Analysis of Politicians' Tweets to Explore Political Communication with Social Network Analysis

Heidi Schuhbauer, Sebastian Schötteler, Johannes Niu, Bernhard Schiffer,
David Wolfarth, Tim Groß, Thomas Kreidl, and Fabian Model

Computer Science Department
Nuremberg Institute of Technology
Nuremberg, Germany
heidi.schuhbauer@th-nuernberg.de; sebastian.schoetteler@th-nuernberg.de; niujo64943@th-nuernberg.de; schifferbe69185@th-nuernberg.de; wolfarthda82341@th-nuernberg.de; grossti74853@th-nuernberg.de; kreidlth74662@th-nuernberg.de; modelfa74656@th-nuernberg.de

***Abstract*—This paper illustrates the practical application of cluster analysis, social network analysis, sentiment analysis, and topic analysis in a case study on Twitter data. These techniques provide insights into the public communication patterns between German Members of Parliament (MPs) on Twitter around the time of the 2021 federal election. The question of this work was to determine whether a potential shift in communication towards the inaugurated "Ampel" coalition, made up of the parties SPD, Greens, and FDP, can be derived from Twitter interactions. Twitter data were collected and separated into two time slots: before and after the election. In distinct scenarios, mention, retweet, and reply interactions are first considered together and then separately. In these scenarios, the Girvan-Newman Algorithm detects clusters of MPs dependent on the interactions observed. Then, the average inbreeding homophily and other network metrics of the pre- and post-election area are compared. An additional scenario focuses on intra- and inter-party sentiments conveyed within tweet texts. In a fourth scenario, MPs are grouped according to their party affiliation, the average inbreeding homophily values of parties, and potential coalitions. A topic analysis handled relevant discussion topics between the successful coalition parties. Changes in communication behavior at these two different time slots are visible. The communication clusters of those MPs differ mostly before and after the election. The average sentiment of the parties towards each other changed positively, although no significant tendency could be derived regarding later coalition formations. A determination of the density of the network for the topics also indicates that the importance of the individual topics before and after the formation of the government has a consistent relevance.**

Keywords-Cluster Analysis; Microblog; Network Metrics; Sentiment Analysis; Social Network Analysis; Topic Analysis.

I. INTRODUCTION

This paper is an extended version of the one presented in SOTICS 2022 [1].

For political communication between parties, politicians, and their constituents, social media platforms play an important role. By communicating through platforms, such as Facebook, Twitter, and Instagram, political actors reach wide audiences within a short period. On these platforms, politicians publicly communicate with each other. As one consequence of this trend, social media platforms are gradually outclassing traditional media outlets – such as television and print media – as the main source of news and information. Case in point, a survey conducted in 2018 determined that social media platforms are the most popular information and news sources for American adults aged between 18 and 29 [2]. Corroborating, another survey conducted in 2021 determined that up to 76% of adults rely on social media as a news and information source, depending on the country [3].

Among those services, Twitter is a very popular microblog that is used by many persons also for professional use. Twitter promotes the dialogue between politicians and between politicians and their constituents via mention and reply interactions, which allow users to engage in direct communication. Consequently, social media have become a central component of political communication.

Relations between individual MPs can be examined in more detail using social network analysis. "A social network consists of a finite set or sets of actors (depicted as nodes) and the relation or relations defined on them (depicted as ties)" [4]. Translated to microblogs, two types of social networks can be derived from such platforms – follower and interaction networks. A follower network can be defined as a set of microblog users connected via follower ties regarding who follows whom. An interaction network can be defined as a set of microblog users connected via interaction ties that diffuse content between said users. The application of social network analysis on Twitter data offers a variety of research possibilities. Interactions can be derived from public tweets referring to other people, i.e., retweets of or replies to another user's tweet, or mentions of a user. Analysis of interaction networks explores these relations, as well as their textual contents, which can be examined through sentiment analysis.

This article applies methods of social network analysis to explore changes to the communication of German MPs from selected political parties around the 2021 federal election. After 16 years of a government led by the Union, made up of the Christian Democratic Union (CDU) and Christian Social Union (CSU), – in coalition with the Social Democratic Party of Germany (SPD) for three and the Free Democratic Party (FDP) for one legislative period – Germany had a change in government after the federal elections in 2021. The new government consisting of the SPD, Alliance 90/The Greens (Greens), and FDP was formed on December 8, 2021. For the first time in its history, a government formed by a coalition of more than two parliamentary groups thus governs the Federal Republic of Germany.

In the past, federal government alliances consisted of coalitions between the Union/SPD, Union/FDP, SPD/FDP, and SPD/Greens. A coalition between the Greens and FDP within a tripartite coalition currently governs in the states of Schleswig-Holstein and Rhineland-Palatinate, in a so-called Jamaica coalition (together with the CDU) and an Ampel ('traffic light') coalition with the SPD, respectively. On a federal level, however, these two parties had yet to form a coalition.

This article applies social network analysis, cluster analysis, sentiment analysis, and topic analysis to explore changes to the communication of German MPs from select political parties around the 2021 federal election.

Section II of this paper presents related works, formulates the research gap, and specifies the hypotheses. Section III introduces the methodology used to aggregate and analyze the data for the network scenarios. Section IV presents the results of each network scenario and discusses them. Section V describes the methods for the topic analysis of relevant subjects discussed by the "Ampel" coalition parties who won the election. Section VI contains the result of the topic analysis. Section VII illustrates the limitations of this research, as well as starting points for possible future work.

II. Research Gap

Virk [5] compares different Social Network Services (SNS) as a type of social media and explores the special role of Twitter in public communication. The author examines the communication patterns between Twitter users and applies the tie strength theory postulated by Granovetter [6] to conclude that interactions on Twitter – unlike other SNS – focus on content rather than user relationships, and thus can reach wider audiences.

Lassen and Brown [7] examine Twitter use by members of congress in the United States of America. They state that SNSs enable politicians to communicate more directly and personally with peers and supporters by eliminating limits on message visibility, allowing content to be redistributed beyond one's followers. The application of social network analysis to political networks shows the fragmentation and clustering of politicians, parties, or political systems.

Boireau [8] identifies communities among Belgian MPs along party and linguistic lines. For this purpose, the Girvan-Newman Algorithm (GNA) was applied on a network generated from the MPs' connections to followers, and retweet interactions to find hidden communities and homogeneous clusters by calculating their homophily indices, which express the degree of similarity of members within a cluster.

Caetano et al. [9] analyze social networks among Twitter users during the 2016 American presidential election by analyzing tweets about the candidates. Users were clustered based on their sentiment towards a candidate with their mentioning behavior and hashtag use. By obtaining homophily indices of these clusters, the authors could identify users with high degrees of relative similarity.

Sentiment analysis attempts to quantify attitudes conveyed in a text. Giachanou and Crestani [10] discuss common procedures for sentiment analysis, as well as their respective limitations, e.g., the detection of irony or emotions. The work explicitly focuses on methods suitable to retrieve sentiments from tweets.

Boras and Singh [11] show that Twitter is the social network most used for political discussions. However, the intensity of the discussion is strongly dependent on the topic. In Meier et al. [12], among other things, the issue of engagement of German politicians on certain topics was examined in the election year 2017. To determine which topics were intensively discussed on Twitter, a hashtag network was created, with which each tweet with a hashtag could then be directly assigned to a specific issue. Using additional metrics, such as the Z-score or the Q-modularity, it was found that the communication behavior of the parties changed significantly before the election, especially in terms of interaction and discussion with other parties. A topic analysis of political discussion topics on Twitter based on hashtags was also done by Boras and Singh [11] for the Indian political landscape. To examine whether discussion on Twitter about the identified topics leads to political polarisation, a mention and retweet network was formed. These networks can indicate discussion and advocacy. Content and sentiment analysis can be used to identify shifts in opinion on specific topics as well as internal network relationships. Garcia-Sanchez et al. [13] used cluster analysis to identify different ideologies concerning various political issues of Brazilian parliamentarians in 2019. Using the degree of centrality, it was possible to identify particularly active users and thus the central key figures. A comparison between possible techniques for topic modeling can be found in Egger et al. [14]. The techniques investigated here were latent Dirichlet allocation (LDA), non-negative matrix factorization (NMF), Top2Vec, and BERTopic. Tweets are used as the data basis of the comparison, and the advantages and disadvantages in the different application areas are discussed.

