefactor.Seniority.Low	8350	7191	1282
nodefactor.Seniority.Medium	11329	9616	1654
nodefactor.House.Senate	4737	3719	631
nodematch.Party	11364	9720	1751
nodematch.State	1831	1649	465
nodematch.Seniority.High	1098	905	171
nodematch.Seniority.Low	1534	1353	240
nodematch.Seniority.Medium	2475	2106	346
nodematch.House	11678	10046	1809

Table 14. Summary of Network Terms. *This table presents a summary of network terms involved in Exponential Random Graph Models (ERGMs) across different networks: Net_all, Net_moderate, and Net_ative. The terms include the number of edges, mutual connections, and node attributes like party affiliation and seniority. By comparing these terms, the table helps to understand the structural differences between these network models.*

The choice of these terms stems from the descriptive analysis above, such as the factors of party and house, which have a significant impact on the network. However, the impact of state and seniority is difficult to observe in descriptive analysis. Here still including them, ERGM is used to explore whether they indeed have a statistically significant impact on the network. Another reason is the premise that the ERGM model with these terms would not degrade or fail due to insufficient MCMC sample size or high correlation. Summarising terms helps us understand their significance and identify potential issues that may lead to poor modelling results. For example, in Table 14, there are 3067 mutual dyads in Net_all, indicating 6134 of the 13289 ties are reciprocated, where both directional ties are present. Another example is that there are 2224 connections of the Republican Party, exceeding the total count of 2008 in Net_active, this is because it aggregates outgoing and incoming links, resulting in this phenomenon. The previous part descriptive analysis has described the patterns of these terms in detail, these values complement the description of the specific number of edges related to the specific term in the networks. All the TERMs have a significant number of edges, except for that about independent parties due to only two members in this party, with small number of edges. Nevertheless, since transitivity corresponding to the term “triangle” leads to model degeneracy in model fitting, more complex terms like geometrically weighted edgewise shared partnerships (“gwesp”) should be used instead. However, due to the high computational complexity, which could take several hours or even a day, it is not considered here. And because there are too many states, only whether they come from the same state is included, without considering the state itself.

MCMC Diagnostics

The MCMC diagnostic checks ensure that the model simulation process accurately reflects the network structure. According to Appendices 1, 3 & 5, sample statistic auto-correlation measures the correlation between sample statistics at different points in the MCMC chain. A well-mixed chain shows low autocorrelation close to 0 after Lag 0 in the results of all networks, indicating a low autocorrelation in sample statistics. Sample statistic burn-in diagnostic (Geweke) shows no low p-values of sample statistics, indicating good MCMC simulation convergence in all networks. Notably, 0.231, 0.582, and 0.337 joint p-values in Net_all, Net_moderate, and Net_all respectively, imply better performance of MCMC simulation on moderate connections. Furthermore, in all three networks, MCMC trace plots demonstrate mixing centred around 0, meaning a good simulation between sample statistics and the observed network at each simulation step. MCMC density plots display a normal distribution centred around 0, implying no obvious difference from the observed network. Overall, consistent trace plots and low autocorrelation suggest the models are all well-calibrated and the estimates are reliable.

Model Fitting Results

As shown in Figure 22, the p-values for medium seniority are much higher than 0.05 across all networks, indicating that this factor does not significantly affect the formation of connections. Additionally, despite showing statistical significance, the effect of the independent party on network connections is negligible due to the small number of members and their limited ties. Other statistically significant factors are:

- **Edges:** it represents the basic tendency of nodes to form connections. A significant negative coefficient across all networks suggests connections are less likely to form than expected by chance, with increasing negative value from -5.58 to -7.73 indicating a lower probability of stronger connection formation.
- **Mutual:** A positive coefficient indicates a strong tendency for reciprocal connections between members.
- **Republican:** Being a Republican increases the likelihood of forming connections compared to Democrats. This effect is particularly strong in forming robust connections, evident from the increase coefficient from 0.095 to 0.261.
- **Low Seniority:** Members with low seniority are less likely to form connections, especially strong ones, as indicated by the more negative coefficient (-0.128) in Net_all.
- **Senate:** Senators are more likely to form connections than Representative members, potentially due to House connections nearing saturation.
- **Party Homophily:** Members are more likely to connect with others from the same party.
- **State Homophily:** Similarly, members from the same state are also more likely to connect. This trend of party and state homophily is more significant in stronger connections, especially for state homophily as seen from the stronger coefficient increase (from 1.15 to 1.72).
- **Seniority Homophily:** Members with similar high seniority tend to connect more, while those with medium seniority are less inclined. However, this pattern is more evident in weaker connections as their p-values in the strong connections show that they are not statistically significant. Nevertheless, members with similar low seniority are more likely to connect with each other in all kinds of strength connections.
- **House Homophily:** Members from the same chamber show a strong tendency to connect, especially in forming strong connections, as seen coefficient rise from 1.34 to 1.60.

Overall, the likelihood of forming connections in these networks is generally low. However, being from the same party, state, or legislative chamber significantly increases the chance of forming relationships, particularly stronger ones. Republicans, senior members, and Senators are more actively engaged in interactions, though seniority plays a weaker role in forming relationships compared to other factors. These all reflect the impact of political affiliation, state, seniority, and legislative chamber on network dynamics. Additionally, the network favours mutual connections.

```
Call:
ergm(formula = net.01 ~ edges + mutual + nodefactor("Party") +
    nodefactor("Seniority") + nodefactor("House") + nodematch("Party") +
    nodematch("State") + nodematch("Seniority", diff = T) + nodematch("House"),
    verbose = TRUE)

Monte Carlo Maximum Likelihood Results:

                            Estimate Std. Error MCMC %  z value Pr(>|z|)
edges                      -5.578636   0.046077      0 -121.072  < 1e-04 ***
mutual                      2.557967   0.031806      0   80.425  < 1e-04 ***
nodefactor.Party.I          0.654649   0.115626      0    5.662  < 1e-04 ***
nodefactor.Party.R          0.095451   0.009127      0   10.458  < 1e-04 ***
nodefactor.Seniority.Low   -0.080545   0.024324      0   -3.311 0.000929 ***
nodefactor.Seniority.Medium -0.024250   0.026691      0   -0.909 0.363594
nodefactor.House.Senate     0.471546   0.013873      0   33.991  < 1e-04 ***
nodematch.Party             1.484681   0.023783      0   62.426  < 1e-04 ***
nodematch.State             1.154560   0.025684      0   44.952  < 1e-04 ***
nodematch.Seniority.High    0.250094   0.043623      0    5.733  < 1e-04 ***
nodematch.Seniority.Low     0.220765   0.040033      0    5.515  < 1e-04 ***
nodematch.Seniority.Medium  -0.074970   0.036244      0   -2.068 0.038594 *
nodematch.House             1.344189   0.026639      0   50.459  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 312124  on 225150  degrees of freedom
 Residual Deviance:  80239  on 225137  degrees of freedom

AIC: 80265  BIC: 80399  (Smaller is better. MC Std. Err. = 5.218)
```

a.

```
Call:
ergm(formula = net.02 ~ edges + mutual + nodefactor("Party") +
    nodefactor("Seniority") + nodefactor("House") + nodematch("Party") +
    nodematch("State") + nodematch("Seniority", diff = T) + nodematch("House"),
    verbose = TRUE)

Monte Carlo Maximum Likelihood Results:

                             Estimate Std. Error MCMC %  z value Pr(>|z|)
edges                      -5.9145784  0.0527349      0 -112.157  < 1e-04 ***
mutual                      2.2475539  0.0348540      0   64.485  < 1e-04 ***
nodefactor.Party.I          0.9022749  0.1240503      0    7.273  < 1e-04 ***
nodefactor.Party.R          0.1237262  0.0098758      0   12.528  < 1e-04 ***
nodefactor.Seniority.Low   -0.0678750  0.0262244      0   -2.588  0.00965 **
nodefactor.Seniority.Medium  0.0004553  0.0304862      0    0.015  0.98809
nodefactor.House.Senate     0.4442365  0.0154867      0   28.685  < 1e-04 ***
nodematch.Party             1.6077175  0.0276692      0   58.105  < 1e-04 ***
nodematch.State             1.2376795  0.0274768      0   45.045  < 1e-04 ***
nodematch.Seniority.High    0.2790289  0.0487034      0    5.729  < 1e-04 ***
nodematch.Seniority.Low     0.2509362  0.0434268      0    5.778  < 1e-04 ***
nodematch.Seniority.Medium  -0.0918940  0.0414360      0   -2.218  0.02657 *
nodematch.House             1.4795841  0.0316839      0   46.698  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 312124  on 225150  degrees of freedom
 Residual Deviance:  72422  on 225137  degrees of freedom

AIC: 72448  BIC: 72582  (Smaller is better. MC Std. Err. = 4.562)
```

b.

```
Call:
ergm(formula = net.03 ~ edges + mutual + nodefactor("Party") +
    nodefactor("Seniority") + nodefactor("House") + nodematch("Party") +
    nodematch("State") + nodematch("Seniority", diff = T) + nodematch("House"),
    verbose = TRUE)

Monte Carlo Maximum Likelihood Results:

                            Estimate Std. Error MCMC % z value Pr(>|z|)
edges                       -7.72917    0.12871      0 -60.052   <1e-04 ***
mutual                       2.45033    0.09637      0  25.426   <1e-04 ***
nodefactor.Party.I           1.80163    0.21202      0   8.497   <1e-04 ***
nodefactor.Party.R           0.26078    0.02353      0  11.083   <1e-04 ***
nodefactor.Seniority.Low    -0.12798    0.06275      0  -2.039   0.0414 *
nodefactor.Seniority.Medium -0.13349    0.07098      0  -1.881   0.0600 .
nodefactor.House.Senate      0.43528    0.03534      0  12.316   <1e-04 ***
nodematch.Party              1.76388    0.06901      0  25.559   <1e-04 ***
nodematch.State              1.72051    0.05487      0  31.355   <1e-04 ***
nodematch.Seniority.High     0.17871    0.11099      0   1.610   0.1074
nodematch.Seniority.Low      0.20489    0.10164      0   2.016   0.0438 *
nodematch.Seniority.Medium  -0.08624    0.09575      0  -0.901   0.3677
nodematch.House              1.60489    0.07860      0  20.419   <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 287628  on 207480  degrees of freedom
 Residual Deviance:  19003  on 207467  degrees of freedom

AIC: 19029  BIC: 19162  (Smaller is better. MC Std. Err. = 1.551)
```
c.

Figure 22. Model Fitting Results. *This figure shows the output of fitting an Exponential Random Graph Model (ERGM) to the three networks: (a) Net_all, (b) Net_moderate, and (c) Net_active. The model includes several factors such as the number of edges, mutual ties, and node attributes like party affiliation, state, seniority, and membership in the House or Senate. The coefficients (Estimate) represent the strength and direction of effect of each factor on the likelihood of network formation. Positive coefficients indicate factors that increase the likelihood of connections, while negative coefficients suggest factors that decrease it. Significant values (marked with ***, **, *) indicate that the corresponding factors have a statistically significant impact on the network formation and structure.*

Network Simulation

	OBS.NET_ALL	SIM MEAN.NET_ALL	OBS.NET_MODERATE	SIM MEAN.NET_MODERATE	OBS.NET_ACTIVE	SIM MEAN.NET_ACTIVE
edges	13289	13364	11263	11316	2008	2027
mutual	3067	3074	2212	2238	213	220
nodefactor.Party.I	68	75	62	59	24	22
nodefactor.Party.R	12477	12579	10843	10827	2224	2199
nodefactor.Seniority.Low	8350	8420	7191	7207	1282	1255
nodefactor.Seniority.Medium	11329	11392	9616	9668	1654	1683
nodefactor.House.Senate	4737	4869	3719	3783	631	632
nodematch.Party	11364	11408	9720	9773	1751	1778
nodematch.State	1831	1893	1649	1673	465	492
nodematch.Seniority.High	1098	1096	905	926	171	162
nodematch.Seniority.Low	1534	1539	1353	1320	240	221
nodematch.Seniority.Medium	2475	2479	2106	2132	346	346
nodematch.House	11678	11744	10046	10102	1809	1832

Table 15. Comparison of Simulated and Observed Network Statistics. *This table compares the observed network statistics with the mean statistics from simulations in different networks: Net_all, Net_moderate, and Net_all. It helps determine whether the simulations closely match the real network data by comparing key metrics such as the number of edges, mutual connections, and various nodal attributes (e.g., party affiliation, seniority, and house membership). The closer the simulated mean values are to the observed values, the more accurately the simulation represents the real network.*

A preliminary and simple validation of the goodness of fit of the model can be achieved by comparing simulated and observed network statistics. Table 15 shows that the average simulated values across all three networks are close to the original values, suggesting a good fit for these specific statistics. However, a more detailed assessment requires using a goodness-of-fit function, as explained below.

Goodness-of-Fit

The Goodness-of-Fit diagnostic evaluates how well the model reproduces global network properties not explicitly included in the model, by comparing the structural measures of the observed network with those from simulated networks based on the fitted parameters. For directed graphs, this typically includes model statistics, out-degree, in-degree, edgewise shared partner, and minimum geodesic distance, though not all ERGM terms are supported. From Appendix 2, it is evident that the simulated model fails to capture the global structure of Net_all, with most observed statistics falling outside the simulated range. However, as weaker connections are removed, the model fit improves. In Net_moderate (Appendix 4), all but two statistics fall within the simulated range, and in Net_active (Appendix 6), all model statistics and other measures show better overlap, indicating a better fit for strong connections. This suggests that the model terms are more effective in capturing the formation of strong connections in this Congressional Twitter network.

However, the poor fit for out-degree, in-degree, edgewise shared partner, and minimum geodesic distance still exists, implying that additional terms may be needed. While in-degree and out-degree terms target specific degree counts (e.g., 1 degree, 2 degrees), more comprehensive terms like "gwidegree" and "gwodegree" might be necessary to cover the full degree distribution. Similarly, K-transitivity and K-twopath, which are related to edgewise shared partner and minimum geodesic distance, require the use of "gwesp" and "gwdsp" terms. However, due to their high computational complexity, these terms are not discussed in the paper. The absence of these terms might contribute to the suboptimal Goodness-of-Fit performance.

5.3. Reflections on Research Questions and Hypotheses of Literature Review

Descriptive Analysis

- **Key Members:** The descriptive analysis highlighted three influential politicians: Kevin McCarthy, C. Scott Franklin, and Nancy Pelosi. McCarthy and Franklin primarily connect within their Republican circles, while Pelosi maintains a broader, cross-party network, reflecting their different interaction patterns and neighbour characteristics.
- **Impact of Partisanship:** Partisanship plays a significant role in shaping interactions, with results showing that members of the same party have more connections with each other. Republicans show higher interaction frequency and tighter connections compared to Democrats, especially in active interactions. This pattern may reflect the success of the Republican Party in reclaiming control of the House by a 9-seat margin in the November 8, 2022, U.S. election (Vakil, 2022).
- **Influence of the House:** Interactions are also influenced by the legislative chamber, with more frequent connections among members of the same chamber. Representatives, especially Republicans, serve as crucial hubs in the network, which may be related to the nature of their roles- Representatives having a large number, a short term, coming from the masses, being

closer to the public, needing to pay more attention to voters, and being responsible for drafting bills, which leads to higher activity on Twitter.

