**Supplementary information to:**

**Does Relational Polarization Entails Ideological Polarization? The Case of the 2017 Norwegian Election Campaign on Twitter**

This supplementary information gives some more details about the Norwegian context of the study, the data collection and proceeding and the process ideological coding, that for space reasons, are summarized in the paper.

1. **Norwegian context**

In *Patterns of democracy,* Lijphart (1999) makes a distinction between two basic types of democracy: majoritarian and consensus. The majoritarian model of democracy is characterized by a concentration of executive power in one-party government, dominance by the executive branch, a two-party system, majoritarian and disproportional representation, and interest-group pluralism. However, the consensus model of democracy is characterized by executive power-sharing in broad coalition cabinets, a balance of power between executive and legislative branches of government, a multiparty system, proportional representation, and interest group corporatism. The differences characterizing these two types of democratic institutions reasonably can be expected to be reflected in the dynamics of political life, political culture, and the political public sphere. Unlike the U.S., which uses the majoritarian model, Norway, like many European countries, employs the consensus model. Norwegian democracy is characterized by a multi-party system, coalition governments, proportional representation, a political culture of consensus, and national media operating across ideological divisions.

Nine parties were elected to the Parliament in 2017 in Norway. The Norwegian parties represented in Parliament are designated by their English name, Norwegian name, and abbreviation in bold: Labor Party (*Arbeiderpartiet*, **A**);Conservative Party (*Høyre*, **H**); Progress Party (*Fremskritspartiet*, **FRP**); Center Party (*Senterpartiet*, **SP**); Liberal Party (*Venstre,* **V**); Christian Democratic Party (*Kristelig Folkeparti*, **KRF**); Green Party (*Miljøpartiet de Grønne*, **MDG**); Socialist Left Party (**SV**); and Red Party (*Rødt,* **R**). The analysis included, in addition to these main parties, an array of minor parties not represented in Parliament, but rather field candidates in the 2017 election: the Christian Party (KRISTNE), a Christian right party; the Alliance (ALLI), a nationalist party; Democrats in Norway (DEMN), a right-wing populist/nationalist party; Health Party (HELSE), a single-issue party; Coastal Party (KYST), a national conservative party; Pirate Party (PIR), promoting “pirate politics”; and the Capitalist Party (LIBS), a libertarian party. The Labor Party, the Center Party, the Green Party, the Socialist Left Party and the Red Party represent the left side of Norwegian block politics, while the Conservative Party, the Progress Party, the Liberal Party, and the Christian Democratic Party represent the right block. As a result of the 2017 national election, the right block won the election (obtaining a majority in Parliament) an formed a coalition government.

Norway is characterized by high use of the internet, and since the early 2010s social media has been the channel where most people express themselves publicly. Nevertheless, few people use social media to discuss politics and social issues. A population-representative survey fielded in 2020 indicated that 58 per cent of respondents never expressed opinions about society and politics on the internet, while 29 per cent answered that they rarely did so. 7 percent did it monthly, 5 percent weekly and 1 percent daily (Mangset et al., 2022). Barberá and Rivero (2014) have shown that Twitter users who write about politics tend to be male, to live in urban areas, and to have extreme ideological preferences. The same characteristics apply for Twitter users in Norway (Eimjhellen & Ljunggren, 2027), entailing that political communication on Twitter does not represent the opinion in the country and is more polarized than the rest of the citizens (Mangset et al., 2022).

**Data**

While most Twitter studies collect data by querying the Twitter search API via key-words, we adopted another data collection strategy aimed at obtaining the entire universe of political tweets during the election campaign period. The data selection strategy has been aimed at selecting politically engaged and active Twitter users in order to assess the degree of segmentation and polarization characterizing this subset of the citizenry. Below, we summarize the main steps taken to collect the data and create two networks––one representing friend/follower relationships and the other representing mentions and interactions.

Considering that we were interested in mapping the political communication network on Twitter related to the national election, we selected, as a starting point, election candidates who had a Twitter account, assuming that politically engaged Twitter users would follow at least one of these candidates. Thus, we made a list (P) of 1,845 Norwegian political actors with Twitter accounts, comprising all the candidates in the 2017 Norwegian general election who had a Twitter account. By querying the open Twitter API, we made a list (U) of all 833,931 Twitter users who followed one or more of the accounts in P and counted how many of the accounts in P they followed. Considering that estimates of ideology would be unreliable for users following less than three politicians, we selected users to be retained in the analysis based on the following steps: First, we acquired a dump of 4.2 million tweets from the Twitter Historical PowerTrack API, comprising all tweets that: (i) were coded as Norwegian language by Twitter[[1]](#footnote-1); (ii) were posted during the seven months leading up to and including the Norwegian general election in 2017 (March-September 2017); and (iii) were posted by one of the 264,853 accounts in U that followed more than one account in P.