Until now, literature does not describe possible changes in Twitter communication behavior between MPs before and after an election. An exploration of the change in tone by analyzing the sentiment of tweets before and after an event has also not yet been described. An analysis of the topics discussed by the parties is missing. Interesting aspects of political communication behavior on social media are expected results of this analysis.

Consequently, this article examines how Twitter interactions (mentions, retweets, replies) between MPs of possible coalition partners (CDU, CSU, SPD, Greens, FDP) changed before and after the 2021 German federal election. It furthermore explores potential differences in intra- and inter-party communication and attempts to show whether the political shift towards the inaugurated "Ampel" coalition could be derived from the observed changes. Changes in intra- and inter-party communication are additionally presented in relation to the various topics.

The following hypotheses form the basis for the communication behavior analysis: The article hypothesizes that different interactions between MPs can be observed during the pre- and post-election period (H1) and that the resulting interaction networks for each period show a difference in intra- and inter-party communication (H2). The article further assumes that “Ampel” MPs’ mutual sentiment changed positively (H3). By analyzing the sentiment between parties, as well as the average homogeneity within parties and party groups, political tendencies towards an “Ampel” coalition can be observed (H4). In the topic networks, intra-party homophily changes to inter-party homophily as a function of the observation time points (H5). For the change in communication, it is also important to consider the density of the social network. The communication intensity of the topic networks depends on the observation times (H6).

Thus, this article attempts to describe the change in communication between MPs by analyzing their Twitter interactions before and after the federal election 2021. It aims to understand whether changing interaction intensities between MPs of potential coalition partners yield conclusions about the emerging “Ampel” coalition. This would be of relevance for future research into the interdependencies of political communication on Social Network Services, such as Twitter.

III. Methods for Scenarios

Mention, retweet, and reply interactions between MPs from the SPD, Greens, FDP, CDU, and CSU were collected to explore changes in communication on Twitter. One MP using another MP’s handle denotes a mention interaction. A Twitter handle, which is commonly known as a username**,** is the name with which a user has registered on Twitter. Since it serves as an account’s identifier, no two usernames on the social network are the same [16]. Retweets refer to the redistribution of another user’s tweet and can contain commentary by the retweeter. A reply is defined as a comment posted under another MP’s tweet. The resulting social networks of MPs connected by their interactions are analyzed in four separate scenarios.

A. Network Scenarios

Scenario 1 considers all interaction types, while in scenario 2, a) mention, b) retweet, and c) reply interactions were examined separately. For each scenario, MPs were grouped using automated cluster detection and examined for modularity and homophily. Modularity measures the strength of division of a network into clusters. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules [17]. Homophily in network structures means the principle that nodes in a network tend to have links to other nodes with similar attributes [14].

In scenarios 3 and 4, MPs were grouped based on their party affiliation. In scenario 3, interactions were examined for the tweet author’s sentiment towards the addressed MP using sentiment analysis. The sentiment for every interaction was evaluated based on the tweet’s text. To determine changes to the inter-party relations, each party’s average sentiment toward all other parties was then calculated and compared between the pre- and post-election networks. Scenario 4 examined the average homophily within each party and party group. Party groups were based on politically and numerically possible coalition compositions (“Ampel”, “Jamaica”) and for the Union parties.

B. Data Aggregation

Publicly available Twitter data can be divided into three categories: (1) User information, such as the username, the Twitter handle (identified by @), or account description; (2) following and liking behavior of a user, and the user’s followers; (3) the user’s tweet timeline, in which all self-published or retweeted tweets appear, as well as the user’s replies to others’ tweets.

As a basis for this study, publicly available tweets from MPs of the 19th (2017-2021) and 20th (2021-2025) legislative sessions were collected for the period from July 26, 2021, 0:00 a.m. to November 26, 2021, 12:00 p.m. The end date was chosen to serve as cut-off due to the official presentation of the coalition agreement between the SPD, Greens, and FDP on November 24, 2021. To collect reactions to this announcement, two more days were added. The period between the closing of polls on September 26, 2021, at 6 p.m., and the end date covers 60 days and is considered as the post-election period. An equally long time before the closing of polls was considered for the pre-election period.

Twitter accounts were selected from all MPs with a public Twitter timeline who are members of the parties SPD, CDU, CSU, Greens, and FDP. Members of the parties “The Left” and AfD were not included in this analysis, as neither party was relevant for coalition negotiations after the election. The timelines of all selected accounts were then scraped from Twitter’s website.

Data Collection. Scraping of timelines was done using the Python package Scweet [15]. Scweet uses the Chrome plugin Selenium [19] to access the desired Twitter page, extract the information of the tweet from the page, and save it to a CSV file.

Data Processing. A custom Java application was developed to generate uniformly formatted and sanitized datasets. The data originally scraped from Twitter included the timelines of all MPs, i.e., all their tweets, retweets of, and replies to other tweets within the time frame. The information generated for each of these messages included the time of publication, the author’s username and handle, the textual contents of the tweet, as well as information on whether it was posted as a retweet of, or reply to another tweet. If other users were mentioned within the tweet, they could be identified through their handles.

Additionally, the application enriches the data with information on party affiliation and membership of the 19th or 20th legislative period. It produced output data in the GEXF format [20], which is limited by specified procedures. First, all tweets that did not represent a connection between two MPs were removed. The dataset was then divided into a pre- and a post-election partition. For this purpose, all tweets that were created before the time of the closing of polls on September 26, 2021, 6:00 p.m. were assigned to a first partition. The elements from the timeline after this date were assigned to a second partition. Additionally, the output is restricted to specific interaction types. This allowed the creation of one pre-election and one post-election dataset for each of the scenarios defined.

Data Description. The data set collected from Twitter consisted of 26,888 German-language tweets from 736 Twitter accounts. 15,770 of these tweets were posted before and 11,118 after election day. 1,030 MPs were elected for the 19th and 20th legislative periods. 71.5% of them maintained a Twitter account. Once filtered, the dataset consisted of 622 accounts and 9,582 tweets. After removing all tweets that did not connect two MPs, 5,766 tweets from 466 MPs remained in the pre-election dataset, and 3,816 from 476 MPs in the post-election dataset. Figure 1 shows the percentage distribution of all tweets among the parties before and after the election.

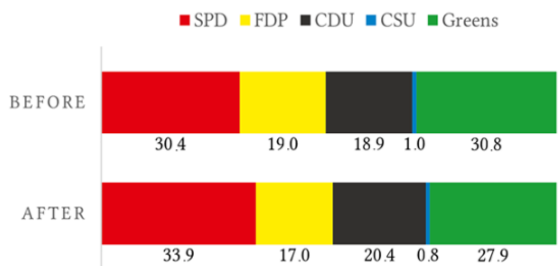

Figure 1. Percentage distribution of MPs' tweets by party

The pre- and post-election data contain nodes and edges depending on the interaction types selected during the data processing step. Scenarios 1 and 4 thus contained all MPs, while scenario 2 contained three separate data sets, differentiated by interaction types. Scenario 3 handled only those interaction types whose tweet text fields were not empty. The aggregated data and source code can be accessed at [14].