- **Four Distinct Subgroups:** The Congressional Twitter network is divided into four clear communities based on party and chamber affiliation: Republican Representatives, Democratic Representatives, Republican Senators, and Democratic Senators.
- **Interaction Patterns by Connection Weights:** Strong connections are scattered, with 2008 strong ties among 456 members. Reciprocity and transitivity are low for strong connections, indicating that people often only connect with specific nodes more frequently, and these connections are one-way and do not form cohesive groups, except among Republican Representatives who show clearer clusters. Key members tend to broadcast information in weak interactions and receive information in strong interactions, which aligns with the intuition of political-social interaction. Moreover, Democratic Representatives are broadening their interactions, while Republican Representatives focus more on key figures from their party.

Preferential Attachment and Homophily Principle

The network displays signs of Preferential Attachment to some extent, where influential politicians are more active in building and maintaining connections. However, contrary to initial hypotheses, seniority does not significantly impact these interactions. In contrast, the Homophily Principle is more evident, particularly regarding political stance (party affiliation) and position (parliamentary membership).

ERGMs Analysis

ERGM, as a statistical tool, provides a more quantitative view of network dynamics. It confirms the strong influence of political parties and parliamentary chambers on network structure, emphasizes the prominent role of the Republicans, supplements the influence of state homophily, which was not fully captured in descriptive analyses, and approves that homophily becomes more pronounced in stronger connections. Although junior members tend to interact more, seniority has only a minor overall impact.

The analysis also asserts a positive tendency for reciprocity in connections, despite the decreasing trend observed in descriptive analysis. This might be because reciprocity can appear in different types of connections—such as a strong link from A to B with a weaker response from B to A. The separation of connections by weight may obscure this, revealing the limitations of this method.

Interestingly, Senators are more likely to form new connections than Representatives, even though Representatives are more active in interactions. Due to computational constraints, the effects of factors like out-degree, in-degree, and transitivity on network formation were not fully validated, possibly resulting in poor model fit. The current ERGM perspective does not fully capture some global network properties, particularly those not explicitly included in the model. Future analyses should consider additional network statistics or member attributes, such as gender and political contributions.

6. Conclusion

6.1. Summary of Findings and Implications for Political Network

This study fills the important gap in political network research by conducting a detailed analysis of the Congressional Twitter network. Specifically, it measures the concentration of political interactions, identifies key influencers and internal groups, examines how various member attributes impact interaction behaviours, and explores the extent of mutual connections and interaction clusters. The study also implies that frequent Twitter interactions may correlate with political success and underscores the differences between political Twitter networks and traditional social networks, as compared to earlier research and theories. These insights are valuable for political network research and practical applications, such as political campaign management and digital political outreach.

In addition to showcasing the value and uniqueness of this network, the study provides a clear introduction to descriptive analysis and Exponential Random Graph Models (ERGMs), two widely used social network analysis tools. By combining both methods, the study offers a well-rounded understanding of the network. The approach of dividing the network based on connection strength allows the use of some metrics and ERGMs that typically do not support edge weights, thereby providing meaningful insights. The study demonstrates their powerful application in network analysis, but also their limitations for further research.

6.2. Limitations of the Study

Analysis Limitation: This network is not like a typical social network, and research cannot determine the underlying reasons for its non-similarity. Two real-world social network applications of Newman and Park (2003) demonstrated that the Collaboration Network has excellent agreement with the positive degree correlations and network transitivity, however, the Board of Directors network also shows a positive degree correlation, but its degree is lower than that of academic collaboration networks, and its motivation for connection is more complex, not solely dependent on simple social relationships. The author speculated that the remaining degree correlations may stem from genuine sociological or psychological factors, rather than just the topological structure in the model. This is consistent with, and even more specific to, the observed network objects studied here. The Congressional Twitter network has a nontrivial clustering and a negative degree correlation but close to 0, which does not conform to the high positive correlation and high transitivity characteristics described by Newman and Park in Methodology. This may suggest that the Congressional Twitter network is more complex than the Board of Directors network. However, the current analysis could not explain it well, as Newman and Park suggest, the underlying causes need to be explored in conjunction with factors such as psycho-social factors outside the network.

Tool Limitations: The analysis is constrained by the tools used, particularly the standard metrics like degree centrality and centralization, which do not account for edge weights. Additionally, ERGM models struggle with fitting complex terms related to out-degree, in-degree, and transitivity due to computational complexity. Issues such as model degeneracy, small sample sizes, and high correlations can also fail models. Accurately capturing global network properties through intricate local configurations can require extensive trial and error, sometimes taking hours or even days.

Data Limitations: The original data only provides basic connection information and its edge weights. Interactions like retweets, replies, and mentions are aggregated without distinguishing between them, and the content of these interactions is unknown. Furthermore, the node attributes included in the analysis are based on assumptions, and other potentially influential factors—such as gender, political reputation, or achievements—are not considered. To gain a more comprehensive understanding of the political Twitter network, more detailed data would be needed, and uncovering these factors may be costly.

6.3. Further Plan

Incorporate Additional Network Metrics and Terms or Explore Alternative Models: Future analyses should use metrics that account for edge weights and apply terms of Curved Exponential Family Models that handle out-degree, in-degree, and transitivity, which could reveal deeper insights into network dynamics and properties. Alternatively, testing different models, such as Stochastic Actor-Oriented Models (SAOMs), might better capture the complex behaviour of the network.

Expand Data Collection: Collect more granular data that distinguishes between different types of interactions, such as retweets, replies, and mentions, and includes content analysis of these interactions. Adding more detailed node attributes, such as gender, political reputation, and achievements, would provide a richer context and potentially uncover new patterns.

Conduct Comparative Studies: Comparing the Congressional Twitter network with other related political networks, such as the Congressional Voting network, can reveal unique characteristics and underlying factors. This comparison will help provide a more complete understanding of political behaviour among members of Congress.

Investigate Motivational Factors: Further research should explore the motivations behind the connections in this network, potentially through qualitative studies or surveys, to determine if factors beyond topological structure, such as political alliances or strategic communications, and assess their impact beyond just network structure.

7. Appendices

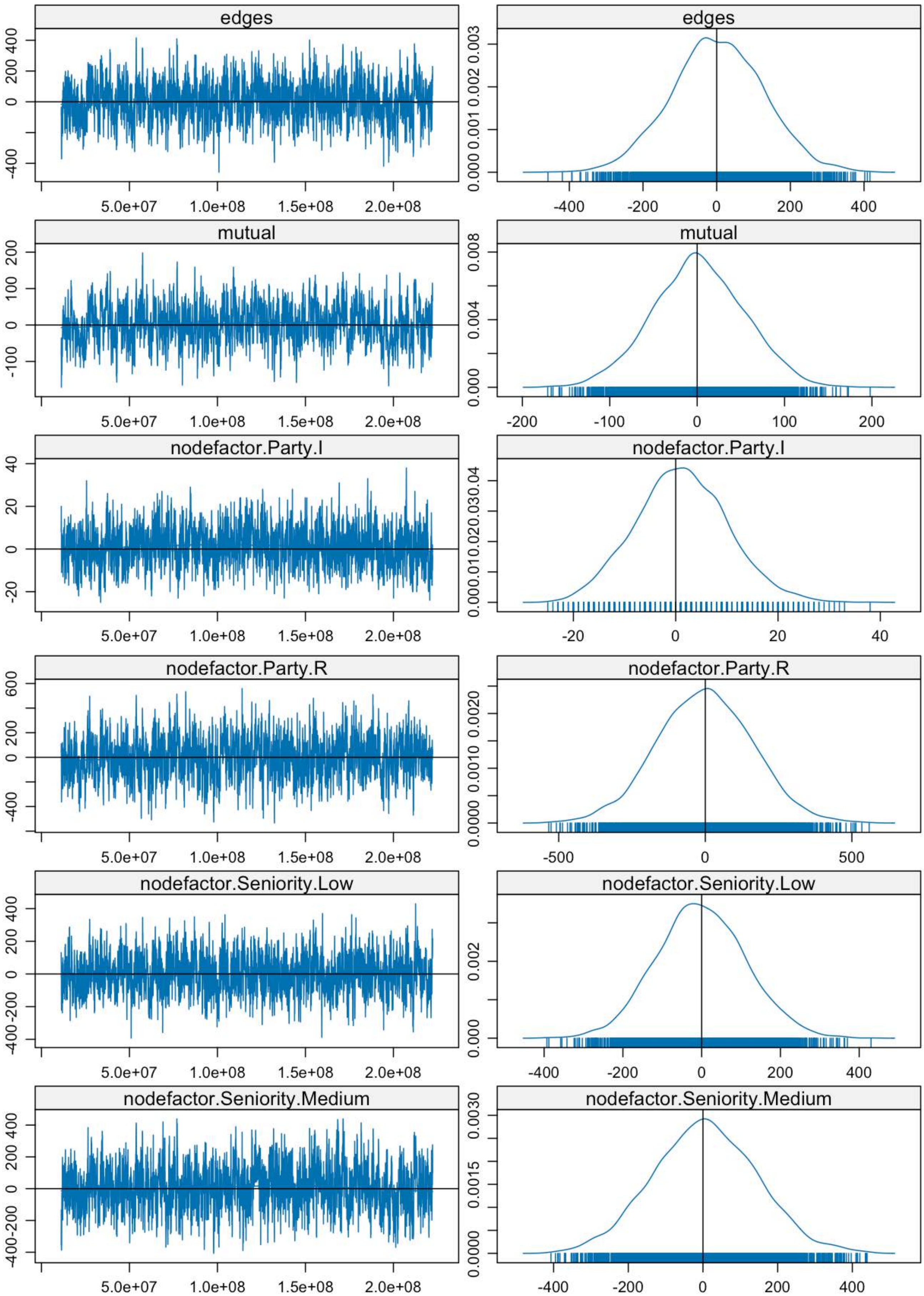
7.1. MCMC Diagnostic and Goodness-of-Fit for ERGM on Net_all

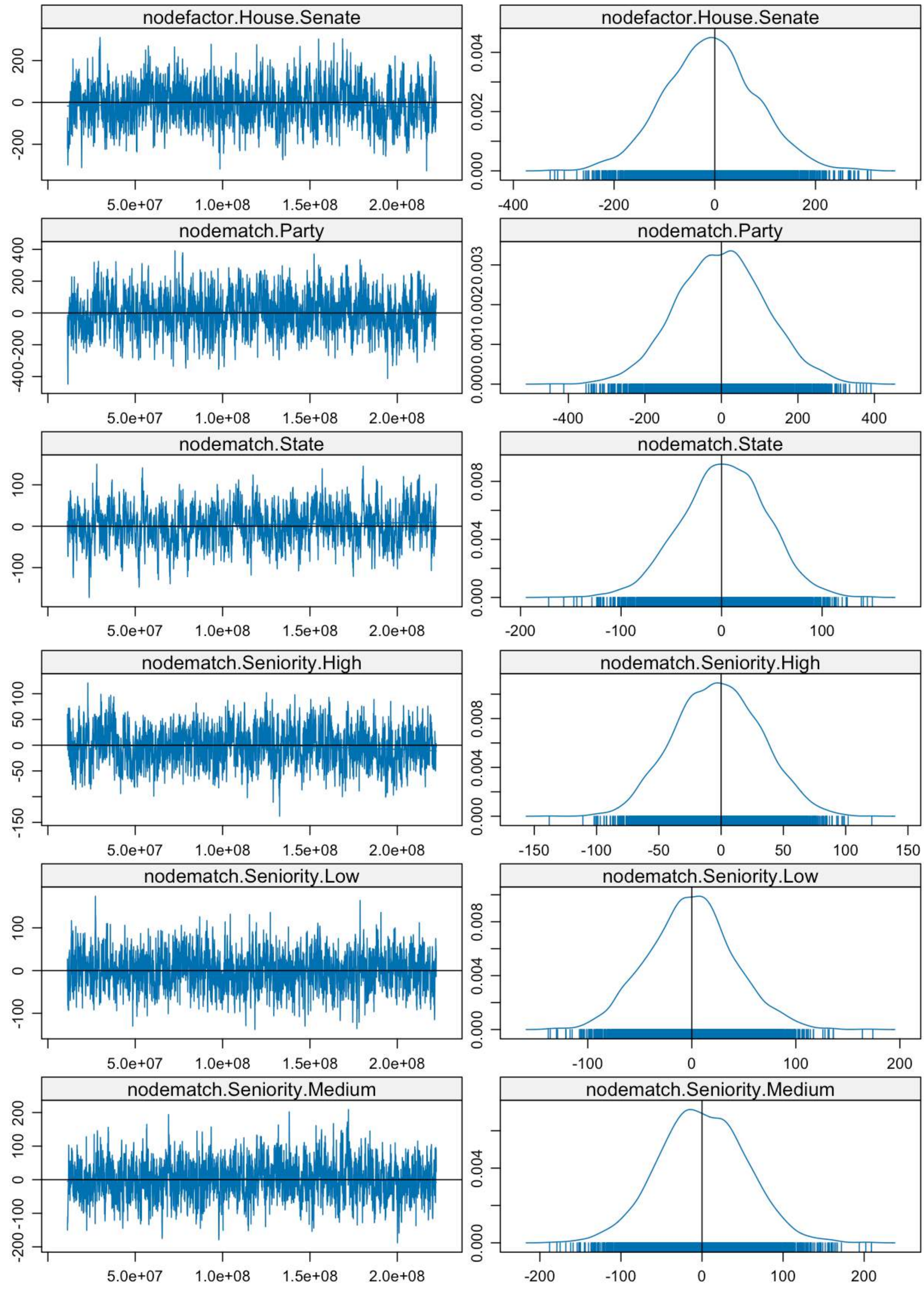
```
Sample statistics auto-correlation:
Chain 1
               edges    mutual nodefactor.Party.I nodefactor.Party.R nodefactor.Seniority.Low nodefactor.Seniority.Medium
Lag 0      1.0000000 1.0000000          1.0000000          1.0000000                1.0000000                   1.0000000
Lag 65536  0.5170849 0.7247938          0.4820617          0.5327944                0.4773448                   0.5116631
Lag 131072 0.3733161 0.5545214          0.3155929          0.3780467                0.3253971                   0.3698607
Lag 196608 0.2882560 0.4251532          0.2311990          0.2834759                0.2257422                   0.2902143
Lag 262144 0.2082029 0.3397605          0.1852443          0.1979661                0.1658130                   0.2144021
Lag 327680 0.1673414 0.2629227          0.1328218          0.1477500                0.1123401                   0.1886873
           nodefactor.House.Senate nodematch.Party nodematch.State nodematch.Seniority.High nodematch.Seniority.Low
Lag 0                    1.0000000       1.0000000       1.0000000                1.0000000              1.00000000
Lag 65536                0.6008101       0.5715968       0.7288725                0.5655947              0.49392326
Lag 131072               0.4585286       0.4220673       0.5930760                0.4036561              0.35300607
Lag 196608               0.3668710       0.3242729       0.4927529                0.3022661              0.24565278
Lag 262144               0.3044882       0.2439281       0.4026530                0.2223031              0.16817337
Lag 327680               0.2373070       0.2056119       0.3180191                0.1721553              0.09105881
           nodematch.Seniority.Medium nodematch.House
Lag 0                       1.0000000       1.0000000
Lag 65536                   0.5053059       0.5501364
Lag 131072                  0.3517236       0.4109478
Lag 196608                  0.2679591       0.3205705
Lag 262144                  0.2133391       0.2315689
Lag 327680                  0.1735165       0.1872495
```