Second, based on this dump and further data about accounts’ followers and friends collected in 2017 after the election using the Twitter API-v1, we selected a sample of accounts that would be the focus of our investigation. We removed accounts that followed less than three of the accounts in P so that we could automate ideological coding of the selected accounts reliably, as explained below.[[2]](#footnote-2) This gave us a set of 179,377 users whom we viewed as having been engaged actively with Norwegian politics on Twitter during the March–September 2017 period. Using criteria similar to Barbera et al. (2015) we removed accounts that had ‘bot-like’ characteristics or appeared to be inactive, i.e., accounts that: (i) sent < 10 or > 2500 tweets in the election period; (ii) had < 25 followers; OR, (iii) followed < 100 accounts. These criteria were retained because while inactive accounts tend to send few tweets and to have few friends and followers, bots tend to tweet a lot. In spite of not guaranteeing that all bots have been eliminated, this approach allows to exclude the most inactive and hyperactive accounts. Further, we removed accounts that followed less than 10 of the politicians accounts, so that we could reliably automate ideological coding of the selected accounts, as explained next. This gave us a set of 11,236 users that we consider to have been actively engaged with Norwegian politics on Twitter in the period March-September 2017.

Considering that we were interested in political communication between the selected users, we needed to be able to classify tweets as political/nonpolitical, in which we adopt a broad definition of “political,” much like “political communication,” but not necessarily containing political content. From this perspective, communication with political actors would be viewed as “political” even if the content was not. A tweet was classified as “political” if: (i) it contained a word, phrase, or hashtag from a pre-compiled list based on Keyness analysis; or (ii) it mentioned, was sent by, or interacted with the account of a political actor (interactions comprised replies, retweets, and quoted tweets).

Keyness (Edmundson & Wyllys, 1961) is a statistic used in computational linguistics that highlights words that are unusually frequent in one set of texts (or corpus) compared with another set of texts (corpus). Here, the comparison was between tweets viewed as containing a term from the initial list of terms or interacting with an account from the initial list of actors, and the set of remaining tweets. Thus, the list of generated key-words was expected to include good candidates for expanding the initial lists. We used the list of comparatively frequent key-words to select the set of political tweets. Compared with the most prevalent way of collecting tweets (using the Twitter search API, which returns tweets containing a key-word or set of key-words), our method allowed us to collect a richer collection of political tweets based on an exhaustive list of key-words.

The lists of political terms and actors were compiled in two steps. First, one of the authors familiar with Norwegian political communication in social media manually compiled a list of 28 words, phrases, and hashtags that defined political topics. The list of the 28 words, phrases and hashtags used for the initial search was the following (in English translation): election, AP, Labor Party, SV, "Socialist Left Party", SP, Center Party Right, FrP, Progress Party, KrF, "Christian People's Party", Liberal Party, Red, MDG, "Erna Solberg" (Prime Minister), "Jonas Gahr Støre" (Leader of the opposition), @Venstre, @arbeiderpartiet, @Partiet, @Hoyre, @KrFNorge, @Senterpartiet, #election2017, #nrkvalg, #election17 #dax18 (News political program), #siste (News political program).

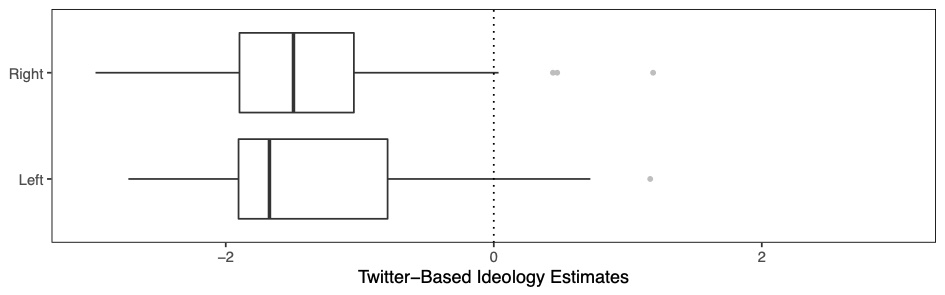
Initially, the list of political actors was taken to comprise the 1,845 accounts from List P above. These lists then were expanded in a semi-automated process, similar to Conover et al. (2011a), using the idea of Keyness analysis (Edmundson & Wyllys, 1961). By scanning the automatically generated list of the most frequent key-words and examining instances of tweets containing the suggested words, phrases, hashtags, and account names, one of the authors added 677 words, phrases, and hashtags to the list of political terms, as well as 249 more accounts viewed as “opinion leaders” to the list of political actors. Using the expanded lists, we filtered our initial set of 4.2 million tweets based on the presence of a political term or interaction with a political actor’s account, resulting in a set of around 1.5 million “political” tweets. The frequency-based nature of the query expansion process means that we are confident that most political tweets were identified, and the use of Keyness analysis helped mitigate researcher bias in the selection of terms and accounts. Finally, we matched the list of political tweets with the list of users following at least three political actors on List P.