C. Cluster Detection

Cluster detection extracts groups of individuals from a network based on the similarity of one or more attributes. This work used connectivity-based clustering, which identifies clusters based on the connections between nodes in the network, as well as the weights of connections. For this purpose, the Girvan-Newman Algorithm [22] was used. This algorithm assumes that members of a cluster have more connections to other members of the same cluster, and fewer connections to other nodes in the remaining network. By iteratively removing connections whose Edge Betweenness Centrality (EBC) is the highest, clusters are separated from each other. The EBC is defined as "the number of [the] shortest paths between pairs of vertices that run along it" [10]. In each step, the edge with the highest EBC is removed from the network and its modularity is calculated. The modularity of a network denotes how well clusters are separated from each other. The iteration continues until every connection between nodes has been eliminated. The intermediate step with the highest modularity is the result of the algorithm.

To guarantee that an MP's allocation to a cluster is based on their interactions and not their party affiliation, a χ^2 test is performed on the network. The test's p-value denotes the probability p of MPs' party affiliation determining the results of the cluster detection.

D. Sentiment Analysis

The textual contexts of MPs' tweets were examined to analyze the sentiment for which the Python package TextBlob [15] was used. The package uses a lexicon-based approach to compute the sentiment. For the analysis of German language texts, the plugin TextBlobDE [24] was used. A predefined dictionary of words associated with positive or negative emotions is used to weigh a text's sentiment. An individual score is assigned to each word in the examined text. The overall sentiment is defined by the average sentiment across all words in the text. The algorithm generates a polarity score from –1.0 to +1.0 for each tweet, which classified the tweet as either positive, neutral, or negative. Each tweet in the data set is then enriched with the polarity value, as well as the polarity class as additional attributes.

E. Homophily

The homophily index H measures a cluster's relative homogeneity. To determine H for a cluster i, the connections of all nodes of the cluster are examined. Caetano et al. [9] calculate $H_i = \frac{s_i}{si+d_i}$ where s_i denotes homogeneous links, i.e., those that connect a node of class i to other nodes of the same class, while d_i denotes heterogeneous connections, i.e., those that connect a node of class i to nodes of another class. By normalizing H_i over the whole network, H can be compared across different clusters. This inbreeding homophily index IH is determined by $IH_i = \frac{H_i - w_i}{1 - w_i}$, where w_i denotes the relation of nodes between cluster i and the total number of nodes in the network. Clusters whose IH_i is greater than 0 are considered homogeneous. The average of IH across all clusters in a network is used to compare the clusters detected in the pre- and post-election networks.

F. Evaluation

The procedure resulted in a set of network pairs, each consisting of a pre- and a post-election network. The two networks created for scenario 1 contained all MPs that have interacted via mentions, retweets, or replies within the respective timeframe. The number of connections between two nodes weighted the edges.

Scenario 2 generated one network pair for each of the three interaction types. Thus, one pre- and one post-election network each were generated, which included all those MPs that a) mentioned each other, b) replied to one another, and c) retweeted each other. Edges represent the connections. They are weighted by the interaction count. These scenarios were examined separately. For each network, automated cluster detection was applied. The H and IH indices were calculated to determine the homogeneity of each cluster. Additionally, the number of nodes and edges in the network, the number of clusters identified by the GNA, as well as their networks' average homophily and inbreeding homophily indices, and the maximum modularity were determined. Statistical significance was ensured using the χ^2 test. The results of these analyses were then compared for the pre- and post-election network pair. To illustrate the results of the automated cluster detection, each pre- and post-network pair is visualized as a cluster graph.

In scenario 3, each party's average sentiment towards all other parties was examined. For this purpose, MPs were clustered according to their party affiliations.

Scenario 4 looked at the inbreeding homophily of each party, as well as the coalition options before and after the election. The *IH* values for the coalitions were also checked for statistical significance using the χ^2 test and its p-value.

IV. RESULTS OF THE SCENARIOS

A. Scenario 1: Multiple Interactions

In scenario 1, automated cluster detection included all interaction types. An overview of the collected metrics can be found in Table I.

TABLE I. NETWORK AND CLUSTER METRICS CONSIDERING ALL INTERACTIONS

Metric	Value (pre)	Value (post)	Difference
Number of nodes	466	476	10
Number of edges	5766	3816	-1950
Number of clusters	256	188	-68
Maximum modularity	0.026	0.356	0.330
Average *IH*	0.0212	0.0571	0.0359
p-value from χ^2-Test	< 0.001	< 0.001	

The number of MPs (nodes) tweeting after the election did not vary much from that before the election. However, the number of connections (edges) was reduced by 33%, which suggests that tweeting activity was distributed more equally among MPs after the election. The GNA identified 256 clusters of the pre-election network with 466 MPs, and very low modularity, homophily, and inbreeding homophily indices. After the election, 476 MPs could be assigned to 188 clusters. The maximum cluster size was reduced by 54.5% to 97. The modularity increased by 1369%, from 0.026 to 0.356, and homophily and inbreeding homophily also increased significantly. Figure 2 shows a visualization of these clusters. Node colors represent each MP's party affiliation. The size of a node depicts the sum of all incoming and outgoing edges, i.e., the node's degree. Edges were omitted from these figures for improved visibility.

Pre-election, the visualization shows a distinctive, large cluster that unites MPs across all parties. Outside of this cluster, many MPs are scattered into tiny groups or unassigned to any notable cluster. Post-election, four large clusters separated along party affiliation can be identified. A heterogeneous group of MPs was not assigned to any notable cluster.

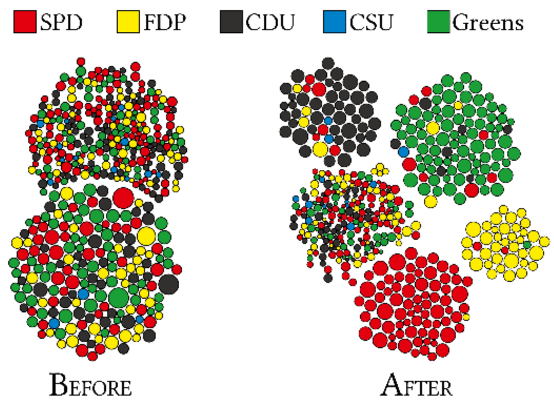

Figure 2. Clusters found by GNA before and after the election considering all interactions

The pre-election results of scenario 1 show that MPs were likely allocated to the dominant cluster based on their general activity on Twitter. Nodes with higher degrees were allocated to the dominant cluster. Post-election, distinct clusters are clearly separable, which consist mainly of MPs of either the SPD, CDU, Greens, or FDP. The number of nodes that could not be allocated to any major cluster decreased. This indicates that post-election, MPs predominantly communicated within their parties, while they communicated much more openly before the election. The overall count of interactions decreased significantly.

B. Scenario 2: Single Interactions

When interaction types are considered separately, these findings can be analyzed in more detail.

Mentions. In this particular scenario, clusters were determined based on mentions only. Table II shows the collected metrics.

TABLE II. METRICS OF NETWORK AND CLUSTERS DERIVED FROM MENTIONS

Metric	Value (pre)	Value (post)	Difference
Number of nodes	433	428	-5
Number of edges	3247	1758	-1489
Number of clusters	95	38	-57
Maximum modularity	0.237	0.441	0.204
Average *IH*	0.1158	0.4550	0.3292
p-value from χ^2-Test	< 0.001	< 0.001	

Almost as many (433 vs 428) MPs mentioned one another in the pre- and post-election period. Interactions decreased by 54%, and the number of detected clusters decreased by 40%. After the election, 38 clusters with a modularity of 0.441 could be identified, compared to 95 clusters with a modularity of 0.237 before the election. The average IH across all clusters in both networks increased by more than 300%. Figure 3 visualizes the detected clusters.