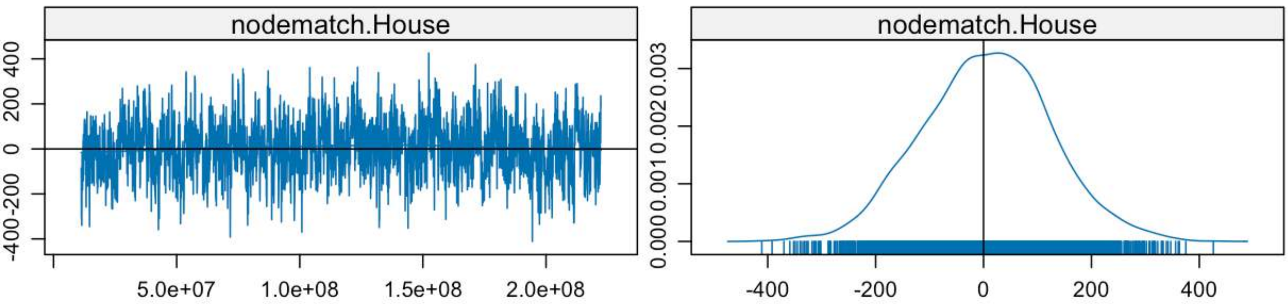
a. Sample Statistic Auto-correlation for Net_all. *This measure checks the correlation between sample statistics at different points in the MCMC chain. A well-mixed chain should have low autocorrelation (close to 0) after Lag 0. The results show a decreasing value to 0, indicating low autocorrelation, which is a positive outcome.*

```
Sample statistics burn-in diagnostic (Geweke):
Chain 1

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

                   edges                      mutual          nodefactor.Party.I          nodefactor.Party.R
              -1.6570483                  -1.4325419                  -1.0366241                  -1.4082348
nodefactor.Seniority.Low nodefactor.Seniority.Medium     nodefactor.House.Senate             nodematch.Party
              -1.0620532                  -1.6012242                  -0.2920503                  -1.5101174
         nodematch.State    nodematch.Seniority.High     nodematch.Seniority.Low  nodematch.Seniority.Medium
              -0.9879102                   1.0246403                   0.8608713                  -1.0006514
         nodematch.House
              -2.1405785

Individual P-values (lower = worse):
                   edges                      mutual          nodefactor.Party.I          nodefactor.Party.R
              0.09750972                  0.15198880                  0.29991110                  0.15906156
nodefactor.Seniority.Low nodefactor.Seniority.Medium     nodefactor.House.Senate             nodematch.Party
              0.28821151                  0.10932727                  0.77024813                  0.13101347
         nodematch.State    nodematch.Seniority.High     nodematch.Seniority.Low  nodematch.Seniority.Medium
              0.32319664                  0.30553294                  0.38930892                  0.31699535
         nodematch.House
              0.03230805
Joint P-value (lower = worse):  0.230778

Note: MCMC diagnostics shown here are from the last round of simulation, prior to computation of final parameter
  estimates. Because the final estimates are refinements of those used for this simulation run, these diagnostics
  may understate model performance. To directly assess the performance of the final model on in-model statistics,
  please use the GOF command: gof(ergmFitObject, GOF=~model).
```

b. Sample Statistic Burn-in Diagnostic (Geweke) for Net_all. *This diagnostic shows the p-values of sample statistics, with higher values indicating better MCMC convergence. Most individual p-values are high, and the joint p-value of 0.23 suggests satisfactory convergence.*

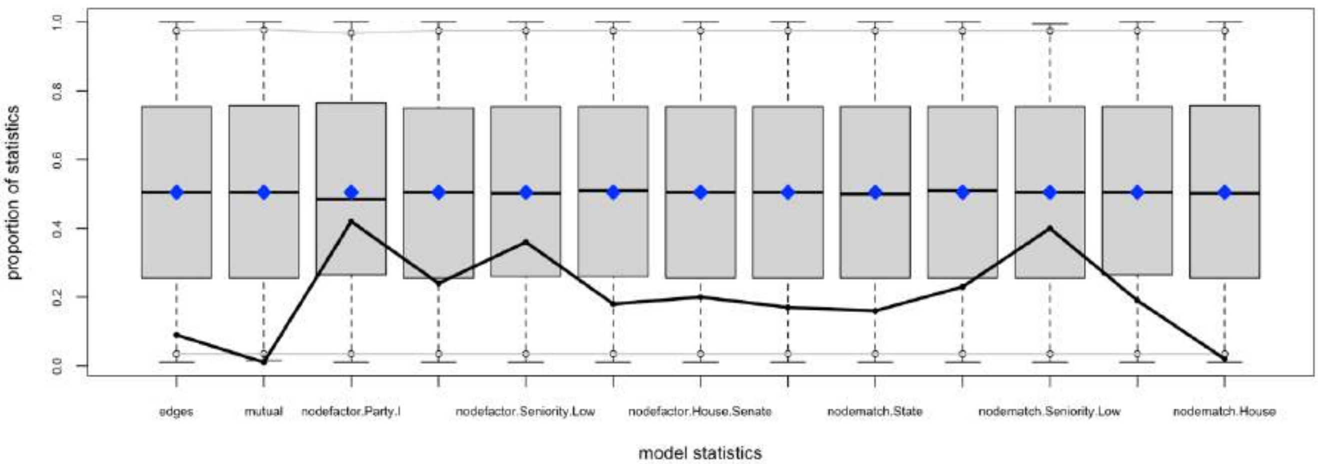

edges
mutual
nodefactor.Party.I
nodefactor.Party.R
nodefactor.Seniority.Low
nodefactor.Seniority.Medium
5.0e+07
1.0e+08
1.5e+08
2.0e+08


nodefactor.House.Senate
nodefactor.House.Senate
nodematch.Party
nodematch.Party
nodematch.State
nodematch.State
nodematch.Seniority.High
nodematch.Seniority.High
nodematch.Seniority.Low
nodematch.Seniority.Low
nodematch.Seniority.Medium
nodematch.Seniority.Medium
5.0e+07
1.0e+08
1.5e+08
2.0e+08

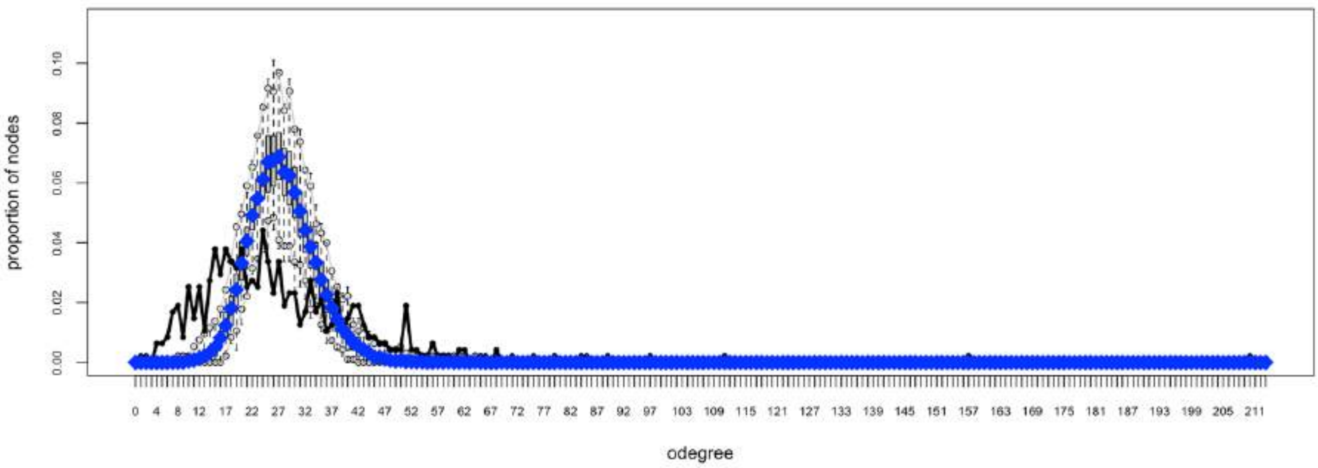


c. MCMC Trace and Density plots for Net_all. *MCMC trace plots demonstrate the sample statistics are well-mixed around 0, indicating a good match between the simulations and the observed network at each step. MCMC density plots, displaying a normal distribution centred around 0, suggest that the simulations closely align with the observed network.*

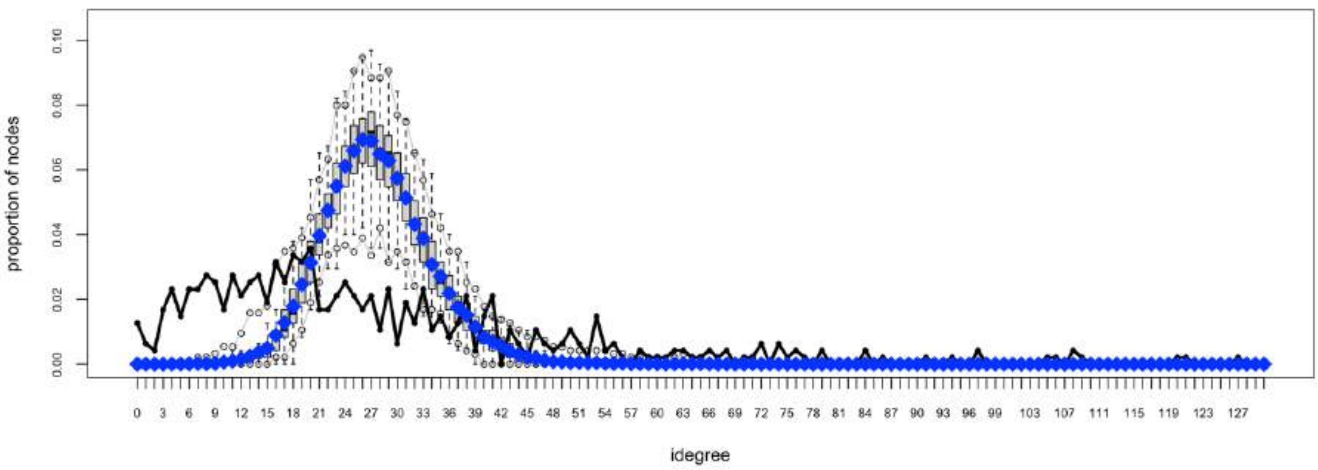
Appendix 1. Markov Chain Monte Carlo (MCMC) Diagnostics for Net_all



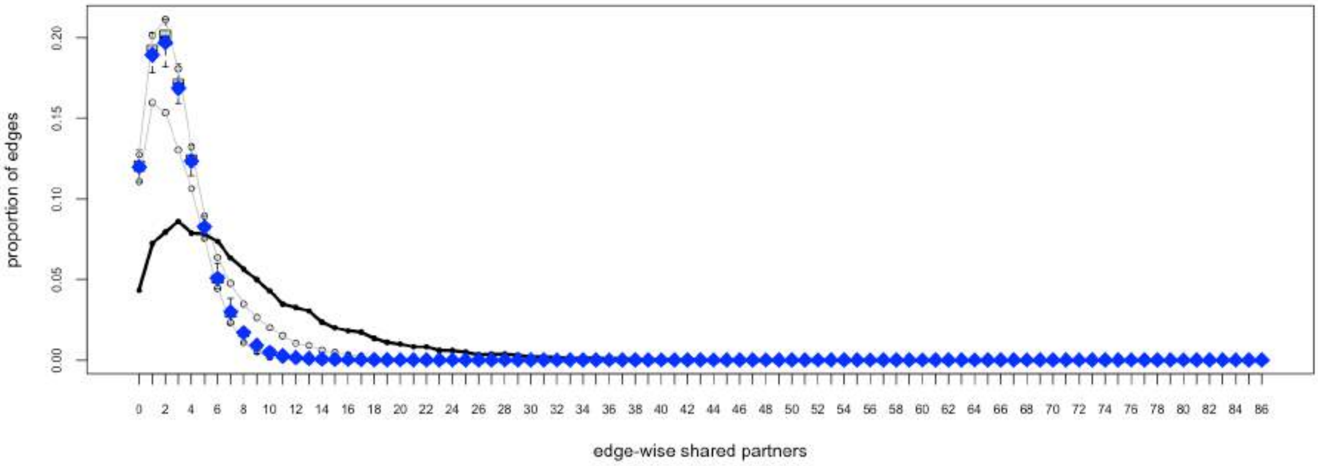
a.



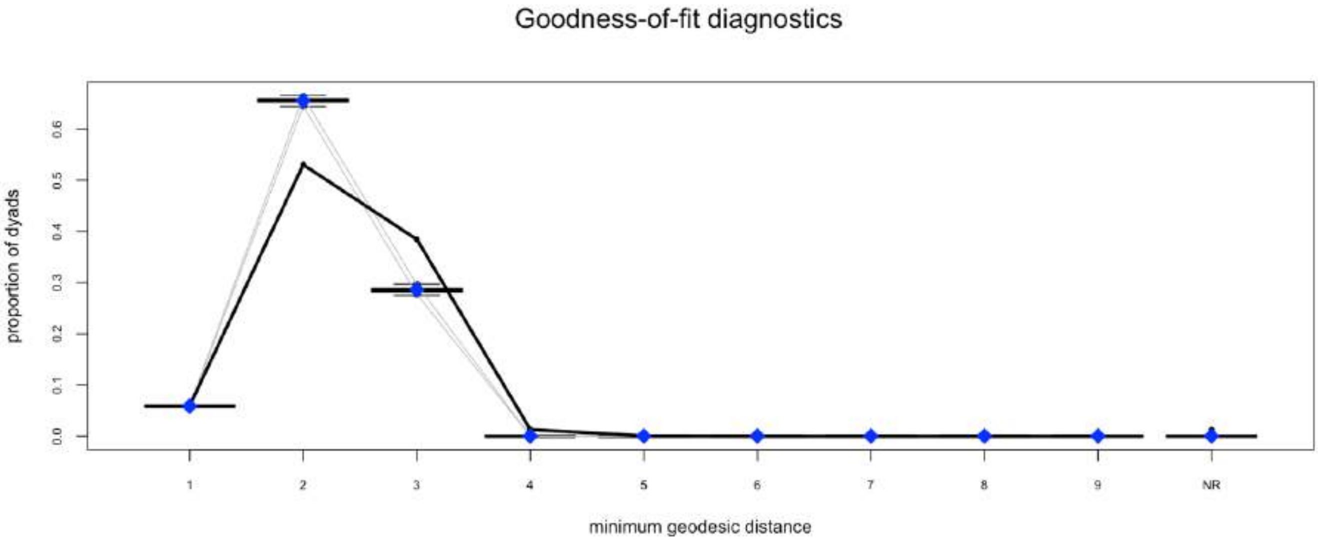
b.



c.



d.



e.

