Claiming the (green) agenda:

Exploring how Danish MPs and parties set their climate agenda on Social Media

Character Count (incl. spaces): 59135

ASDSII/Digital Methods: Fri-753376-3

Members: 11/16, 32/17, 34/28, 9/38

June 17, 2021

Contents

1 Introduction (All)

Within the last few years the debate surrounding climate change and appropriate measures to stop and/or mitigate its consequences has developed into being of significant importance in Danish politics (Øyen, 2019; Ladrech & Little, 2019). As is required by the Paris Agreement, Denmark has agreed to reduce the emission of greenhouse gases, a process that arguably requires a transition from current ways of production, consumption, investment and resource exploitation to more sustainable infrastructures (Blau, 2017). The fundamental challenge of climate change politics is perhaps best defined in the Brundtland report, namely how to ensure that societies develop in a manner *"... that meets the needs of the present without compromising the ability of future generations to meet their own needs"* (Brundtland, 1987). In a representative democracy, such a fundamental transition and defining its substance inevitably sparks debates, political disputes and agenda claiming conducts by political parties and their representatives, as public opinion and international agreements place it on the political agenda (Ladrech & Little, 2019). There is ample evidence demonstrating that climate policy preferences align with left-right divisions in social and economic preferences (Ladrech & Little, 2019), and in the 2020s, digital platforms allow for communicating such preferences at a new scale. In Denmark, Twitter in particular is used for political agenda-setting (Blach-Ørsten, Eberholst, & Burkal, 2017). Thus, the issue of climate change and the use of digital platforms to communicate politics offer new ways of social-scientific inquiry, as the agenda setting leaves digital traces for researchers to examine. Indeed, recent research into public policy and agenda setting emphasize the increasing importance of studying political communication on social media (Blok, 2020-working paper; Kaiser, Rhomberg, Maireder, & Schlögl, 2016). Through applying an integration of qualitative and quantitative methods, we set out to examine the climate

change topical substance and its distribution among Danish political parties on social media, understood through the theoretical lense of agenda-setting and issue-competition (Green-Pedersen & Mortensen, 2010).

1.1 Research Question (11/16)

In a study of issue-competition among Danish political parties, Green-Pedersen and Mortensen (2010) define an agenda as a hierarchy of issues that relevant actors have to pay attention to. Thus, an agenda operates as a constraining structure by limiting possibilities to address additional issues. The authors further introduce the concept of the *party-setting agenda*, which emphasizes the reciprocal dimension of party-political issue-competition. This implies that political parties alternately communicate their views on issues on the agenda to influence the composition and the substance of these issues (Green-Pedersen & Mortensen, 2010). Consequently, the parties in government are more responsive to this agenda, as their position implies an expectation about ability to provide policies. On the contrary, parties of the opposition are less obligated to address all issues and are free to criticize the government (ibid.). According to the constraining nature of political agendas, climate issues are something that all parties have to address at least to some degree because of its prevalence on the political agenda. This leads us to the following research question and two enclosing sub-questions:

How do political parties and their MPs communicate their political agenda in relation to climate change on social media?

To concretize our research question, we ask two sub-questions that explore specific parts of the climate-oriented communication. Previous research suggests that the Danish parties have widely different approaches of communicating is-

sues on social media, both in terms of what is emphasized, and how this is done (Blach-Ørsten et al., 2017). We are interested in whether this is also the case for the discussion surrounding climate issues. We aim to examine this in our study by taking departure in the following sub-questions, where the first concerns the content of the communication:

1. What are the topics in the Danish climate debate on Twitter and Facebook, and how do the parties differ topic-wise?

The second sub-question examines the aspect of social media practice related to communication of these sub-issues, as previous research has shown that the platform-specific affordances shape the communication (Housley, Webb, Edwards, Procter, & Jirotka, 2017):

2. How do Danish MPs and parties use social media to set their agenda?

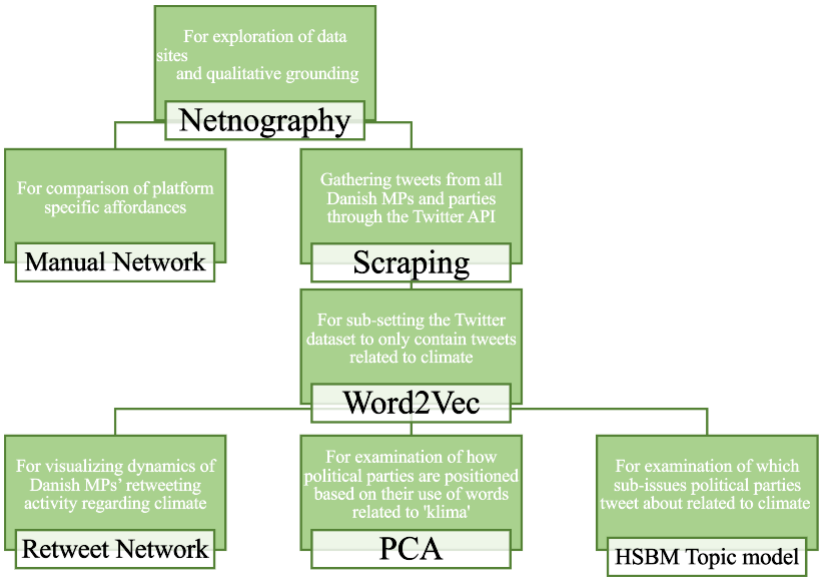
The two posed sub-questions are interrelated as they cover dimensions that put together are essential to consider to understand the agenda setting dynamics related to the key issue of climate-related communication.