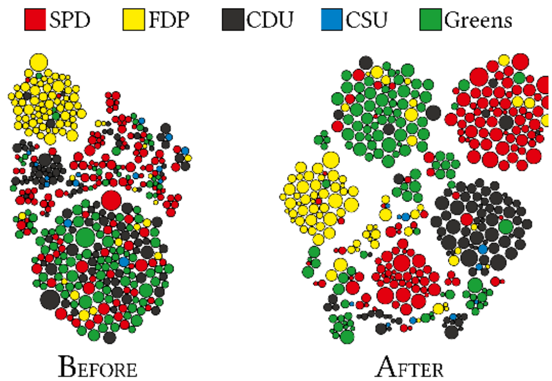

Figure 3. Clusters found by GNA before and after the election considering only mentions

Pre-election, three distinct clusters can be identified, one portraying a large cluster mainly dominated by Greens but including MPs across all parties, one dominated by FDP MPs, and a smaller one dominated by CDU MPs. The large, heterogeneous cluster dominated by Green MPs could be caused by many mentions of the Greens' chancellor candidate, Annalena Baerbock.

Distinct clusters are detected in the post-election network separated along party lines. Two SPD clusters are found, as well as several smaller but still homogeneous clusters. The number of mentions increased. Subsequent analysis revealed that the distinct party clusters might be caused by MPs congratulating their party peers.

In the next analysis, only retweet connections were examined. The metrics on this can be found in Table III.

TABLE III. METRICS OF NETWORK AND CLUSTERS DERIVED FROM *RETWEETS*

Metric	Value (pre)	Value (post)	Difference
Number of nodes	178	144	-34
Number of edges	253	151	-102
Number of clusters	19	26	7
Maximum modularity	0.713	0.825	0.112
Average *IH*	0.8809	0.9567	0.0758
p-value from χ^2-Test	< 0.001	< 0.001	

The number of nodes was reduced by 19.1% from 178 to 144, and the number of edges by 40% (253 to 151). Nineteen clusters could be identified before the election, and twenty-six after, which explains the reduction of the median cluster size from 6 to 2.5. Conversely, the modularity increased slightly, just like the average IH. Figure 4 shows a visual representation of both networks' clusters.

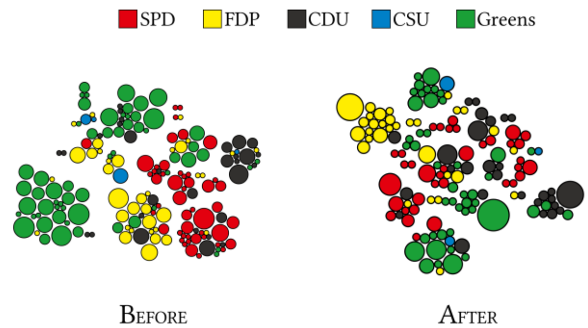

Figure 4. Clusters found by GNA before and after the election considering only retweets

Clustering detected several well-separated clusters with relatively high homogeneity both before and after the election, with smaller clusters found in the latter. A possible explanation is that MPs attempted to promote the tweets of party peers. The clusters in the post-election network were smaller. Retweets play a smaller role in communication among MPs.

The last analysis of this section considered reply connections only. The metrics can be found in Table IV.

TABLE IV. METRICS OF NETWORK AND CLUSTERS DERIVED FROM *REPLIES*

Metric	Value (pre)	Value (post)	Difference
Number of nodes	351	388	37
Number of edges	2266	1907	-359
Number of clusters	233	24	-209
Maximum modularity	0.113	0.401	0.288
Average *IH*	0.0636	0.7112	0.6476
p-value from χ^2-Test	< 0.05	< 0.001	

More nodes could be identified in the post-election network, but fewer connections between them. The algorithm identified 233 clusters before and 24 after the election, a reduction of 89.7%. Modularity and IH increased significantly. A visualization of identified clusters can be found in Figure 5.

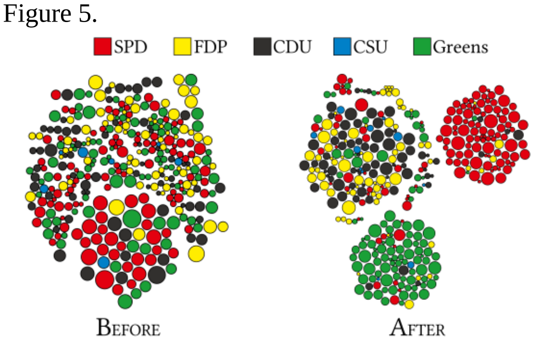

Figure 5. Clusters found by GNA before and after the election considering only replies

Pre-election, one large and many small clusters were found, with the main cluster containing many nodes with a high in– and out-degree.

Post-election, two main clusters were identified, notably consisting mainly of SPD and Green party members respectively. Another large cluster contained a heterogeneous mix of MPs.

Solely considering reply interactions, one large and many small clusters were found in the pre-election network. The main cluster contains many nodes with a high in– and out-degree. In the post-election network, more nodes are identified but fewer connections between them are found. Two main clusters were identified, notably consisting mainly of SPD and Green party members. One cluster of CDU and FDP MPs indicates active conversations between these two parties, potentially on the FDP's willingness to enter coalition negotiations with the SPD and Greens shortly after the election, which supports hypothesis H1.

C. Scenario 3: Sentiment Analysis

Each interaction's textual content was analyzed to retrieve the parties' mutual sentiments. The average sentiment of interactions from MPs of one party towards MPs of the other parties was calculated. The results are shown in Table V. Notably, polarity does not score very highly overall, except for the sentiment from MPs of the CSU towards MPs from the CDU. FDP MPs communicated neutrally in general. The SPD scores positively towards the "Ampel" parties. On average, Green party MPs showed positive polarities only towards other MPs of their party.

TABLE V. AVERAGE SENTIMENT BETWEEN PARTIES BEFORE THE ELECTION

Source \ Target	SPD	FDP	CDU	CSU	Greens
SPD	0.25001	0.21293	0.00002	-0.11499	0.35683
FDP	0.05095	-0.01008	0.06981	0.09734	0.01032
CDU	0.00070	0.02997	0.13179	-0.12469	0.00483
CSU	0.04297	-0.01875	0.70728	0.10625	0.11405
Greens	-0.16582	0.03257	-0.16458	0.00053	0.35588

Table VI shows the average sentiment between parties after the election. The post-election sentiments between parties notably tend towards an overall positive sentiment. The SPD received overall positive interactions, especially from the CDU. The SPD communicated relatively neutrally, both internally, as well as towards their subsequent coalition partners. The polarity of the interactions among MPs of the Greens and interactions from MPs of the CSU towards CDU MPs did not change significantly from their pre-election scores. The overall sentiment across all parties after the election was on average more positive than before the election. The FDP especially shows notable increases in positive sentiments towards the SPD and the Greens, considering that the FDP moved towards the "Ampel". This strongly hints at successful coalition negotiations, which ended with the signing of the coalition contract.

TABLE VI. AVERAGE SENTIMENT BETWEEN PARTIES AFTER THE ELECTION

Source \ Target	SPD	FDP	CDU	CSU	Greens
SPD	0.00166	0.25408	0.31106	0.84063	0.06433
FDP	0.33102	0.54495	-0.11953	0.00391	0.27281
CDU	0.79865	-0.00598	0.09291	-0.08487	0.67012
CSU	-0.16250	0.24688	0.59688	0.12500	0.39146
Greens	0.43225	0.09978	0.62791	-0.06024	0.38109

D. Scenario 4: Party and Group-Dependent Clustering

In this scenario, MPs were clustered along party affiliation. Additionally, the two potential government coalitions, "Ampel" (SPD, Greens, FDP) and Jamaica (CDU, CSU, Greens, FDP), as well as the Union (CDU, CSU), were clustered. To compare the homogeneity within each cluster, the average *IH* before and after the election was calculated and compared. Table VII displays the average *IH* values of each party, as well as the coalition and union clusters for the pre- and post-election networks.