Appendix 2. Goodness-of-Fit Diagnostics for Net_all. *The diagnostic compares the structural measures of the observed network with those from networks simulated using the fitted parameters. Box plots display the overlay between the observed and simulated network statistics. Although not all ERGM terms are supported, the default for directed graphs includes (a) model statistics, (b) out-degree, (c) in-degree, (d) edgewise shared partner, and (e) minimum geodesic distance. The plots indicate a poor fit.*

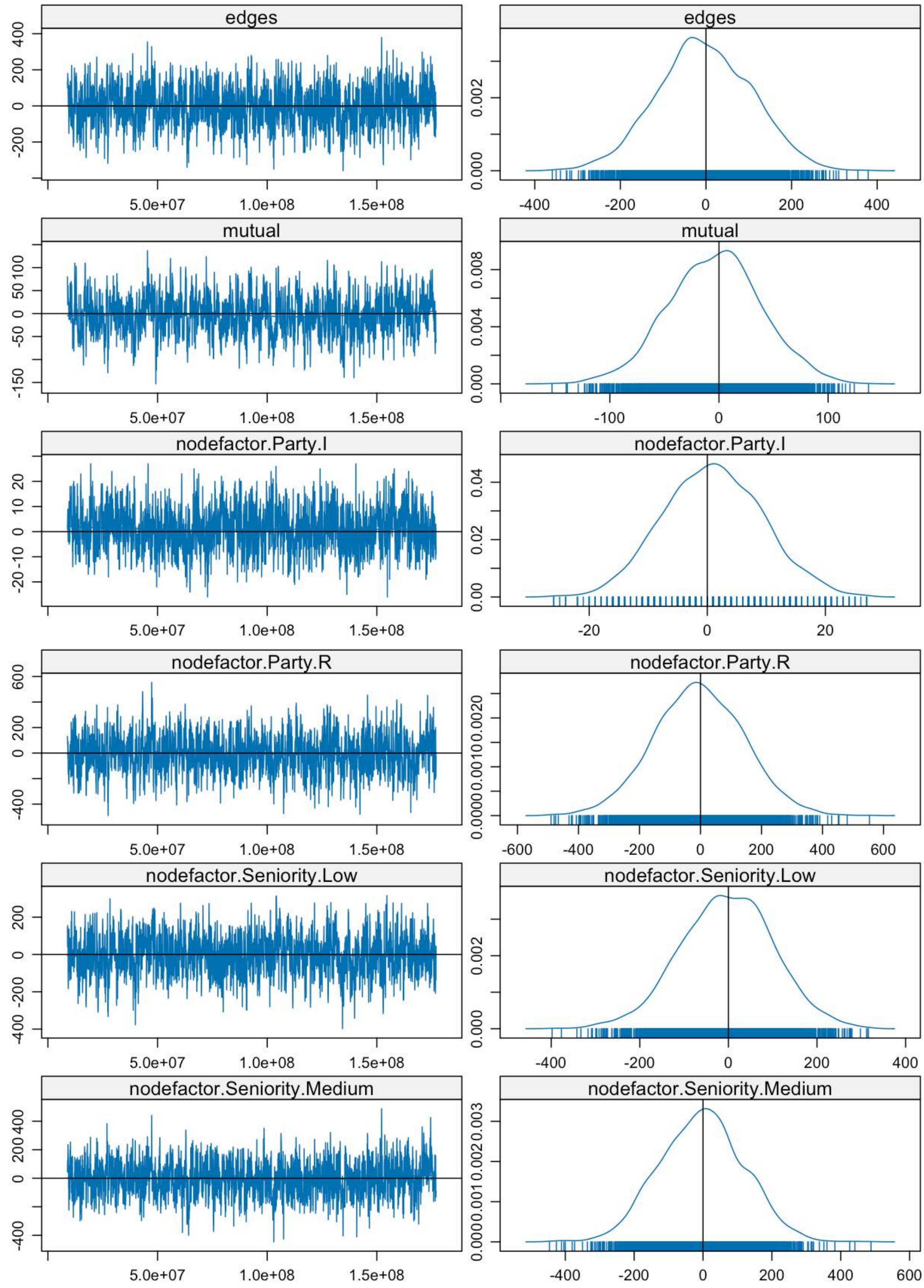
7.2. MCMC Diagnostic and Goodness-of-Fit for ERGM on Net_moderate

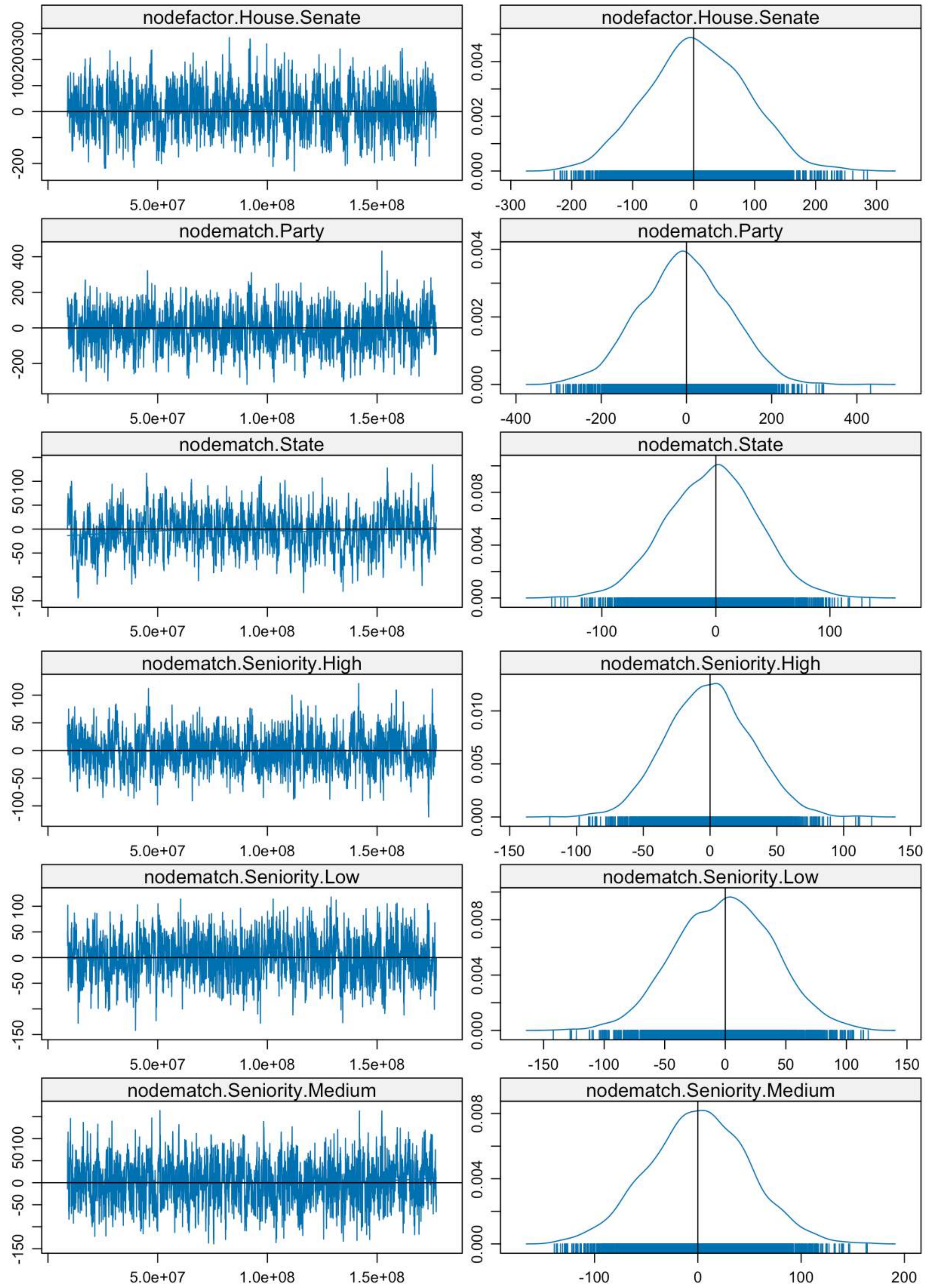
```
Sample statistics auto-correlation:
Chain 1
              edges    mutual nodefactor.Party.I nodefactor.Party.R nodefactor.Seniority.Low nodefactor.Seniority.Medium
Lag 0      1.00000000 1.0000000         1.00000000         1.00000000               1.00000000                  1.00000000
Lag 65536  0.40546784 0.6309974         0.38238426         0.43850244               0.43852387                  0.38605712
Lag 131072 0.23833843 0.4474315         0.24025497         0.26692784               0.27330171                  0.23392757
Lag 196608 0.15954862 0.3169530         0.13484872         0.17093018               0.16857855                  0.16605113
Lag 262144 0.11125712 0.2358642         0.11410280         0.09740758               0.10426683                  0.12447412
Lag 327680 0.07840075 0.1668166         0.07581421         0.07089547               0.06867204                  0.09475581
           nodefactor.House.Senate nodematch.Party nodematch.State nodematch.Seniority.High nodematch.Seniority.Low
Lag 0                    1.0000000       1.0000000       1.0000000                1.0000000               1.0000000
Lag 65536                0.5223648       0.4641102       0.6954157                0.5219793               0.4856499
Lag 131072               0.3663173       0.2992335       0.5580062                0.3293818               0.3428395
Lag 196608               0.2816990       0.2071849       0.4548358                0.2447801               0.2336368
Lag 262144               0.2069323       0.1538063       0.3667352                0.1924133               0.1601870
Lag 327680               0.1742727       0.1075878       0.3158326                0.1595305               0.1154454
           nodematch.Seniority.Medium nodematch.House
Lag 0                      1.00000000       1.0000000
Lag 65536                  0.41312758       0.4426989
Lag 131072                 0.25445021       0.2691934
Lag 196608                 0.18552628       0.1755462
Lag 262144                 0.11737340       0.1231376
Lag 327680                 0.08856349       0.0823921
```

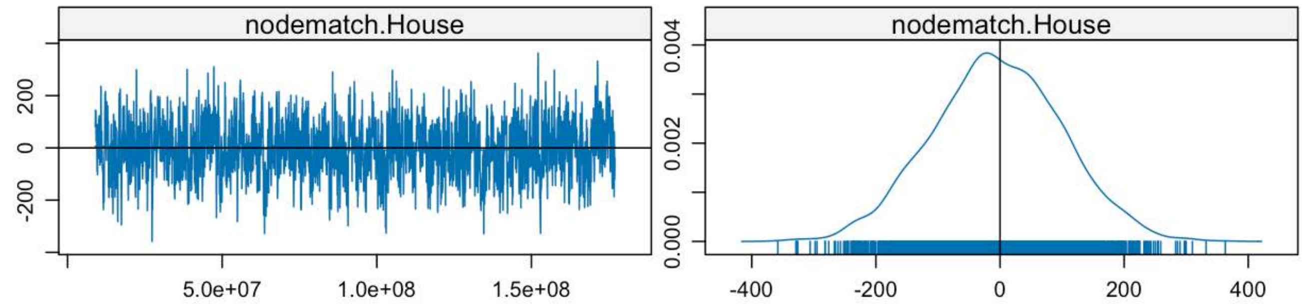
a. Sample Statistic Auto-correlation for Net_moderate. *This measure checks the correlation between sample statistics at different points in the MCMC chain. A well-mixed chain should have low autocorrelation (close to 0) after Lag 0. The results show a decreasing value to 0, indicating low autocorrelation, which is a positive outcome.*

```
Sample statistics burn-in diagnostic (Geweke):
Chain 1

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

                    edges                     mutual          nodefactor.Party.I          nodefactor.Party.R
               0.07614513                 0.13419333                  1.21453482                 -0.08899313
 nodefactor.Seniority.Low nodefactor.Seniority.Medium     nodefactor.House.Senate             nodematch.Party
              -0.27360699                -0.08693215                  0.19905574                 -0.13072621
          nodematch.State    nodematch.Seniority.High     nodematch.Seniority.Low  nodematch.Seniority.Medium
              -1.52883434                 0.43985378                 -0.12077272                  0.32929480
          nodematch.House
               0.70497216

Individual P-values (lower = worse):
                    edges                     mutual          nodefactor.Party.I          nodefactor.Party.R
                0.9393036                  0.8932497                   0.2245436                   0.9290874
 nodefactor.Seniority.Low nodefactor.Seniority.Medium     nodefactor.House.Senate             nodematch.Party
                0.7843867                  0.9307254                   0.8422191                   0.8959919
          nodematch.State    nodematch.Seniority.High     nodematch.Seniority.Low  nodematch.Seniority.Medium
                0.1263055                  0.6600430                   0.9038711                   0.7419329
          nodematch.House
                0.4808276
Joint P-value (lower = worse):  0.5815412

Note: MCMC diagnostics shown here are from the last round of simulation, prior to computation of final parameter
  estimates. Because the final estimates are refinements of those used for this simulation run, these diagnostics
  may understate model performance. To directly assess the performance of the final model on in-model statistics,
  please use the GOF command: gof(ergmFitObject, GOF=~model).
```

b. Sample Statistic Burn-in Diagnostic (Geweke) for Net_moderate. *This diagnostic shows the p-values of sample statistics, with higher values indicating better MCMC convergence. Most individual p-values are high, and the joint p-value of 0.58 suggests satisfactory convergence.*

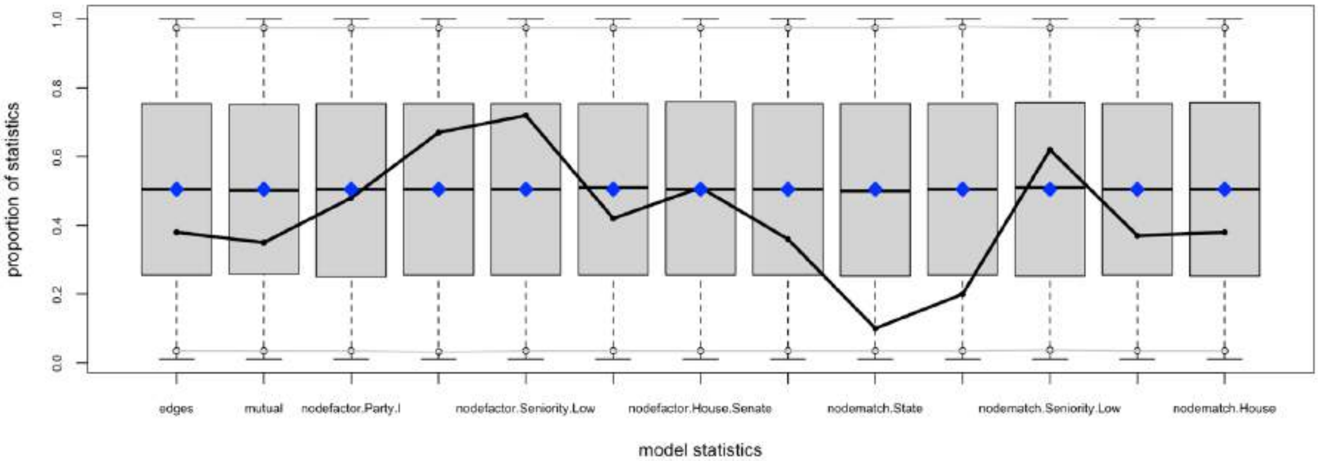

edges
mutual
nodefactor.Party.I
nodefactor.Party.R
nodefactor.Seniority.Low
nodefactor.Seniority.Medium
5.0e+07
1.0e+08
1.5e+08


nodefactor.House.Senate
nodefactor.House.Senate
nodematch.Party
nodematch.Party
nodematch.State
nodematch.State
nodematch.Seniority.High
nodematch.Seniority.High
nodematch.Seniority.Low
nodematch.Seniority.Low
nodematch.Seniority.Medium
nodematch.Seniority.Medium
5.0e+07
1.0e+08
1.5e+08