1. **Estimation of ideology**

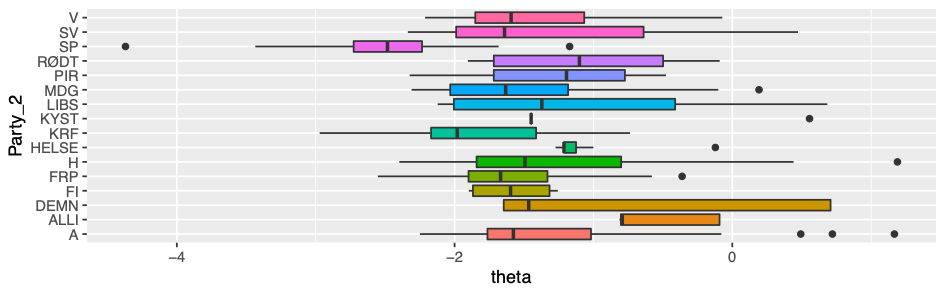
*Ideology estimation: The Barberá (2015) method*

Following Barberá’s (2015) method, we identified the ideological latent space and estimated ideological ideal points for 179,377 political Twitter users—including 1,756 political accounts (candidates) for which we have collected followers. Figure 1 displays the distribution of Twitter-based ideology estimates of candidates on the ideological spectrum based on their positions on the left-right political spectrum, computed on the basis of their party membership, while Figure 2 displays this distribution based on their party affiliation. With this estimation method, the average ideal points for right-wing and left-wing candidates are relatively close to each other, and distribution based on party does not reflect the Norwegian parties’ positions on the left-right ideological spectrum, with, e.g., the Center Party (SP), a party situated at the center-left of Norwegian politics, being more leftist than the radical-left party Rødt. To sum up, this method does not function well with our data, possibly because, contrary to Barberá et al. (2015), we did not initially restrict our matrix of connections to a subset of accounts with high ideological discrimination, but rather estimated ideological ideal points based on the set comprising all candidates and their followers.

***Figure 1: Twitter-based ideology estimates of candidates: the Barberá method***



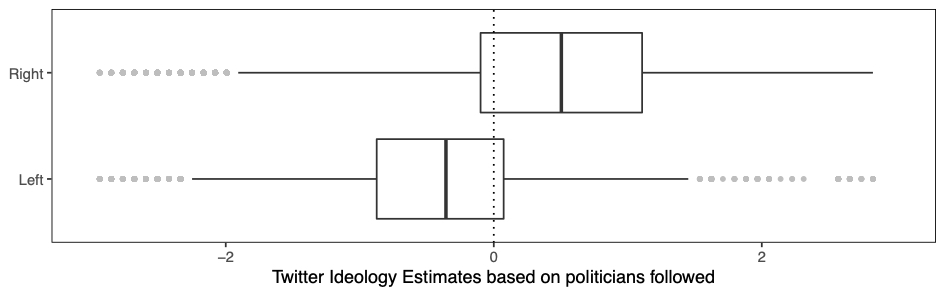
***Figure 2: Twitter-based ideology estimates of candidates by party: Barberá method***



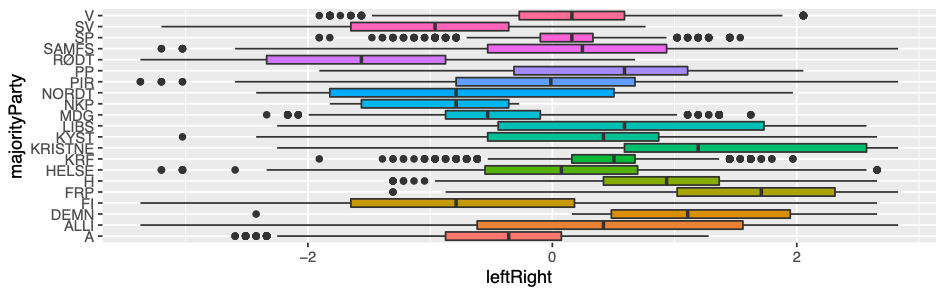
*Ideology estimation - Halberstam and Knight (2016) method*