TABLE VII. RELATIVE IH IN PARTIES AND PARTY GROUPS

	Before	After	Difference
CDU	0.4749	0.5596	0.0848
CSU	0.0721	0.0516	-0.0205
SPD	0.5272	0.6392	0.1120
Greens	0.5682	0.5729	0.0047
FDP	0.5397	0.4618	-0.0779
"Ampel" Coalition	0.4519	0.6272	0.1754
Jamaica Coalition	0.5209	0.5307	0.0098
Union Group	0.4632	0.5422	0.0791
p-value from χ^2-Test	0.057764	0.106983	

The biggest differences are between the SPD and CDU. Their relative homophily increased. CSU and FDP decreased in *IH*. SPD received the biggest increase in homogeneity. This could be explained by their win of the election, and the positive feedback MPs received from their peers, as well as the election of SPD MPs Olaf Scholz as chancellor and Bärbel Bas as president of the parliament. The biggest positive change among grouped MPs took place in the "Ampel" coalition, but *IH* increased for the Jamaica and Union clusters as well. However, a significant statistical independence of these findings is not reliably provable, as the χ^2-test results in relatively high p-values for the pre- and post-election homophily.

V. METHODS FOR TOPIC ANALYSIS

The change of government was preceded by a volatile election campaign, which was interspersed with multifaceted debates on various topics [25]. The change of working communities, which also include public decision-making bodies, is often accompanied by a change in the working and communication climate because the inner diversity changes [26]. The analysis of these changes helps to understand how the patterns of interaction within and across parties change over this period. In the context of this study, a qualitative analysis of the interactions between the members of the factions of the governing parties is done based on exploratory research. Changes in the social network for certain topics are examined. Changes in intra- and inter-party communication are presented in relation to relevant topics.

A. *Data Work*

The topic analysis focuses on the communication between the traffic light parties because this was the successful coalition. For this purpose, a new data acquisition had to be done:

The tweets of all members of the traffic light coalition are used as the data basis for further network analyses. In the first step, the Twitter handles of the parliamentarians are fetched from a list of all members of the Bundestag, which is dynamically managed and made available by Twitter [27]. This data scratch is implemented using the Twitter API and the Python library Tweepy [28]. The data of the first period refers to the time from 26 September 2020 00:00 to 25 September 2021 23:59. The comparison period is defined similarly, offset by one year, from 26 September 2021 00:00 to 25 September 2022 23:59. This period is longer than the period of the previous analysis to get more data. The data contains information on the date, the text, the user ID, and the tweet ID. In addition, information on retweets, replies, and mentions are included, which are important for the formation of the networks in the further process of the study. Since party affiliation plays an important role, this information is subsequently added to the individual tweets with the help of master data from the German Parliament [29]. With this information, only tweets from members of the Greens, the SPD, or the FDP are kept to sort out the tweets from the opposition parties. In total, the remaining dataset contains tweets from 118 members of the three parties with approximately 120,000 tweets. No tweets could be found for the remaining 298 members of the government parties.

To carry out a topic analysis, the data is first pre-processed. In this process, only tweets with their conversation ID are retained and tweets with a line length of less than 100 are sorted out. In addition, interfering words and characters such as links or Twitter-specific characters are removed from the tweets. Pre-processing reduces the size of the data to approximately 80,000 tweets, which are analyzed with the help of the topic modeling technique BERTopic [30] and classified into 10 topics. Only 10 topics are determined to prevent fragmentation into several smaller topics. This is necessary so that representative networks of an appropriate size can be formed. Clustering into many topics greatly reduces the number of tweets on a topic.

Figure 6. Word Scores per Topic

The Topic Word Scores in modelfa74656@th-nuernberg.de 6 show the distribution of the detected topics. The most frequent five words are listed for each of the 10 topics. The division of the topics results in approximately 65,000 tweets that can be assigned to a topic. This results in approximately 55,000 tweets that are present in the original data set of 120,000 tweets but could not be assigned to a topic. In the next step, connections are detected between the assigned and unassigned tweets to be able to assign these tweets to a topic. With the help of retweets and replies, about 10,000 unassigned tweets can be assigned to a topic. It is assumed that a reply and a retweet mean that the tweet is about the same topic as the original tweet. This procedure results in the topics, which are shown in Table VIII and the respective number of tweets they contain.

TABLE VIII. TOPIC EVALUATION

Topic	Amount	IRR	Precision	Topic Assessment
0	**6500**	**92 %**	**95 %**	**✓ (Climate)**
1	6500	96 %	93 %	× (Ukraine-Conflict)
2	3000	75 %	81 %	× (Education)
3	**2000**	**98 %**	**98 %**	**✓ (Vaccination)**
4	4500	89 %	20 %	× (Covid)
5	**5000**	**92 %**	**100 %**	**✓ (Finance)**
6	30000	88 %	7 %	× (Elections)
7	5500	73 %	27 %	× (Disasters)
8	5000	80 %	8 %	× (Europe)
9	3000	68 %	14 %	× (Conflicts)

To select suitable topics, it is necessary to determine their quality in advance. For this purpose, suitable generic terms for the respective topics are sought at the beginning. In cooperation with a political scientist, suitable terms are defined for the keywords in Figure 6.

The evaluation of the assessment is done using the interrater reliability (IRR) value. This involves verifying whether the tweets thematically match the keywords and the generic terms. The procedure is based on Newman et *al.* [31]. The IRR value is used to check whether all observers agree on the assessment. A representative sample of 50 tweets is reviewed for each topic and the IRR value is calculated from this. Only topics whose IRR value is high enough are of sufficient data quality and are therefore taken into account in the subsequent analyses. Values in the range of at least 80 percent are considered high enough.

To determine the accuracy of the assignment of the tweets to the topics, all tweets of a topic are included for which all observers agree that the generic term and the keywords match the respective tweet or not. The IRR for the tweet under investigation must therefore be 100 percent. The values for calculating the accuracy are 0 or 1, depending on whether the topic matches the tweet (1) or not (0). The mean value is determined from these binary values, which corresponds to the accuracy in percent.

The result of the evaluation indicates that 7 from 10 topics have a sufficiently high IRR value. However, the accuracy is only high enough for topics climate, the Ukraine conflict, education, vaccination, and finance. Topic 1 (Ukraine-Conflict) is omitted, as this topic is almost exclusively discussed after election day, and thus no comparison is possible concerning the two observation periods.

Finally, this leads to the topics of climate, vaccination, and finance, which are analyzed because their IRR values and accuracies are high enough.

B. Data Analysis

First, graphs are generated for visualization and analysis with the software Gephi [14a] using the Python package NetworkX [32]. These contain all nodes and edges of a respective topic. A node corresponds to a participant in the communication network of the topic, thus a Twitter account. An edge corresponds to a connection between two participants in the topic network. The connection can take the form of a mention, a retweet, or a reply. During the creation process, self-loops are removed from the graph, as they have no relevance in the study of interactions between politicians. In addition, the attribute "party" with the value FDP, SPD, or Greens is added to each node for the subsequent analyses. All remaining nodes that cannot be assigned to an “Ampel” party are removed. The resulting graphs are visualized in Gephi and are exemplarily shown for Topic 0 (Climate) in Figure 7 before (on the left side) and after (on the right side) the election. The colors of the points visible in the following images represent the party colors of the SPD (red), FDP (yellow), and Greens (green).