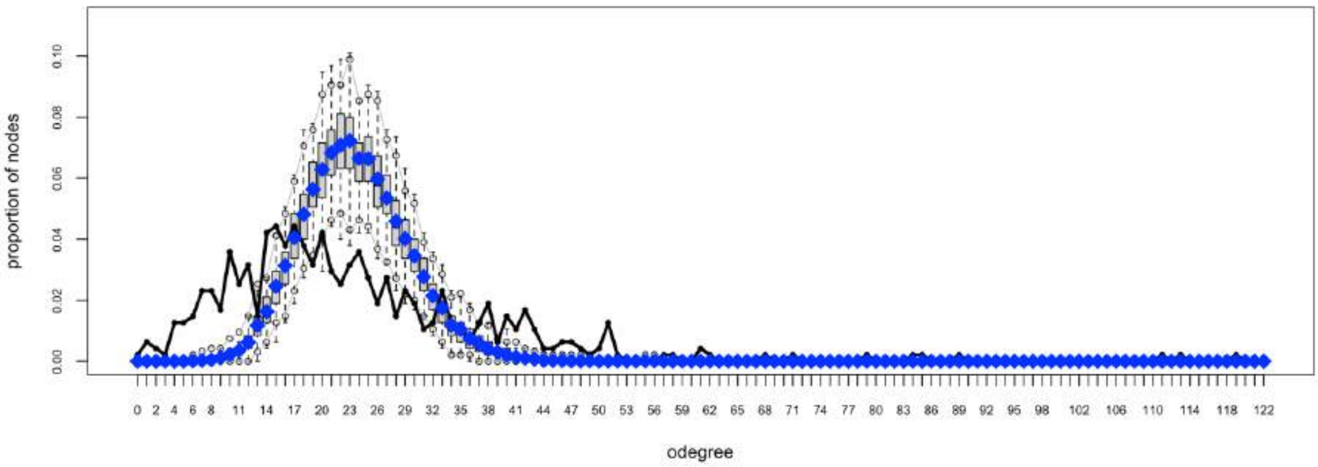


c. MCMC Trace and Density plots for Net_moderate. *MCMC trace plots demonstrate the sample statistics are well-mixed around 0, indicating a good match between the simulations and the observed network at each step. MCMC density plots, displaying a normal distribution centred around 0, suggest that the simulations closely align with the observed network.*

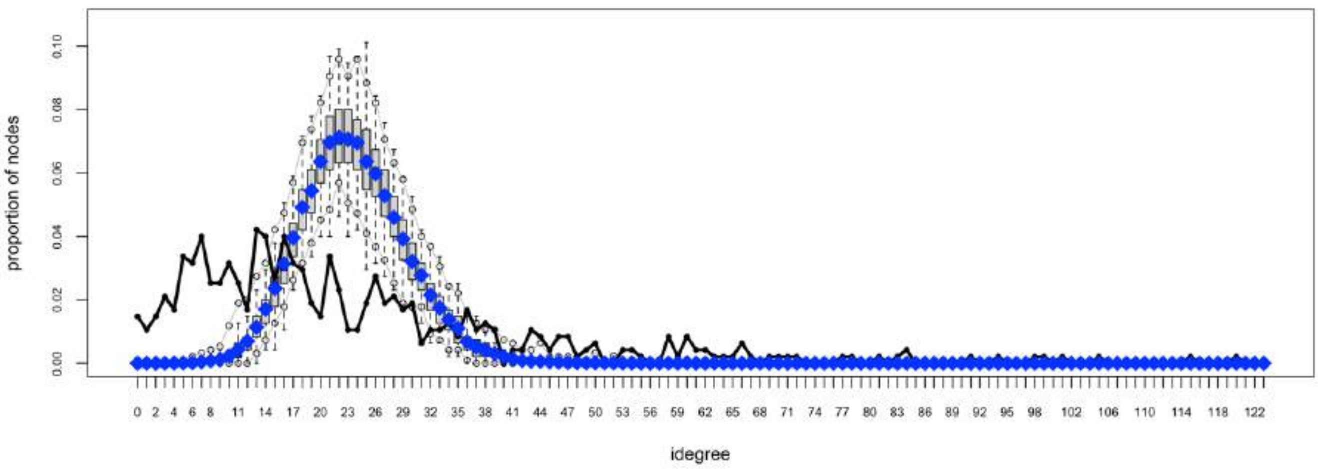
Appendix 3. Markov Chain Monte Carlo (MCMC) Diagnostics for Net_moderate.



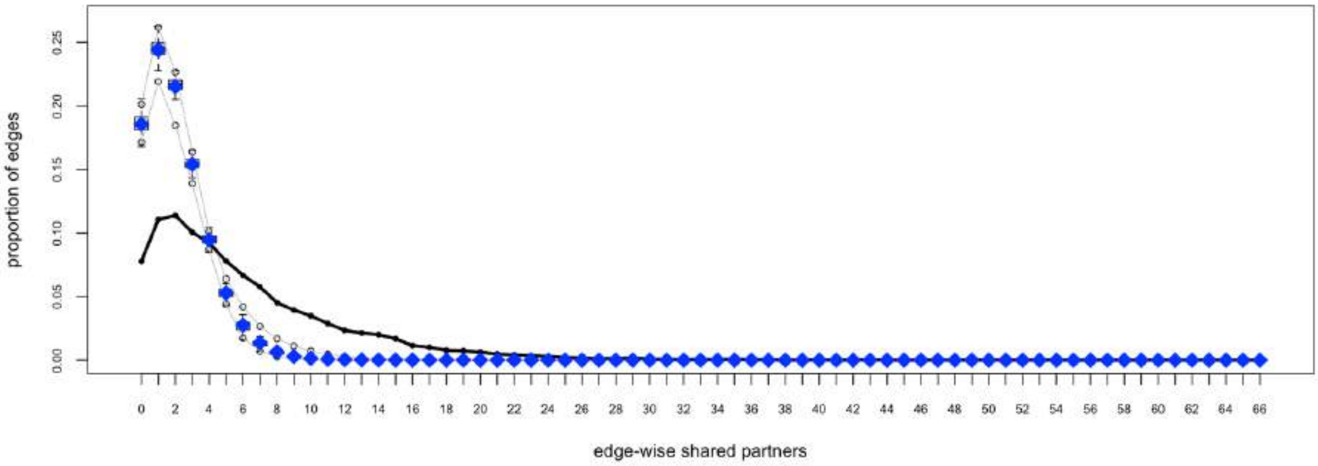
a.



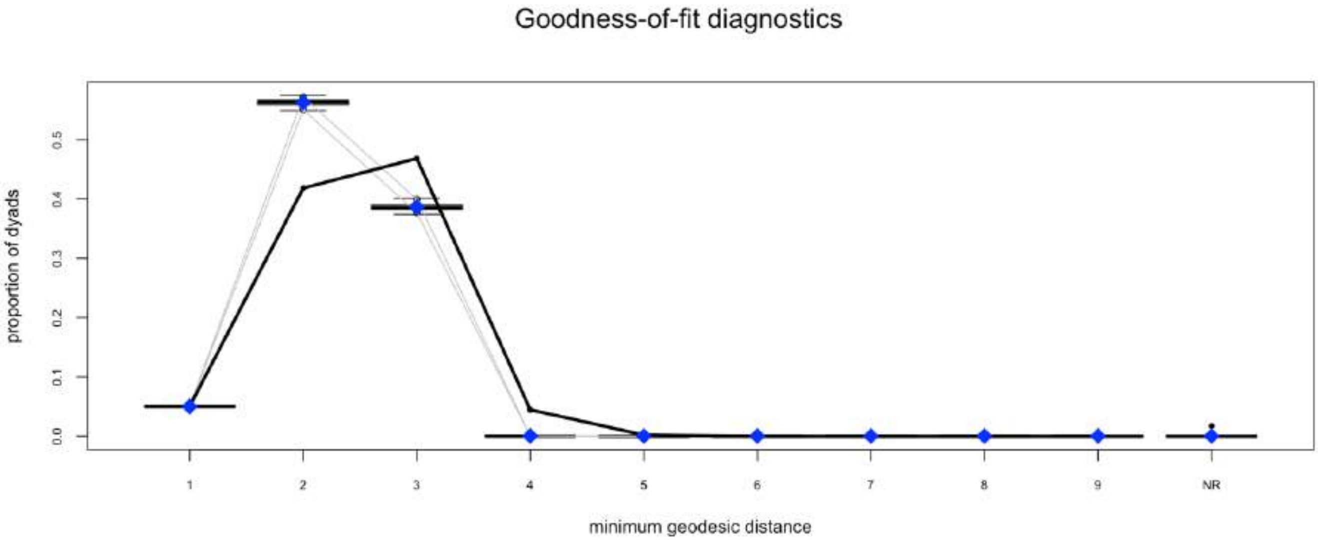
b.



c.



d.



e.

Appendix 4. Goodness-of-Fit Diagnostics for Net_moderate. *The diagnostic compares the structural measures of the observed network with those from networks simulated using the fitted parameters. Box plots display the overlay between the observed and simulated network statistics. Although not all ERGM terms are supported, the default for directed graphs includes (a) model statistics, (b) out-degree, (c) in-degree, (d) edgewise shared partner, and (e) minimum geodesic distance. While most model statistics fit well, with observed values falling within the simulated range, except two statistics. Other plots still show poor overlap, indicating a poor fit overall.*

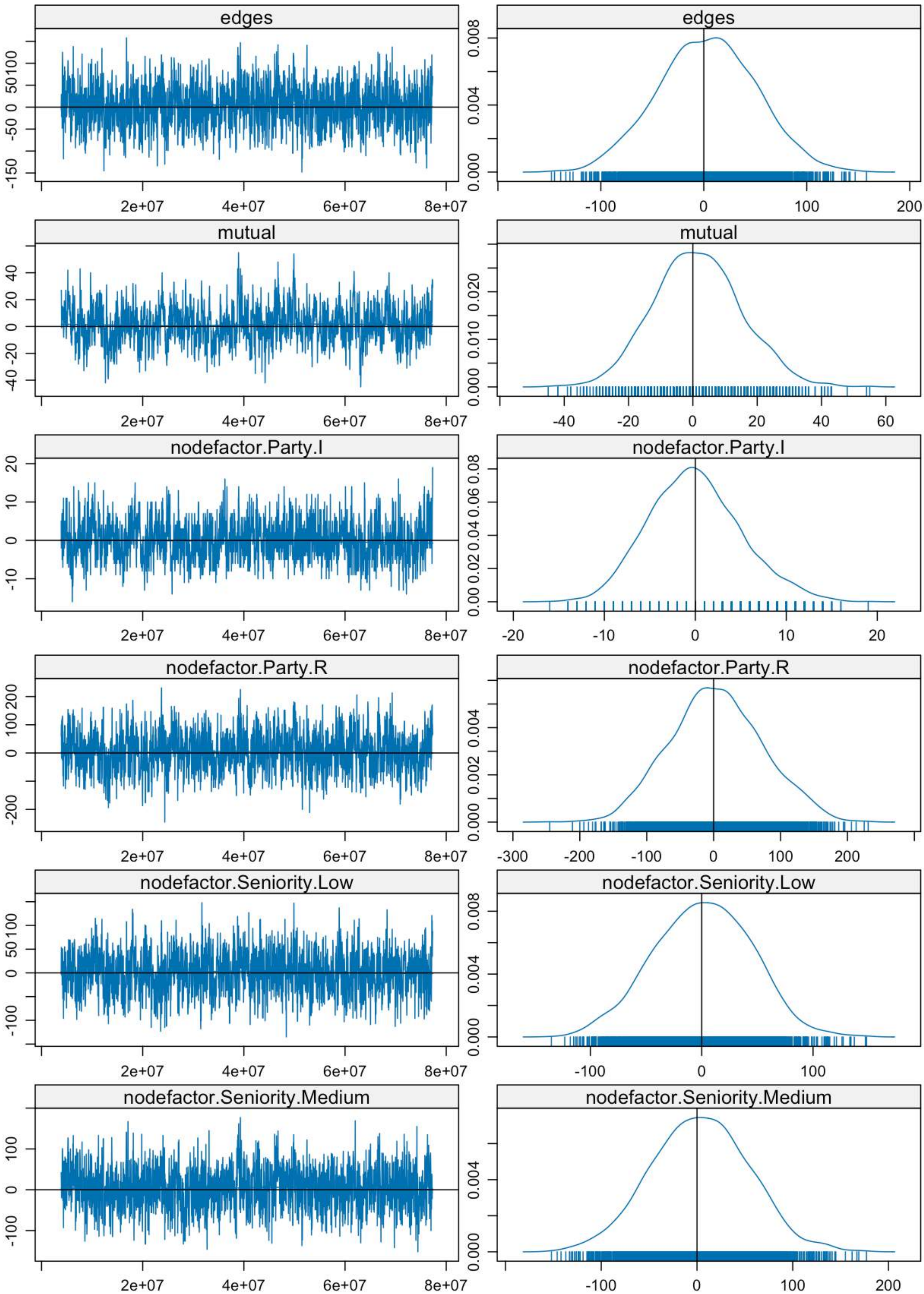
7.3. MCMC Diagnostic and Goodness-of-Fit for ERGM on Net_active