2 Research design (32/17)

For answering our research questions, we employ a mixed methods research design. The qualitative methods involve netnography and manual networks, which have proven suitable for analyses of social media (Kozinets, 2020; Decuypere, 2020). The quantitative methods applied are a retweet network for exploration of interaction, and unsupervised learning methods namely PCA, HSBM, and Word2Vec for exploration of textual data. The methods and the process are visualised in Figure 1.

Inspired by Carlsen and Raulund (2021-working paper)'s computer assisted

Figure 1: Research design



learning model (CALM), we aim at a design that involves the strengths of both methods in terms of human interpretation and computer-driven methods with the goal of mitigating the biases of both. In doing so, a meta-reflection is in order: Using digital traces from online platforms calls for an awareness of how the data is produced through a business-model logic rather than for the purpose of research. We carry this in mind throughout our project, and will discuss this further in section 12. Our research design entails an abductive approach in which we switch between different methods as we gain new insights for further analysis in a process we will discuss further in section 11.

We find that the parties use different strategies to claim the agenda of climate, and that their communication differ across the political spectrum, where some parties position themselves around traditional left-right party lines, others reveal counter-intuitive ways of political positioning in regards to climate politics as shown in section 10. We also find that political negotiations extend into

the digital realm, and that the parties differ both in how they communicate, in the content of their communication in section 10, and in their interactions, as shown in sections 4 and 5. Through this comprehensive combination of methods we argue that we find rather robust findings regarding the agenda-setting behaviour of the MPs and parties. We will discuss these findings as we present the different methods and the particular insights they provide throughout the paper.

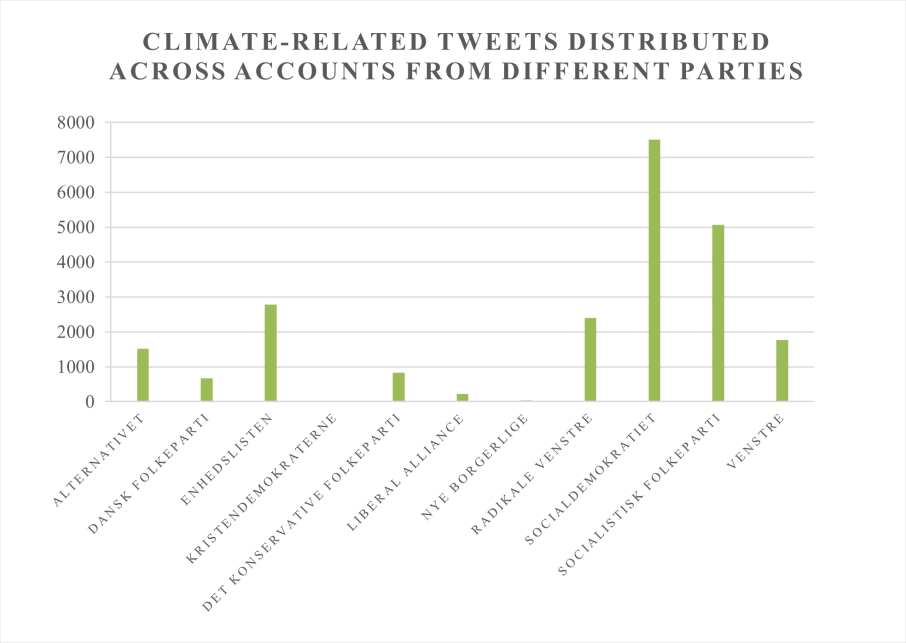
3 Sampling on Twitter (34/28)

In a Scandinavian context, Twitter has proven to be widely used among politicians for agenda setting efforts (Blach-Ørsten et al., 2017), making Twitter a suitable data site to gain insights about practices of agenda setting among the digital political public. Through this project we are interested in the whole population of Danish parties and MPs, but are at the same time aware that we will only retrieve information from the actors who have an active Twitter account (out of 179 MPs and 11 parties, 166 MPs and 10 parties have a Twitter account) (Appendix 1). We are therefore working with what Rafail (2017) defines a bounded population, restricted by a specific topic and selected accounts. We have retrieved all content the actors have produced on Twitter since 1st of January 2018 which was the year before the last election in Denmark. Hence, we are able to capture topics in the communication before, during and after the election period, arguably a dynamic time in politics which can offer fruitful and relevant data for our project. In addition, the election of 2019 was widely coined 'the climate election' (Davidsen-Nielsen, 2019), which we argue increases the possibility of finding comprehensive data relevant for our research question.