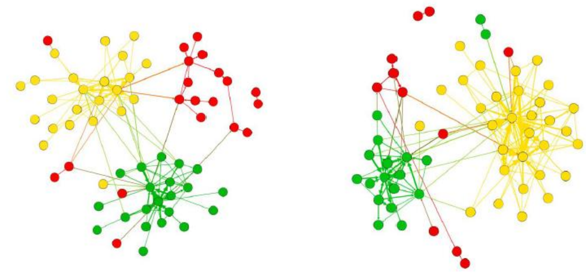

Figure 7. Network Visualisation Topic 0 (Climate)

NetworkX offers a variety of functions for calculating metrics, which are used to determine basic network properties for each topic before and after the election. These include the number of nodes and edges of the respective topics, which provide information about the change in the number of nodes and edges over the observation period and thus the size of the network. Based on the nodes and edges, the connectivity of each topic area is analyzed using the node density metric. This corresponds to the ratio between the number of edges in each network and the theoretical number of edges if the network were fully connected and is expressed in *Stegbauer et al.* [33] using $p = \frac{Actual\ number\ edges}{Expected\ number\ edges}$. The expected number of edges in a loop-free directed network is defined as *N(N-1)*, where *N* is the number of nodes. A node density of one would correspond to a fully interconnected network. The node density is determined across parties at the topic level as well as at the party level for each topic. This allows the analysis of the change in network node density for the entire topic as well as a party-specific breakdown of the change. Node density is a metric that enables the measurement of the connectedness of a network. The density can be used to compare networks with each other and to determine the intensity of social dynamics within a network.

Subsequently, the dyadicity at the party level is determined for each topic network. Dyadicity represents the connectedness between nodes of the same party relative to the standard connectedness of the network. In *Wang et al.* [34] to calculate dyadicity, first, the average connectedness in the network of two nodes is determined with $p = \frac{2M}{N(N-1)}$, where *N* is the number of nodes and *M* is the number of edges. Within a party *i*, the dyadicity is expressed by $D_i = \frac{Actual\ number\ edges}{Expected\ number\ edges}$. The actual number of edges is the number of edges between nodes of the same party. The expected number of edges is determined by $\frac{N_i(N_i-1)}{2}p$, where N_i represents the nodes of a party. A network is dyadic if the nodes belonging to the same group are more connected to each other than in a random network. This is the case for $D > 1$. In addition to the intra-party determination of the dyadicity for each topic, the average of the intra-party dyadicities for each topic is calculated. This provides information about the dyadicity of the topic as a whole.

The determination of the homophily of the network is based on the calculations in *Currarini et al.* [35]. First, the ratio of nodes of a party to the total number of nodes is represented as $w_i = \frac{N_i}{N}$, where i represents the party under consideration. The homophily within a party is calculated with $H_i = \frac{s_i}{s_i + d_i}$, where s_i represents the edges to nodes of the same party and d_i the edges to nodes of a different party. Since this measurement is susceptible to bias and different group sizes, the result is normalized to the internal homophily with $IH_i = \frac{H_i - w_i}{1 - w_i}$. Here the distortion is placed in relation to the maximum possible distortion $1 - w_i$. If $IH_i < 0$, it is a heterophilic network. A homophilic network is defined with $IH_i > 0$. With $\overline{H}_{Total} = \frac{1}{n}\sum_{i=1}^{n} IH_i$, the homophily is then built over the entire network, where n corresponds to the number of parties.

VI. Results of Topic Analysis

Figure 8 shows the number of nodes for each topic and both periods under consideration. It can be observed that the number of nodes for the topic areas is almost similar in both periods. Only for topic 3 (Vaccination), there is a slightly higher deviation in the number of nodes.

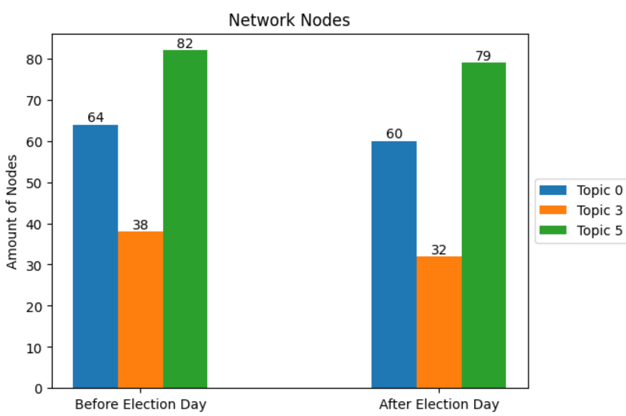

Figure 8. Network Nodes Distribution

Figure 9 shows the number of edges of the selected topics for both periods. The number of edges refers to the actual sum of interactions between accounts. Multiple interactions between the same nodes are also listed as multiple edges. It should be emphasized that the number of edges has only increased for topic 0 (Climate) over the course of the observation period. For all other topic areas, there is a reduction in the number of edges in the second period compared to the first period. The popularity of topics 3 (Vaccination) and 5 (Finance) decreased over time, whereby a lower need for discussion has led to a lower number of edges. For topic 3 (Vaccination), this development can be explained by the course of the pandemic countermeasures.

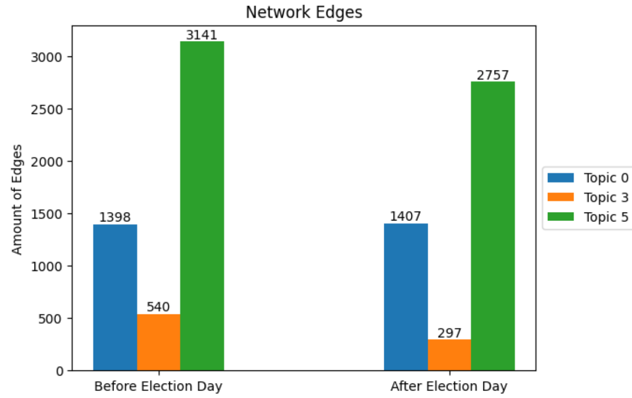

Figure 9. Network Edges Distribution

The results of the network node density analysis are shown in Figure 10. The node density is determined based on a directed network; it is the directed node density of the entire network. For topic 0 (Climate), the network node density increases slightly over time, this topic is examined in more detail. Figure 11 serves as an example of a breakdown of the party node density based on a topic.

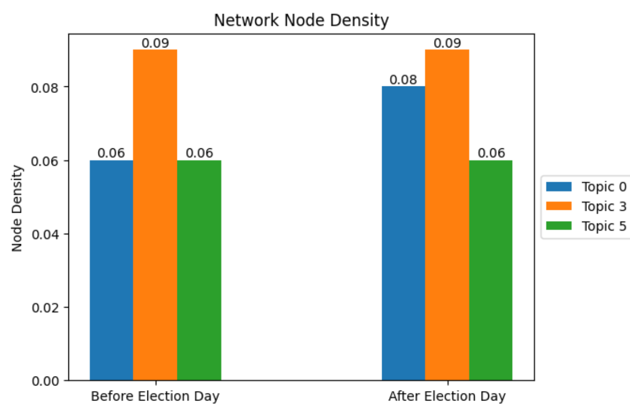

Figure 10. Network Node Density

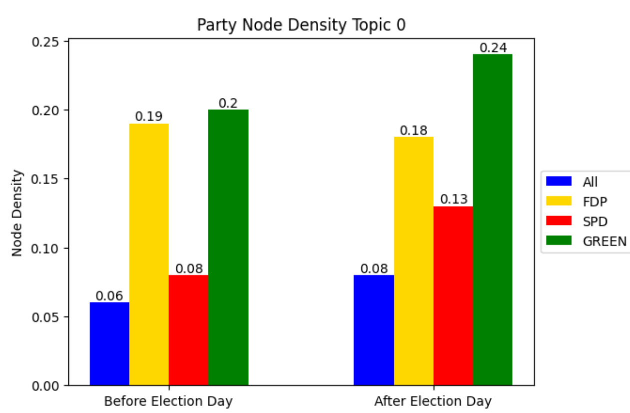

Figure 11. Party Node Density Topic 0 (Climate)

Table IX shows the average value of the node density for each party as well as for the entire network over the period for all topics.