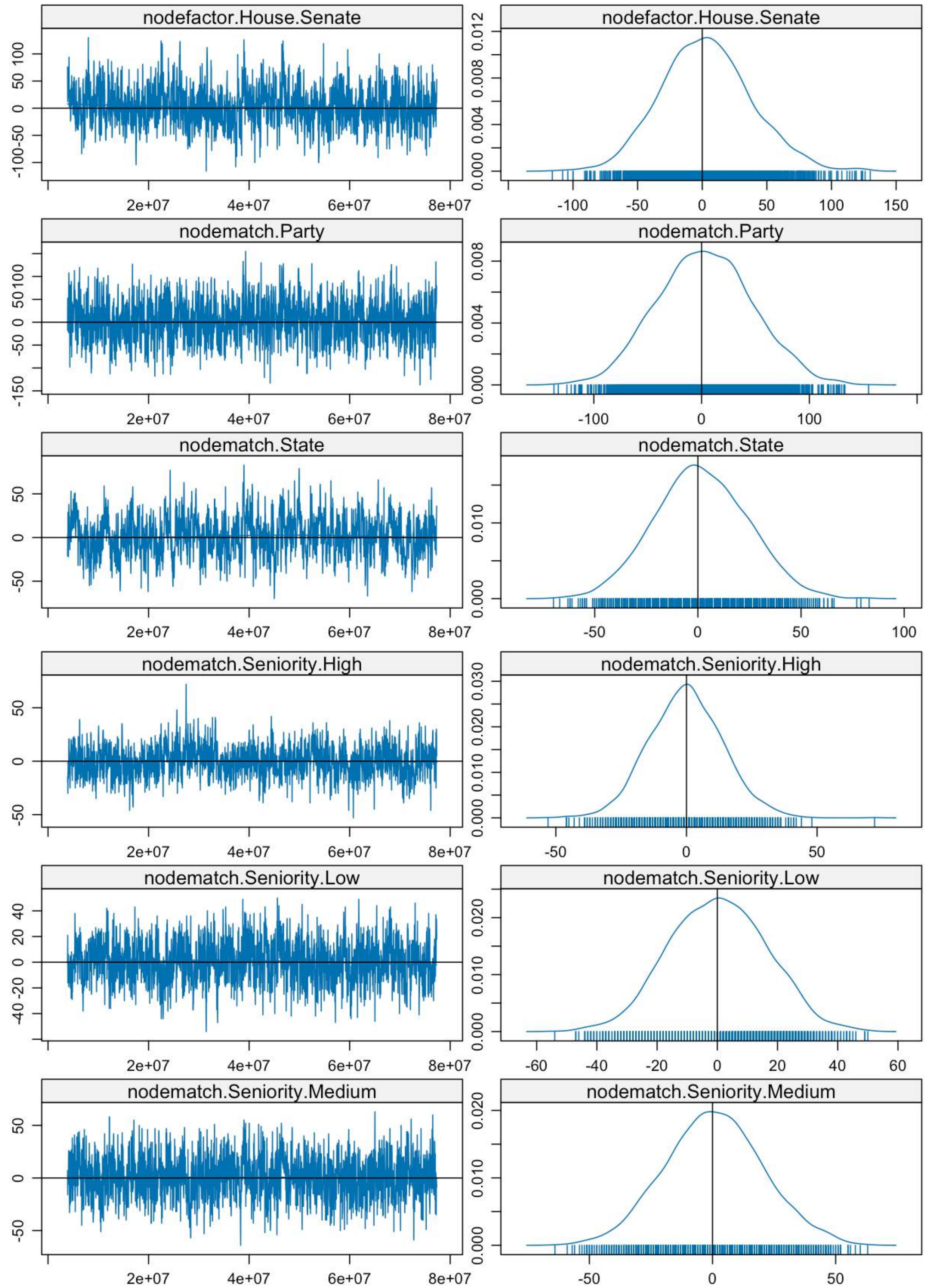
```
Sample statistics auto-correlation:
Chain 1
                edges    mutual nodefactor.Party.I nodefactor.Party.R nodefactor.Seniority.Low nodefactor.Seniority.Medium
Lag 0      1.00000000 1.0000000          1.0000000          1.0000000               1.00000000                  1.00000000
Lag 32768  0.26302621 0.6039551          0.3255110          0.3144524               0.26654392                  0.23752385
Lag 65536  0.15641749 0.4147147          0.2334588          0.2249970               0.18522207                  0.12146663
Lag 98304  0.11146602 0.3043337          0.2018687          0.1495727               0.13568215                  0.07352188
Lag 131072 0.07862908 0.2406412          0.1467121          0.1379134               0.11039439                  0.04605752
Lag 163840 0.06035971 0.2089837          0.1327045          0.0991906               0.06867953                  0.03967067
           nodefactor.House.Senate nodematch.Party nodematch.State nodematch.Seniority.High nodematch.Seniority.Low
Lag 0                    1.0000000      1.00000000       1.0000000                1.0000000               1.0000000
Lag 32768                0.3599049      0.28124932       0.6024134                0.3307719               0.2934404
Lag 65536                0.2467579      0.17574198       0.4595509                0.2434373               0.2079861
Lag 98304                0.1946727      0.10607376       0.3507947                0.2046585               0.1702983
Lag 131072               0.1416201      0.07911755       0.2709149                0.1595306               0.1286746
Lag 163840               0.1005315      0.06643895       0.2321577                0.1385990               0.0981741
           nodematch.Seniority.Medium nodematch.House
Lag 0                      1.00000000      1.00000000
Lag 32768                  0.24514297      0.28423653
Lag 65536                  0.14580427      0.17110912
Lag 98304                  0.09633053      0.11713603
Lag 131072                 0.09548340      0.09992937
Lag 163840                 0.04575000      0.08105854
```

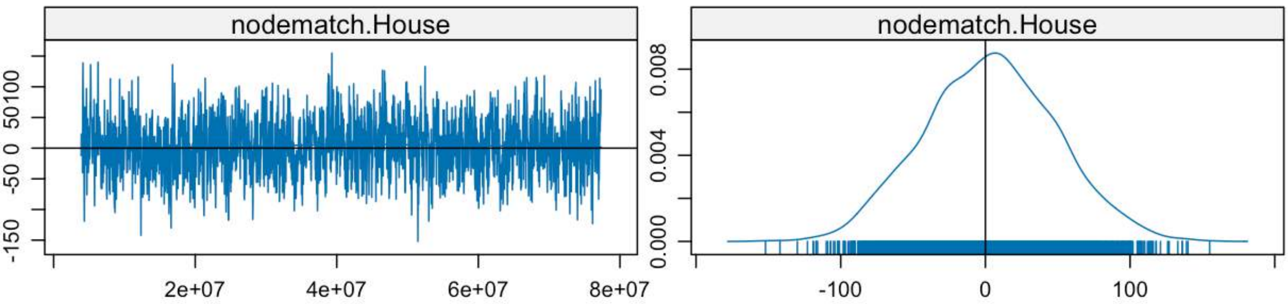
a. Sample Statistic Auto-correlation for Net_active. *This measure checks the correlation between sample statistics at different points in the MCMC chain. A well-mixed chain should have low autocorrelation (close to 0) after Lag 0. The results show a decreasing value to 0, indicating low autocorrelation, which is a positive outcome.*

```
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

                   edges                     mutual          nodefactor.Party.I          nodefactor.Party.R
              0.81278564                 0.43851823                  0.67169965                  1.36268071
nodefactor.Seniority.Low nodefactor.Seniority.Medium     nodefactor.House.Senate             nodematch.Party
              1.42137305                 0.90219378                  1.97083797                  0.67175438
         nodematch.State   nodematch.Seniority.High     nodematch.Seniority.Low  nodematch.Seniority.Medium
             -0.05773528                -0.09521032                  0.45407195                  0.35517699
         nodematch.House
              0.41914618

Individual P-values (lower = worse):
                   edges                     mutual          nodefactor.Party.I          nodefactor.Party.R
              0.41634097                 0.66101066                  0.50177492                  0.17298316
nodefactor.Seniority.Low nodefactor.Seniority.Medium     nodefactor.House.Senate             nodematch.Party
              0.15520834                 0.36695394                  0.04874241                  0.50174008
         nodematch.State   nodematch.Seniority.High     nodematch.Seniority.Low  nodematch.Seniority.Medium
              0.95395949                 0.92414777                  0.64977704                  0.72245700
         nodematch.House
              0.67510931
Joint P-value (lower = worse):  0.3365909

Note: To save space, only one in every 2 iterations of the MCMC sample used for estimation was stored for
  diagnostics. Sample size per chain was originally around 4484 with thinning interval 16384.

Note: MCMC diagnostics shown here are from the last round of simulation, prior to computation of final parameter
  estimates. Because the final estimates are refinements of those used for this simulation run, these diagnostics
  may understate model performance. To directly assess the performance of the final model on in-model statistics,
  please use the GOF command: gof(ergmFitObject, GOF=~model).
```

b. Sample Statistic Burn-in Diagnostic (Geweke) for Net_active. *This diagnostic shows the p-values of sample statistics, with higher values indicating better MCMC convergence. Most individual p-values are high, and the joint p-value of 0.34 suggests satisfactory convergence.*

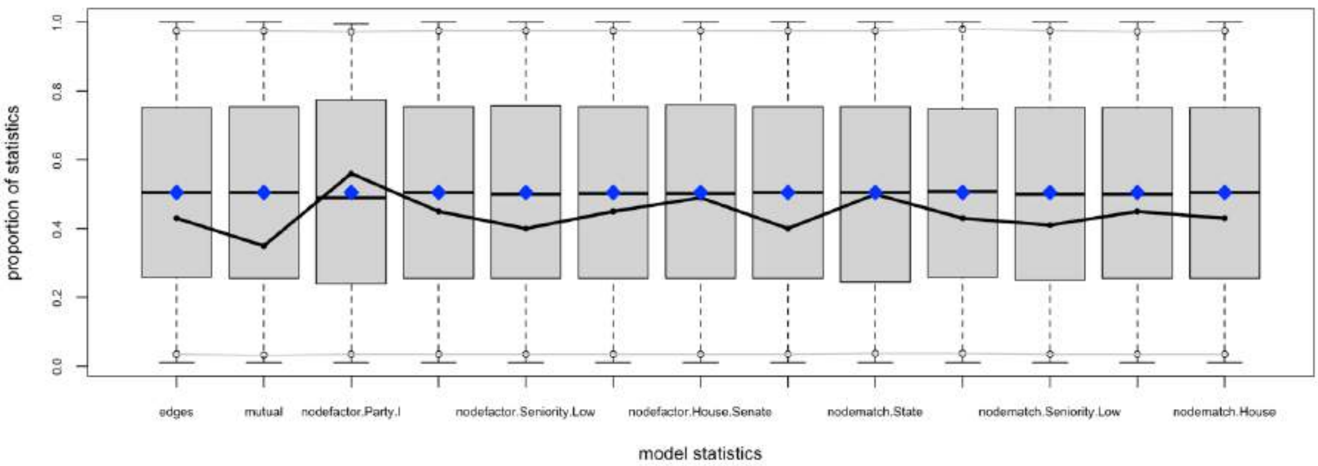

edges
-150
-50 0
50100
2e+07
4e+07
6e+07
8e+07
0.000
0.004
0.008
-100
0
100
200
mutual
-40 -20
0
20
40
0.000 0.010 0.020
-40
-20
0
20
40
60
nodefactor.Party.I
-10
0
10
20
0.00 0.02 0.04 0.06 0.08
-20
-10
0
10
20
nodefactor.Party.R
-200
0
100 200
0.000 0.002 0.004
-300
-200
-100
0
100
200
nodefactor.Seniority.Low
-100
0
50 100
0.000
0.004
0.008
-100
0
100
nodefactor.Seniority.Medium
-100
0
100
0.000
0.004
-100
0
100
200


nodefactor.House.Senate
-100 -50 0 50 100
2e+07 4e+07 6e+07 8e+07
nodefactor.House.Senate
0.000 0.004 0.008 0.012
-100 -50 0 50 100 150
nodematch.Party
-150 -50 0 50100
2e+07 4e+07 6e+07 8e+07
nodematch.Party
0.000 0.004 0.008
-100 0 100
nodematch.State
-50 0 50
2e+07 4e+07 6e+07 8e+07
nodematch.State
0.000 0.010
-50 0 50 100
nodematch.Seniority.High
-50 0 50
2e+07 4e+07 6e+07 8e+07
nodematch.Seniority.High
0.000 0.010 0.020 0.030
-50 0 50
nodematch.Seniority.Low
-40 -20 0 20 40
2e+07 4e+07 6e+07 8e+07
nodematch.Seniority.Low
0.000 0.010 0.020
-60 -40 -20 0 20 40 60
nodematch.Seniority.Medium
-50 0 50
2e+07 4e+07 6e+07 8e+07
nodematch.Seniority.Medium
0.000 0.010 0.020
-50 0 50

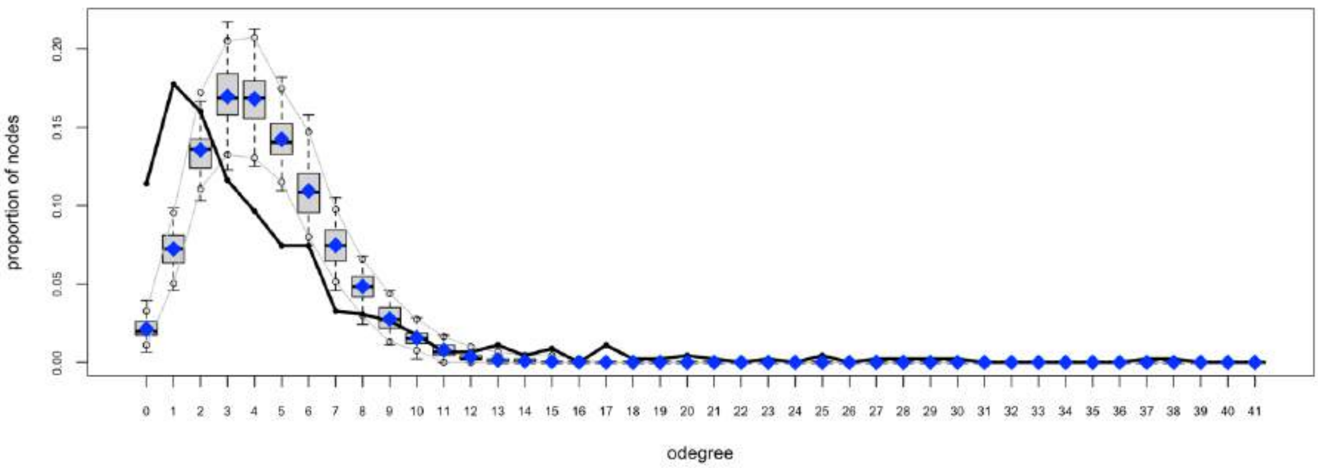


c. MCMC Trace and Density plots for Net_active. *MCMC trace plots demonstrate the sample statistics are well-mixed around 0, indicating a good match between the simulations and the observed network at each step. MCMC density plots, displaying a normal distribution centred around 0, suggest that the simulations closely align with the observed network.*