Furthermore, following Halberstam and Knight’s (2016) method, we coded our selected users for party ideology, a discrete class (one of the parties presenting candidates to the Norwegian election), and for ideology left-right, a scalar value (between 0 and 10) normalized for the analysis. The party ideology variable is calculated as the most common party of the political actors (from List P above) that the user follows. The ideology left-right variable is computed as the mean average of the values for the parties of the political actors that the user follows. The values used to position the parties on the left-right spectrum are based on the averaged self-identification of the candidates for each party during the 2017 national parliamentary election based on a candidate survey realized by Hesstvedt (2017). As is the case with the estimates based on the Barberà (2015) method, the ideological points have been normalized. As noted previously, Figure 3 provides the distribution of Twitter-based ideology estimates of candidates on the ideological spectrum based on their position on the left-right political spectrum computed on the basis of their party membership, while Figure 4 displays this distribution based on their party membership. This estimation method provides average ideological points for candidates belonging to the right and left sides of the political spectrum, respectively, that are more distant from each other than with the ideal point estimation method. Furthermore, the distribution of ideological estimates, according to the parties, reflects, to a greater extent, party positions on the left-right spectrum characterizing Norwegian politics, with the left-wing parties being positioned correctly on the left and the right-wing parties on the right.

***Figure 3: Twitter-based ideology estimates of candidates: Halberstam and Knight method***



***Figure 4: Twitter-based ideology estimates of candidates by party: Halberstam and Knight method***



Given these results, we will use ideology estimates based on Halberstam and Knight (2016) to analyze Twitter networks.

**References**

Barberá, P., & Rivero, G. (2014). Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review, 33*(6), 712-729. doi:10.1177/0894439314558836

Barbera, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science, 26*(10), 1531-1542. doi:10.1177/0956797615594620

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). *Political Polarization on Twitter*: In Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM), 2011. ICWSM Test of Time Award 2021.

Edmundson, H. P., & Wyllys, R. E. (1961). Automatic abstracting and indexing—survey and recommendations. *Commun. ACM, 4*(5), 226–234. doi:10.1145/366532.366545

Eimhjellen, Ivar, og Jørn Ljunggren. *Kollektiv handling i digitale medier. Nye digitale skiller?*. Rapport 3, 2017. Oslo: Senter for forskning på sivilsamfunn og frivillig sektor.

Halberstam, Y., & Knight, B. G. (2014). Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter. *NBER Working Paper Series*.

Hesstvedt, S., Karlsen, R. . (2017). *Stortingskandidater. Rapport fra Stortingskandidaterundersøkelsen 2013*. Retrieved from Oslo: <https://samfunnsforskning.brage.unit.no/samfunnsforskning-xmlui/handle/11250/2450099>

Lijphart, A. (1999). *Patterns of democracy. Government forms and performances in thirty-six countries*. New Haven: Yale University Press.

Mangset, M., Midtbøen, A., & Thorbjørnsrud, K. (2022). Ytringsfrihet i en ny offentlighet. Universitetsforlaget. <https://www.idunn.no/doi/book/10.18261/9788215051017-2022>

1. We manually checked a sample of tweets in order to ensure that the Twitter API did not misclassified tweets’ language. [↑](#footnote-ref-1)
2. Of the 833,931 accounts that followed at least one of the political actors in P: 264,853 followed > 1; 43,659 followed > 10; and 1,238 followed > 100. The actual dump that we received from the Twitter Historical PowerTrack API contained c. 25 million tweets. We then filtered these to get the 4.2 million tweets that Twitter coded as “Norwegian language.” After skimming the list of accounts that were removed for being “bot-like,” it appears that most of these were removed based on the criterion of having posted < 10 tweets during the seven-month period, but some accounts elicited thousands of tweets and very few friends and followers. We tested and validated the estimation of ideology following the Halberstam and Knight (2016) method for different thresholds of followed political actors (ranging from 1 to 10) and concluded that following at least three political actors is the threshold for providing valid ideological coding. [↑](#footnote-ref-2)