TABLE IX. PARTY NODE DENSITY PARTY AVERAGES

Party	Before Election	After Election	Difference
FDP	0.20	0.19	-0.01
SPD	0.12	0.17	+0.05
Greens	0.20	0.22	+0.02
All	0.07	0.08	+0.01

The results of the analysis of the dyadicity of the parties are explained through the lens of the individual topics. Figure 12 shows the dyadicities of the networks of the individual topics. For topic 3 (Vaccination), the dyadicity increases over the period under observation. For topics 0 (Climate) and 5 (Finance), the dyadicity value decreases slightly. An increase in dyadicity indicates an increase in intra-party communication for the respective topic, while a decrease in dyadicity indicates an increase in inter-party communication.

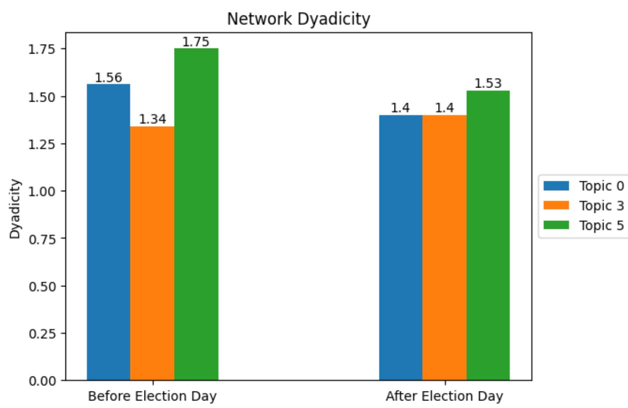

Figure 12. Network Dyadicity Distribution

Table X shows the development of the average dyadicity of the parties across all topics. The dyadicity values of the individual parties lead to the conclusion that the Greens and the FDP cultivate strong intra-party communication, while the SPD prefers inter-party communication. The developments shown in Table X suggest that the SPD has slightly increased its intra-party communication over time, while the FPD and the Greens have developed a slight trend towards inter-party communication.

TABLE X. PARTY DYADICITY PARTY AVERAGES

Party	Before Election	After Election	Difference
FDP	1.85	1.54	-0.31
SPD	0.80	1.08	+0.28
Greens	1.99	1.70	-0.29

Figure 13 shows the analysis of party dyadicity for topic 0 (Climate). Here it can be seen that the dyadicity decreases at the FPD and the Greens. The dyadicity of the SPD increases for the same topic. This suggests that, compared to the previous period, the FDP and the Greens have stronger inter-party communication than the SPD. However, the SPD has strengthened its intra-party communication.

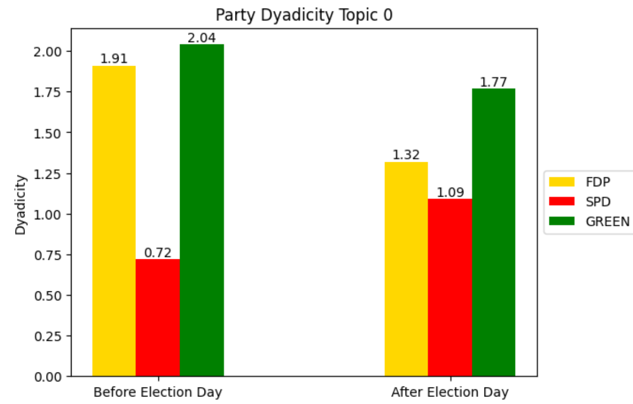

Figure 13. Party Dyadicity Topic 0 (Climate)

Table XI shows the change in the average homophily of the parties for all topics. A decrease in homophily can be observed for all parties. It indicates a reduction in intra-party debate in relation to inter-party communication. This means that the parties developed a trend towards communication between each other rather than communication within their party boundaries.

TABLE XI. INBREED HOMOPHILY PARTY AVERAGES

Party	Before Election	After Election	Difference
FDP	0.85	0.83	-0.02
SPD	0.20	0.07	-0.13
Greens	0.75	0.64	-0.11

Overall, there have been changes in the parties' communication behavior concerning the analyzed topics. The SPD's homophily declines. At both other parties, the changes are small and the homophily remains at a high level, which does not align with the initial expectations. An answer to the research question can therefore be formulated as follows:

The communication behavior of the traffic light parties for the period under consideration concerning the selected topics has not changed as a result of the formation of the government in 2021. From the perspective of a political scientist, this can be attributed to the way compromises are reached in government circles. While compromises are discussed and found in the respective committees, social media platforms such as Twitter usually serve to profile successes, at least among coalition partners. In general, public criticism is avoided among coalition partners and rather expressed towards the opposition.

VII. CONCLUSION

This paper illustrates the application of techniques from social network analysis, sentiment analysis, cluster analysis, and topic analysis in combination to explore communication on social media, especially on microblogs.

H1 is proven, as differences are found for mention and reply interactions. The networks for each interaction type yield differences in both intra- and inter-party interactions, which is shown by the results of the GNA. These findings are statistically significant due to the low p-values. H2 can therefore be considered true. The p-value of the χ^2 test indicates a low likelihood that party affiliation influences the assigned cluster.

H3 cannot be answered clearly. MPs' mutual sentiment changed positively. The FDP's positive change towards the coalition partners SPD and Greens can be considered a sign of a generally improved attitude towards these parties. However, the notable overall increase in positivity across most parties could indicate that the findings of the FDP are not unique. The generally positive attitude between parties after the election can be caused by MPs congratulating one another. A lack of German language sentiment analysis models for short text fragments limits this research. Improved models utilize machine learning techniques and so can comprehend sentiments on a broader level and can also recognize nuances.

Statements about H4 are not reliable. However, while positive tendencies towards an "Ampel" coalition can be shown from both the sentiment analysis and the inter-party and intra-coalition homogeneity, neither can be proven as statistically significant.

Concerning hypothesis H5, the specific metrics do not indicate a change in homophily. No significant change in homophily can be detected in the topic networks depending on the observation period.

Concerning hypothesis H6, the intensity of communication has decreased across all topics. The communication intensity is determined by the number of edges per topic.

Different interactions between MPs can be observed during the pre- and post-election periods and the resulting interaction networks for each period show a difference in intra- and inter-party communication. However, this paper handles political communication only via Twitter. Results are partially transferable to other countries.

Future work may include "The Left" and AfD in these considerations to produce more information. Expanding the evaluated timeframes or continuous monitoring would produce more data. Analyzing follower and friend networks and MPs' liking behavior in combination with the findings of this article would yield insights into differences in parties' mutual relationships around elections.

The change of communication networks within German parties based on political issues could be of interest to several actors. The results of this study could contribute to a better understanding and analysis of the political climate within the "traffic light parties" in the context of political analysis. The results of this study can help to understand the behavior of people in groups and their communication within political parties. This could be of interest to social scientists working on questions of group dynamics and political socialization. Findings from this study could also be useful for communication practitioners by providing insights into the way communication networks develop and change within political parties. This could be of interest to communication professionals involved in the design and management of communication strategies.