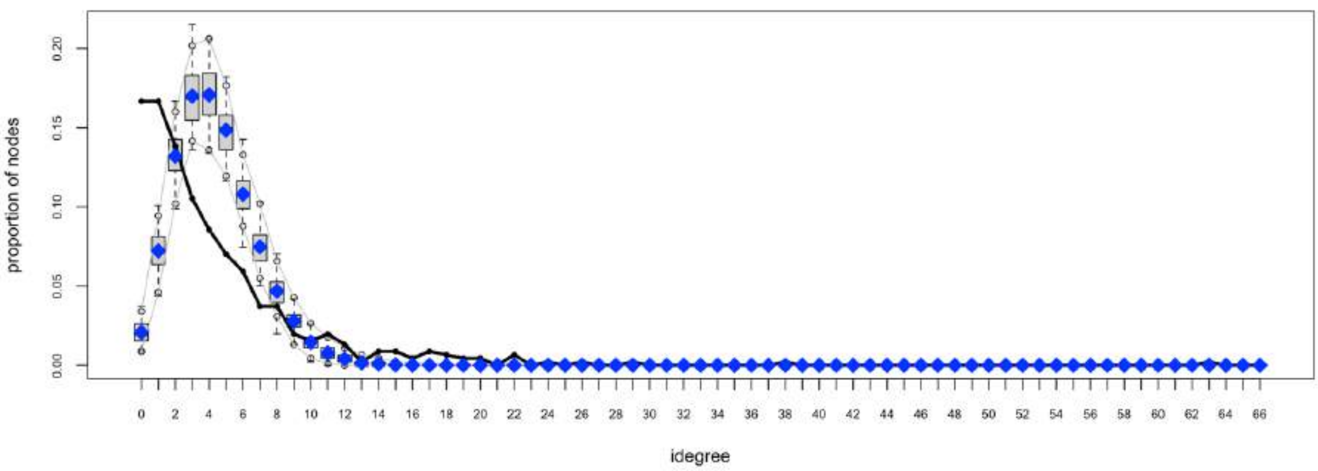
Appendix 5. Markov Chain Monte Carlo (MCMC) Diagnostics for Net_active



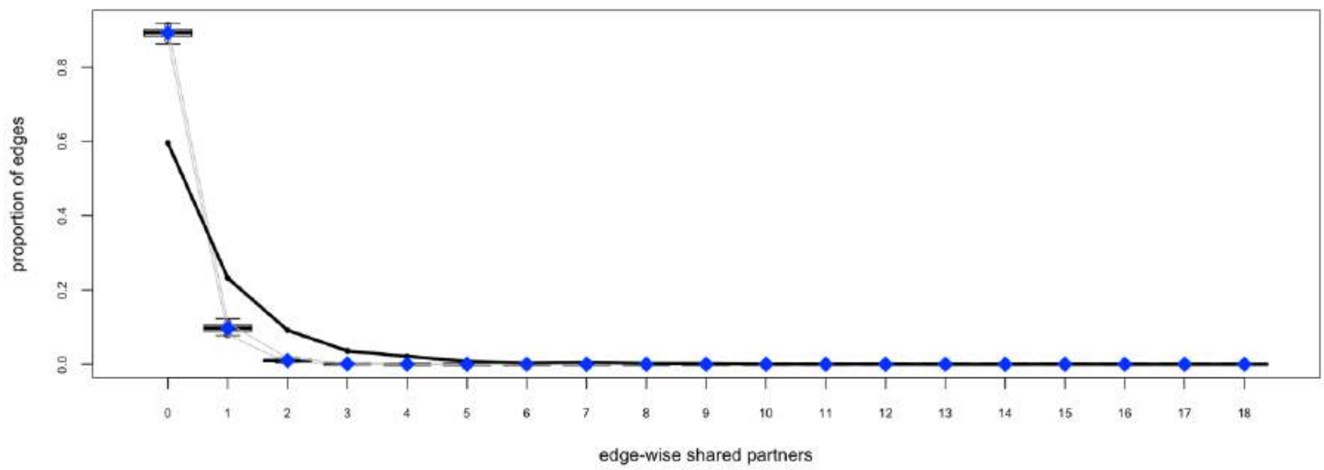
a.



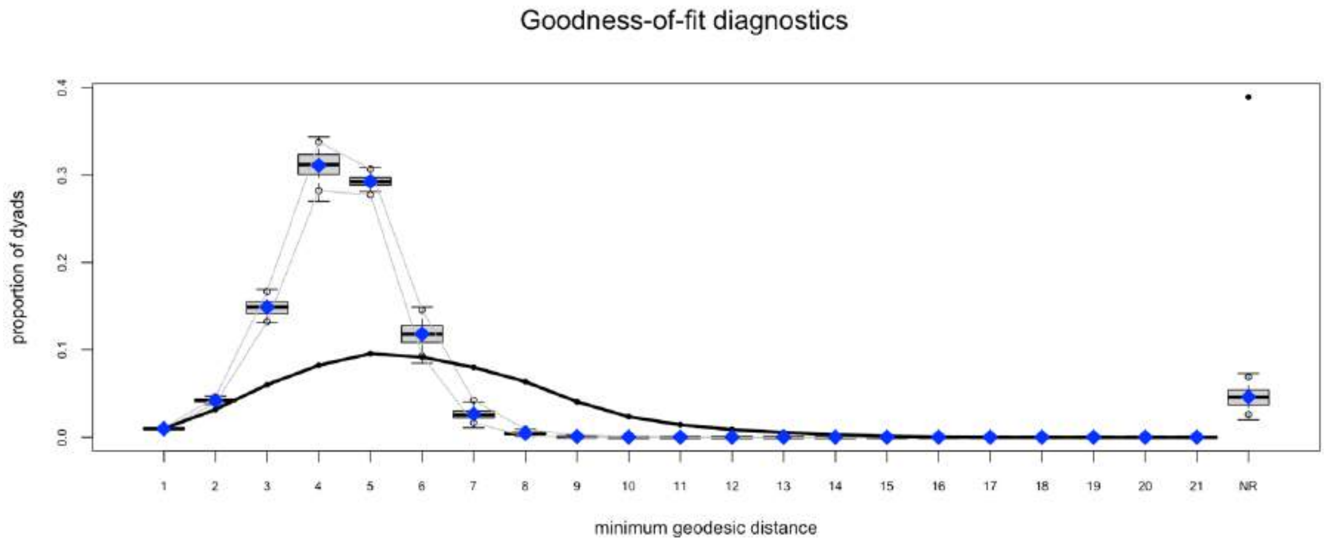
b.



c.



d.

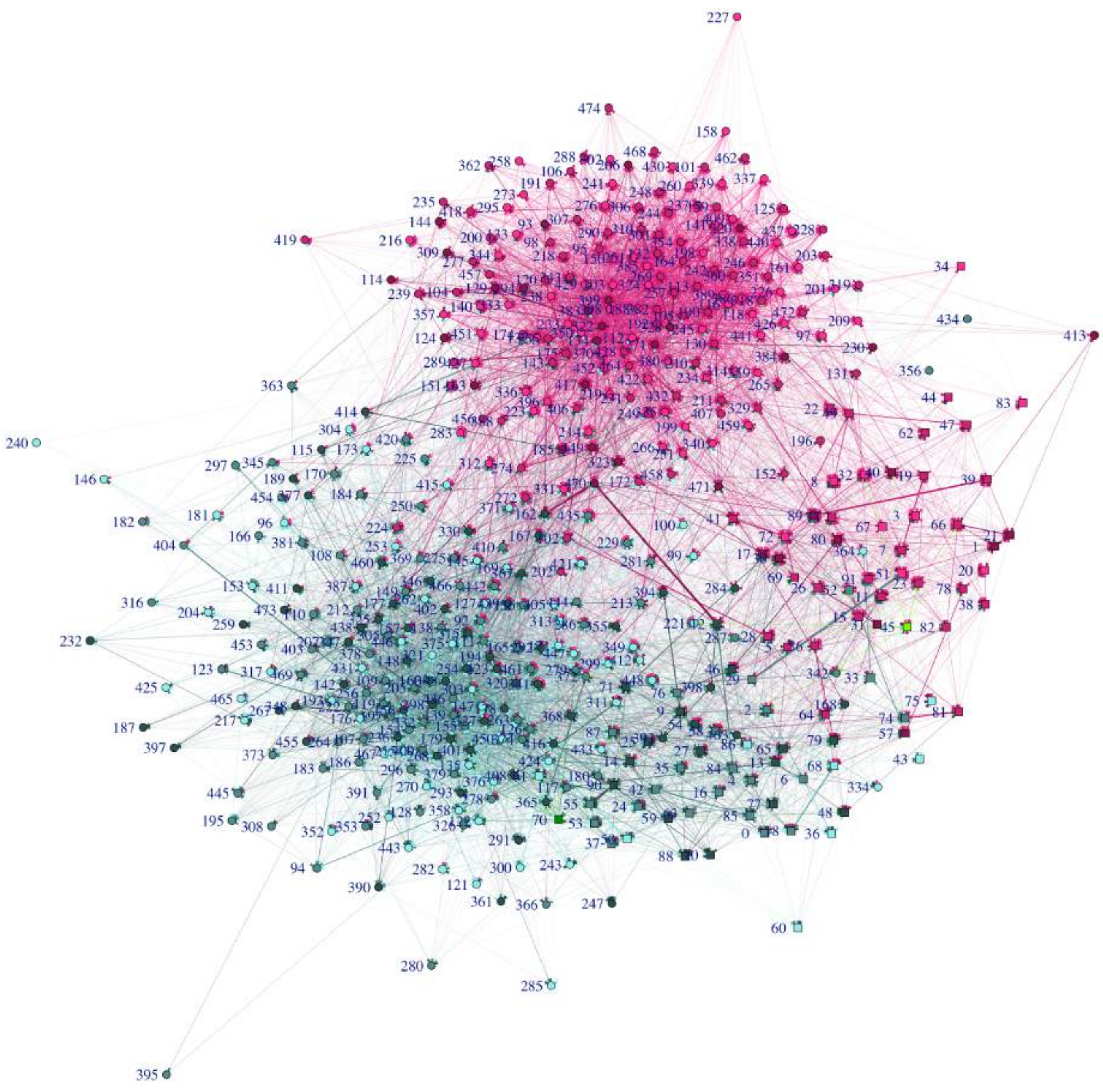


e.

Appendix 6. Goodness-of-Fit Diagnostics for Net_active. *The diagnostic compares the structural measures of the observed network with those from networks simulated using the fitted parameters. Box plots display the overlay between the observed and simulated network statistics. Although not all ERGM terms are supported, the default for directed graphs includes (a) model statistics, (b) out-degree, (c) in-degree, (d) edgewise shared partner, and (e) minimum geodesic distance. All model statistics fit well, with observed values falling within the simulated range. Other plots show better overlap, indicating a better fit for strong connections.*