References

[1] H. Schuhbauer, S. Schötteler, J. Niu, B. Schiffer, and D. Wolfarth, "A Quantitative Social Network Analysis of Politicians' Tweets to Explore Political Communication". Proceedings of The Twelfth International Conference on Social Media Technologies, Communication, and Informatics (SOTICS 22). Lisbon, Portugal, Oct. 2022.

[2] E. Shearer, "Social Media Outpaces Print Newspapers in the U.S. as a News Source", Pew Research Center, [Online]. Available from https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/, 2018.

[3] A. Watson, "Usage of Social Media as a News Source Worldwide 2021", Statista, [Online]. Available from https://www.statista.com/statistics/718019/social-media-news-source/, 2021.

[4] S. Wasserman and K. Faust, "Social Network Analysis – Methods and Applications", 1st edn. Cambridge University Press, Cambridge, USA, 1994.

[5] A. Virk, "Twitter: The Strength of Weak Ties", University of Auckland Business Review, Vol. 13, No. 1. University of Auckland, Auckland, AUK, NZL, pp. 19-21, Jan. 2011.

[6] M. S. Granovetter, "The Strength of Weak Ties", American Journal of Sociology, Vol. 78, No. 6. The University of Chicago Press, Chicago, IL, USA, pp. 1360-1380, May 1973.

[7] D. S. Lassen and A. R. Brown, "Twitter: The Electoral Connection?", Social Science Computer Review, Vol. 29, No. 4. SAGE Publications, Thousand Oaks, CA, USA, pp. 419-436, Nov. 2011.

DOI: https://doi.org/10.1177%2F0894439310382749

[8] M. Boireau, "Uncovering Online Political Communities of Belgian MPs through Social Network Clustering Analysis", Proceedings of the 2015 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '15). Association for Computing Machinery, New York, NY, USA, pp. 150-163, Nov. 2015. DOI: https://doi.org/10.1145/2846012.2846049.

[9] J. A. Caetano, H. S. Lima, M. F. Santos, and H. T. Marques-Neto, "Using sentiment analysis to define Twitter political users' classes and their homophily during the 2016 American presidential election", Journal of Internet Services and Applications, Vol. 9, Article 18, Sep. 2018. Springer Open, DOI: https://doi.org/10.1186/s13174-018-0089-0.

[10] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods", ACM Computing Surveys, Vol 49, No. 28. Association for Computing Machinery, New York, NY, USA, pp. 1-41, Jun. 2017. DOI: https://doi.org/10.1145/2938640.

[11] A. Boras and S. R. Singh, "Investigating political polarization in India through the lens of Twitter", Springer, 2022. DOI: https://doi.org/10.1007/s13278-022-00939-z.

[12] F. Meier, A. Bazo, and D. Elsweiler, "Using Social Media Data to Analyse Issue Engagement During the 2017 German Federal Election", Commun. ACM 22, 1, 25, pp. 1-25, Feb. 2021. DOI: https://doi.org/10.1145/3467020.

[13] E. Garcia-Sanchez, P. R. Benetti, G. L. Higa, M. C. Alvarez, and E. Gomez-Nieto, "Political discourses, ideologies, and online coalitions in the Brazilian Congress on Twitter during 2019". New Media & Society, 25(5), p.1130-1152, 2021. DOI: https://doi.org/10.1177/14614448211017920.

[14] R. Egger and J.A. Yu, "Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts". Front. Sociol. 2022, 7, 886498. https://doi.org/10.3389/fsoc.2022.886498

[15] Y. A. Jeddi, Scweet (Version 1.6), [Online]. Available from: https://github.com/Altimis/Scweet, Dec. 2021.

[16] Twitter Handle, [Online]. Available from: https://influencermarketinghub.com/glossary/twitter-handle/#:~:text=What%20is%20Twitter%20Handle%3F%20A%20Twitter%20handle%2C%20which,usernames%20on%20the%20social%20network%20are%20the%20same, Dec. 2022.

[17] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "On Modularity Clustering". IEEE Transactions on Knowledge and Data Engineering. 20 (2): 172–188, Feb. 2008. doi:10.1109/TKDE.2007.190689 . S2CID 150684.

[18] A. Evtushenko and J. Kleinberg, "The paradox of second-order homophily in networks", Sci Rep 11, 13360, Jun. 2021. https://doi.org/10.1038/s41598-021-92719-6

[19] The Selenium Project, Selenium (Version 4.1.0), [Online]. Available from: https://github.com/seleniumhq/selenium, Dec. 2021.

[20] GEXF Working Group, GEXF File Format (Version 1.2), The Gephi Community Project, 2009. [Online]. Available from: http://gexf.net, Dec. 2021.

[21] Data work, [Online]. Available from: https://git.informatik.fh-nuernberg.de/wolfarthda82341/sna-germanys-members-of-parliament-on-twitter, Oct. 2022.

[22] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proceedings of the National Academy, Vol. 99, No. 12. National Academy of Sciences of the United States of America, Washington DC, USA, pp. 7821-7826, Jun. 2002.
DOI: https://doi.org/10.1073/pnas.122653799.

[23] S. Loria, TextBlob (Release v0.16.0) Documentation. TextBlob: Simplified Text Processing. [Online]. Available from: https://textblob.readthedocs.io/en/dev/index.html#, Dec. 2021.

[24] M. Killer, textblob-de. (Version 0.4.3), German language support for TextBlob by Steven Loria. [Online]. Available from: https://github.com/markuskiller/textblob-de, Dec. 2019.

[25] E. Quadbeck, "Früh und schmutzig – der Bundestagswahlkampf 2021," RedaktionsNetzwerk Deutschland, [Online]. Available from: https://www.rnd.de/politik/bundestagswahlkampf-2021-haerter-und-schmutziger-MAAJZPIDTFAD7OMUJXUXE2OE5I.html, May 2023.

[26] S. K. Horwitz and I. B. Horwitz, "The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography", Journal of Management, 33(6), pp. 987–1015, 2007. https://doi.org/10.1177/0149206307308587

[27] Twitter Account Deutscher Bundestag, List members [Online]. Available from: https://mobile.twitter.com/i/lists/912241909002833921/members, Nov. 2022

[28] J. Roesslein, "Tweepy Documentation" [Online]. Available from: https://docs.tweepy.org/en/stable/index.html, Jan. 2023.

[29] Deutscher Bundestag, "The Open Graph Viz Platform" [Online]. Available from: https://www.bundestag.de/services/opendata, Jan. 2023

[30] M. Grootendorst, "Bertopic: Neural topic modeling with a class-
based tf-idf procedure". arXiv preprint arXiv:2203.05794, 2022.

[31] D. Newman, Y. Noh, E. Talley, S. Karimi, and T. Baldwin, "Evaluating topic models for digital libraries", Proceedings of the 10th annual joint conference on Digital libraries (JCDL '10). Association for Computing Machinery, New York, NY, USA, pp. 215–224, 2010. https://doi.org/10.1145/1816123.1816156

[32] NetworkX Developers, Software for Complex Networks [Online]. Available from: https://networkx.org/documentation/stable/index.html, Jan. 2023

[33] C. Stegbauer and R. Häußling. Handbuch Netzwerkforschung. Verlag für Sozialwissenschaften, 2010. ISBN: 9783531158082.

[34] X. Wang, O. Varol, and T. Eliassi-Rad, "Information access equality on generative models of complex networks", Appl. Netw. Sci. 7, 54, 2022. https://doi.org/10.1007/s41109-022-00494-8

[35] S. Currarini, M.O. Jackson, and P. Pin, "An Economic Model of Friendship: Homophily, Minorities, and Segregation". Econometrica, 77: pp. 1003-1045, 2009. https://doi.org/10.3982/ECTA7528