7.4. Dataset and Code

https://github.com/DwayneHuang/Social-Network-Analysis.git



0	Tammy Baldwin	119	Suzanne Bonamici	238	Michael Guest	357	Jay Obernolte
1	John Barrasso	120	Mike Bost	239	S. Brett Guthrie	358	Ilhan Omar
2	Michael F. Bennet	121	Carolyn Bourdeaux	240	Josh Harder	359	Burgess Owens
3	Marsha Blackburn	122	Jamaal Bowman	241	Diana Harshbarger	360	Steven Palazzo
4	Richard Blumenthal	123	Brendan Boyle	242	Vicky Hartzler	361	Frank Pallone Jr.
5	Roy Blunt	124	Kevin Brady	243	Jahana Hayes	362	Gary Palmer
6	Cory A. Booker	125	Mo Brooks	244	Kevin Hern	363	Jimmy Panetta
7	John Boozman	126	Anthony Brown	245	Yvette Herrell	364	Chris Pappas
8	Mike Braun	127	Shontel Brown	246	Jody Hice	365	Bill Pascrell Jr.
9	Sherrod Brown	128	Julia Brownley	247	Brian Higgins	366	Donald Payne Jr.
10	Maria Cantwell	129	Vern Buchanan	248	Clay Higgins	367	Nancy Pelosi
11	Shelley Moore Capito	130	Ken Buck	249	French Hill	368	Ed Perlmutter
12	Benjamin L. Cardin	131	Larry Bucshon	250	Jim Himes	369	Scott Peters
13	Thomas R. Carper	132	Ted Budd	251	Ashley Hinson	370	August Pfluger
14	Robert P. Jr. Casey	133	Tim Burchett	252	Steven Horsford	371	Dean Phillips
15	Bill Cassidy	134	Michael Burgess	253	Chrissy Houlahan	372	Chellie Pingree
16	Christopher Coons	135	Cori Bush	254	Steny H. Hoyer	373	Stacey Plaskett
17	John Cornyn	136	Cheri Bustos	255	Richard Hudson	374	Mark Pocan
18	Catherine Cortez Masto	137	Kat Cammack	256	Jared Huffman	375	Katie Porter
19	Tom Cotton	138	Salud Carbajal	257	Bill Huizenga	376	Ayanna Pressley
20	Kevin Cramer	139	Tony Cardenas	258	Darrell Issa	377	David Price
21	Mike Crapo	140	Mike Carey	259	Sheila Jackson Lee	378	Mike Quigley
22	Ted Cruz	141	Jerry L. Carl	260	Ronny Jackson	379	Jamie Raskin
23	Steve Daines	142	Andre Carson	261	Chris Jacobs	380	Guy Reschenthaler
24	Tammy Duckworth	143	Buddy Carter	262	Sara Jacobs	381	Kathleen Rice
25	Richard J. Durbin	144	John Carter	263	Pramila Jayapal	382	Tom Rice
26	Joni Ernst	145	Troy Carter	264	Hakeem Jeffries	383	Cathy McMorris Rodgers
27	Dianne Feinstein	146	Ed Case	265	Bill Johnson	384	Mike Rogers
28	Deb Fischer	147	Sean Casten	266	Dusty Johnson	385	John W. Rose
29	Kirsten E. Gillibrand	148	Kathy Castor	267	Eddie Bernice Johnson	386	Matthew M. Rosendale
30	Lindsey Graham	149	Joaquin Castro	268	Henry C. "Hank" Jr. Johnson	387	Deborah K. Ross
31	Chuck Grassley	150	Madison Cawthorn	269	Mike Johnson	388	David Rouzer
32	Bill Hagerty	151	Steve Chabot	270	Mondaire Jones	389	Chip Roy
33	Margaret Wood Hassan	152	Liz Cheney	271	Jim Jordan	390	Lucille Roybal-Allard
34	Josh Hawley	153	Sheila Cherfilus-McCormick	272	David Joyce	391	Raul Ruiz
35	Martin Heinrich	154	Judy Chu	273	John Joyce	392	C. A. Dutch Ruppersberger
36	John W. Hickenlooper	155	Katherine Clark	274	John Katko	393	Bobby L. Rush
37	Mazie K. Hirono	156	Yvette D. Clarke	275	William Keating	394	Tim Ryan
38	John Hoeven	157	Emanuel Cleaver	276	Fred Keller	395	Gregorio Sablan
39	Cindy Hyde-Smith	158	Ben Cline	277	Mike Kelly	396	Maria Elvira Salazar
40	James M. Inhofe	159	Michael Cloud	278	Robin Kelly	397	Linda Sanchez
41	Ron Johnson	160	James E. Clyburn	279	Ro Khanna	398	John P. Sarbanes
42	Tim Kaine	161	Andrew S. Clyde	280	Dan Kildee	399	Steve Scalise
43	Mark Kelly	162	Steve Cohen	281	Derek Kilmer	400	Mary Gay Scanlon
44	John Kennedy	163	Tom Cole	282	Andy Kim	401	Jan Schakowsky
45	Angus S. Jr. King	164	James Comer	283	Young Kim	402	Adam Schiff
46	Amy Klobuchar	165	Gerald E. "Gerry" Connolly	284	Ron Kind	403	Bradley Schneider
47	James Lankford	166	J. Luis Correa	285	Ann Kirkpatrick	404	Kurt Schrader
48	Patrick J. Leahy	167	Jim Costa	286	Raja Krishnamoorthi	405	Kim Schrier
49	Mike Lee	168	Joe Courtney	287	Ann Kuster	406	David Schweikert
50	Ben Ray Luján	169	Angie Craig	288	David Kustoff	407	Austin Scott
51	Cynthia M. Lummis	170	Charlie Crist	289	Darin LaHood	408	Robert C. Scott
52	Joe, III Manchin	171	Jason Crow	290	Doug LaMalfa	409	Pete Sessions
53	Edward J. Markey	172	John R. Curtis	291	Jim Langevin	410	Terri A. Sewell
54	Bob Menendez	173	Sharice Davids	292	Rick Larsen	411	Brad Sherman
55	Jeff Merkley	174	Warren Davidson	293	John B. Larson	412	Mikie Sherrill
56	Jerry Moran	175	Rodney Davis	294	Robert E. Latta	413	Mike Simpson
57	Lisa Murkowski	176	Madeleine Dean	295	Jake LaTurner	414	Albio Sires
58	Christopher Murphy	177	Peter DeFazio	296	Brenda Lawrence	415	Elissa Slotkin
59	Patty Murray	178	Diana DeGette	297	Al Lawson	416	Adam Smith
60	Jon Ossoff	179	Rosa L. DeLauro	298	Barbara Lee	417	Adrian Smith
61	Alex Padilla	180	Suzan DelBene	299	Susie Lee	418	Jason Smith
62	Rand Paul	181	Antonio Delgado	300	Teresa Leger Fernandez	419	Lloyd Smucker
63	Gary C. Peters	182	Val Demings	301	Debbie Lesko	420	Darren Soto
64	Rob Portman	183	Mark DeSaulnier	302	Julia Letlow	421	Abigail Spanberger
65	Jack Reed	184	Ted Deutch	303	Andy Levin	422	Victoria Spartz
66	James E. Risch	185	Mario Diaz-Balart	304	Mike Levin	423	Jackie Speier
67	Mitt Romney	186	Debbie Dingell	305	Ted Lieu	424	Melanie Stansbury
68	Jacky Rosen	187	Lloyd Doggett	306	Billy Long	425	Greg Stanton
69	Marco Rubio	188	Byron Donalds	307	Barry Loudermilk	426	Pete Stauber
70	Bernard Sanders	189	Michael Doyle	308	Alan Lowenthal	427	Michelle Steel
71	Charles E. Schumer	190	Jeff Duncan	309	Frank Lucas	428	Elise Stefanik
72	Rick Scott	191	Neal Dunn	310	Blaine Luetkemeyer	429	Bryan Steil
73	Tim Scott	192	Tom Emmer	311	Elaine Luria	430	W. Gregory Steube
74	Jeanne Shaheen	193	Veronica Escobar	312	Nancy Mace	431	Haley Stevens
75	Kyrsten Sinema	194	Anna A. Eshoo	313	Tom Malinowski	432	Chris Stewart
76	Tina Smith	195	Adriano Espaillat	314	Nicole Malliotakis	433	Marilyn Strickland
77	Debbie Stabenow	196	Ron Estes	315	Carolyn Maloney	434	Thomas Suozzi
78	Dan Sullivan	197	Dwight Evans	316	Sean Patrick Maloney	435	Eric Swalwell
79	Jon Tester	198	Pat Fallon	317	Kathy E. Manning	436	Mark Takano
80	John Thune	199	Randy Feenstra	318	Thomas Massie	437	Claudia Tenney
81	Thom Tillis	200	A. Drew Ferguson	319	Brian Mast	438	Bennie G. Thompson
82	Patrick J. Toomey	201	Michelle Fischbach	320	Doris O. Matsui	439	Mike Thompson
83	Tommy Tuberville	202	Brian Fitzpatrick	321	Lucy McBath	440	Thomas P. Tiffany
84	Chris Van Hollen	203	Chuck Fleischmann	322	Kevin McCarthy	441	William Timmons
85	Mark R. Warner	204	Lizzie Fletcher	323	Michael T. McCaul	442	Dina Titus
86	Raphael G. Warnock	205	Bill Foster	324	Lisa C. McClain	443	Rashida Tlaib
87	Elizabeth Warren	206	Virginia Foxx	325	Betty McCollum	444	Paul D. Tonko
88	Sheldon Whitehouse	207	Lois Frankel	326	A. Donald McEachin	445	Norma Torres
89	Roger F. Wicker	208	C. Scott Franklin	327	James McGovern	446	Ritchie Torres
90	Ron Wyden	209	Russ Fulcher	328	Patrick T. McHenry	447	Lori Trahan
91	Todd Young	210	Matt Gaetz	329	David McKinley	448	David Trone
92	Alma Adams	211	Mike Gallagher	330	Gregory W. Meeks	449	Mike Turner
93	Robert Aderholt	212	Ruben Gallego	331	Peter Meijer	450	Lauren Underwood
94	Pete Aguilar	213	John Garamendi	332	Grace Meng	451	David G. Valadao
95	Rick Allen	214	Andrew R. Garbarino	333	Daniel Meuser	452	Beth Van Duyne
96	Colin Allred	215	Jesús "Chuy" García	334	Kweisi Mfume	453	Juan Vargas
97	Kelly Armstrong	216	Mike Garcia	335	Mariannette Miller-Meeks	454	Marc Veasey
98	Jodey Arrington	217	Sylvia Garcia	336	Carol Miller	455	Nydia M. Velazquez
99	Jake Auchincloss	218	Bob Gibbs	337	Mary E. Miller	456	Ann Wagner
100	Cindy Axne	219	Carlos A. Gimenez	338	Alex Mooney	457	Tim Walberg
101	Brian Babin	220	Louie Gohmert	339	Barry Moore	458	Jackie Walorski
102	Don Bacon	221	Jared Golden	340	Blake D. Moore	459	Michael Waltz
103	Jim Baird	222	Jimmy Gomez	341	Gwen Moore	460	Debbie Wasserman Schultz
104	Troy Balderson	223	Tony Gonzales	342	Joseph Morelle	461	Bonnie Watson Coleman
105	Jim Banks	224	Jennifer Gonzalez	343	Markwayne Mullin	462	Daniel Webster
106	Andy Barr	225	Vicente Gonzalez	344	Gregory Francis Murphy	463	Peter Welch
107	Nanette Barragán	226	Bob Good	345	Stephanie Murphy	464	Brad Wenstrup
108	Karen Bass	227	Lance Gooden	346	Jerrold Nadler	465	Jennifer Wexton
109	Joyce Beatty	228	Paul A. Gosar	347	Grace Napolitano	466	Susan Wild
110	Ami Bera	229	Josh Gottheimer	348	Richard E. Neal	467	Nikema Williams
111	Don Beyer	230	Kay Granger	349	Joe Neguse	468	Roger Williams
112	Stephanie I. Bice	231	Garret Graves	350	Troy E. Nehls	469	Frederica Wilson
113	Andy Biggs	232	Al Green	351	Dan Newhouse	470	Joe Wilson
114	Gus M. Bilirakis	233	Mark Green	352	Marie Newman	471	Robert J. Wittman
115	Sanford D. Bishop Jr.	234	Marjorie Taylor Greene	353	Donald Norcross	472	Steve Womack
116	Dan Bishop	235	Morgan Griffith	354	Ralph Norman	473	John A. Yarmuth
117	Lisa Blunt Rochester	236	Raul Grijalva	355	Eleanor Holmes Norton	474	Lee Zeldin
118	Lauren Boebert	237	Glenn Grothman	356	Tom O'Halleran		

Appendix 7. Visualization of the Original Weighted Network with Member List

8. Reference

Aric A. Hagberg, Daniel A. Schult, Pieter J. Swart, 2008. Exploring network structure, dynamics, and function using NetworkX [WWW Document]. Pasadena CA USA Pp 11–15. URL https://networkx.org/documentation/stable/index.html

Ausserhofer, J., Maireder, A., 2013. NATIONAL POLITICS ON TWITTER: Structures and topics of a networked public sphere. Inf. Commun. Soc. 16, 291–314. https://doi.org/10.1080/1369118X.2012.756050

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., Fowler, J.H., 2012. A 61-million-person experiment in social influence and political mobilization. Nature 489, 295–298. https://doi.org/10.1038/nature11421

Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G., 2009. Network Analysis in the Social Sciences. Science 323, 892–895. https://doi.org/10.1126/science.1165821

Bossaert, G., Meidert, N., 2013. "We Are Only as Strong as We Are United, as Weak as We Are Divided" a Dynamic Analysis of the Peer Support Networks in the Harry Potter Books. Open J. Appl. Sci. 03, 174–185. https://doi.org/10.4236/ojapps.2013.32024

Buccoliero, L., Bellio, E., Crestini, G., Arkoudas, A., 2020. Twitter and politics: Evidence from the US presidential elections 2016. J. Mark. Commun. 26, 88–114. https://doi.org/10.1080/13527266.2018.1504228

Christakis, N.A., Fowler, J.H., 2008. The collective dynamics of smoking in a large social network. N. Engl. J. Med. 358, 2249–2258. https://doi.org/10.1056/NEJMsa0706154

Conway, B.A., Kenski, K., Wang, D., 2015. The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. J. Comput.-Mediat. Commun. 20, 363–380. https://doi.org/10.1111/jcc4.12124

Desmarais, B.A., Moscardelli, V.G., Schaffner, B.F., Kowal, M.S., 2015. Measuring legislative collaboration: The Senate press events network. Soc. Netw. 40, 43–54. https://doi.org/10.1016/j.socnet.2014.07.006

Enli, G.S., Skogerbø, E., 2013. PERSONALIZED CAMPAIGNS IN PARTY-CENTRED POLITICS: Twitter and Facebook as arenas for political communication. Inf. Commun. Soc. 16, 757–774. https://doi.org/10.1080/1369118X.2013.782330

Fink, C.G., Fullin, K., Gutierrez, G., Omodt, N., Zinnecker, S., Sprint, G., McCulloch, S., 2023a. A centrality measure for quantifying spread on weighted, directed networks. Phys. Stat. Mech. Its Appl. 626, 129083. https://doi.org/10.1016/j.physa.2023.129083

Fink, C.G., Omodt, N., Zinnecker, S., Sprint, G., 2023b. A Congressional Twitter network dataset quantifying pairwise probability of influence. Data Brief 50, 109521. https://doi.org/10.1016/j.dib.2023.109521

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Soc. Netw. 1, 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Ganguly, M., Dey, P., Roy, S., 2024. Influence maximization in community-structured social networks: a centrality-based approach. J. Supercomput. 80, 19898–19941. https://doi.org/10.1007/s11227-024-06217-3

Golbeck, J., Grimes, J.M., Rogers, A., 2010. Twitter use by the U.S. Congress. J. Am. Soc. Inf. Sci. Technol. 61, 1612–1621. https://doi.org/10.1002/asi.21344

Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M., 2008. **ergm** : A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. J. Stat. Softw. 24. https://doi.org/10.18637/jss.v024.i03

James Holland Jones, Elspeth Ready, Ashley Hazel, 2020. Introduction to ERGMs [WWW Document]. Soc. Netw. Anal. Anthropol. URL https://eehh-stanford.github.io/SNA-workshop/ergm-intro.html (accessed 9.5.24).

Jayawickrama, T.D., 2021. Community Detection Algorithms [WWW Document]. Medium. URL https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae (accessed 9.5.24).

Leinwand, B., Lyzinski, V., 2024. ACRONYM: Augmented degree corrected, Community Reticulated Organized Network Yielding Model.

Luce, R.D., Perry, A.D., 1949. A method of matrix analysis of group structure. Psychometrika 14, 95–116. https://doi.org/10.1007/BF02289146

Mao, J., Zou, D., Sheng, L., Liu, S., Gao, C., Wang, Y., Li, Y., 2024. Identify Critical Nodes in Complex Network with Large Language Models.
Maoz, Z., 2012. Preferential Attachment, Homophily, and the Structure of International Networks, 1816–2003. Confl. Manag. Peace Sci. 29, 341–369. https://doi.org/10.1177/0738894212443344
Newman, M.E.J., 2008. Mathematics of Networks, in: Palgrave Macmillan (Ed.), The New Palgrave Dictionary of Economics. Palgrave Macmillan UK, London, pp. 1–8. https://doi.org/10.1057/978-1-349-95121-5_2565-1
Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113. https://doi.org/10.1103/PhysRevE.69.026113
Newman, M.E.J., Park, J., 2003. Why social networks are different from other types of networks. Phys. Rev. E 68, 036122. https://doi.org/10.1103/PhysRevE.68.036122
Nickerson, D.W., 2008. Is Voting Contagious? Evidence from Two Field Experiments. Am. Polit. Sci. Rev. 102, 49–57.
Ognyanova, K., 2016. Network analysis with R and igraph: NetSci X Tutorial. Katya Ognyanova. URL https://kateto.net/networks-r-igraph/ (accessed 9.5.24).
Otte, E., Rousseau, R., 2002. Social Network Analysis: A Powerful Strategy, also for the Information Sciences. J. Inf. Sci. 28, 441–453. https://doi.org/10.1177/016555150202800601
Padgett, J.F., Ansell, C.K., 1993. Robust Action and the Rise of the Medici, 1400-1434. Am. J. Sociol. 98, 1259–1319. https://doi.org/10.1086/230190
Pavel N. Krivitsky, Martina Morris, Mark S. Handcock, Carter T. Butts, David R. Hunter, Steven M. Goodreau, Chad Klumb, Skye Bender de-Moll, Michał Bojanowski, 2024. Exponential Random Graph Models (ERGMs) using statnet [WWW Document]. statnet.org. URL https://statnet.org/workshop-ergm/ergm_tutorial.html#4_Assessing_convergence_for_dyad_dependent_models:_MCMC_Diagnostics
Seidman, S.B., 1983. Network structure and minimum degree. Soc. Netw. 5, 269–287. https://doi.org/10.1016/0378-8733(83)90028-X
Vakil, C., 2022. Cortez Masto wins in Nevada, securing Democratic control of Senate [WWW Document]. URL https://thehill.com/homenews/campaign/3709607-cortez-masto-wins-in-nevada-securing-democratic-control-of-senate/ (accessed 9.6.24).
Wang, Y., Zhang, L., Yang, J., Yan, M., Li, H., 2024. Multi-factor information matrix: A directed weighted method to identify influential nodes in social networks. Chaos Solitons Fractals 180, 114485. https://doi.org/10.1016/j.chaos.2024.114485
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511815478
Yang, S., Keller, F.B., Zheng, L., 2017. Social Network Analysis: Methods and Examples. SAGE Publications, Inc. https://doi.org/10.4135/9781071802847