In order to retrieve the relevant data, we apply computational methods to subset tweets related to the climate debate. Through this method, we end up

with a sample of 31.243 climate-related tweets. In order to validate these tweets qualitatively, we draw a random sample containing 100 of the tweets. Two coders independently validate the tweets, to asses whether the computational methods provides us with representative tweets. We thereby end up with a precision score of 87.5% and an inter rater reliability of 97%. (Appendix 2).

Figure 2: Climate related tweets distributed across MPs and parties



From Figure 2 we see that there are large party-wise differences in the frequency of green tweets produced. While Socialdemokratiet, which is by far the largest party, are the most productive, right wing parties produce fewer green tweets. The parties with the largest proportion of green tweets are from the left wing.

4 Netnography (9/38)

As we aim to understand our field and our actors, we use netnography as it allows us a deep and qualitative understanding of the field (Kozinets, 2020). In addition, this knowledge serves to ground our automated methods as it adds to the ground truth, the empirically assessed data, which in turn allows us to qualitatively validate our automated methods (ibid.).

Netnography is a qualitative research method that is characterized by its cultural focus in that it seeks to understand and interpret meaning in a focal site. Another characteristic is that the method has a focus on social media data, as it is an explicit method for online spaces (Kozinets, 2020). Furthermore, it entails an immersive engagement with data, as the researcher is actively immersed in the datasite, engaging intellectually with the content in the datasites (ibid.). Lastly, it follows a set of traditional ethnographic procedures which are customized for the digital realm, coined *netnographic praxis* (ibid.).

The steps taken through our netnographic campaign align with the five steps suggested by Kozinets (2020) for netnographic collection of social media data, although with some deviations. These steps include *simplifying, searching, scouting, selecting and saving* (Kozinets, 2020):

4.1 Simplifying and searching: putting our research question into searchable terms (11/16)

Initiating our netnographic campaign, as a first step we simplify our research question into searchable terms. As the research question explicitly states that we are interested in how politicians communicate climate related issues, our search queries were defined by default, and we could start searching for the politicians by their names on Twitter and Facebook (Kozinets, 2020). As the Danish parliamentary system consists of 179 MPs, we define a subset of these

including the following actors: the party leaders, the climate spokespersons, and the official party accounts. We believe that these are actors that actively contribute to the agenda-setting on climate related issues, given their positions.

4.2 Scouting for data (32/17)

Having simplified and searched, we start scouting (Kozinets, 2020). We scroll through the timelines of our actors on Twitter and Facebook, through a process that proves to be an iterative, inductive and explorative one. As we immerse into our datasites, we notice that a 'one size fits all' way of information extraction from all our actors is not feasible due to vast differences in the behaviour: some tweet often and some almost never. For some actors, tweets regarding climate are plentiful; for others, many months pass with no such observation. We therefore scroll until we can qualitatively grasp the notion of the climate issue agenda for each party. These findings we will elaborate in section 4.4 and 4.5.

4.3 Saving and organizing our immersion journal (34/28)

We save the relevant data in our immersion journal, a comprehensive document in which we record observations and reflections from our datasite. We organize the immersion journal as follows. In addition to the netnographers' name, name of actor and date, and datasite, we have the following fields: focus, observation, and reflection. In the focus column we note observations about the general focus of the actors' timeline, and which themes are prevalent regarding climate. In the observation column we add one or more selected example(s) that we based on our scouting find representative for their communicated climate agenda. The reflection column includes thoughts and reflections that occur as we are scouting, noting our intellectual engagement with the content (Appendix 3) (Kozinets,

2020).

After having collected the data, we systematically code our observations from the immersion journal. We start by individually doing an open and explorative coding guided by the focus from our research question. This is in line with the coding process suggested by MacQueen, McLellan, Kay, and Milstein (1998), who furthermore suggests that each coder's proposed codes are circulated in order for the group to define definitive codes and develop a structured codebook. We followed this approach and after developing a joined codebook, we tested it on a sample of our observations and adjusted until we ended up with a list of the seven following codes, which are thoroughly described in our codebook (Appendix 4). Five of the codes covers topical related aspects of social media content, while the last two cover how the platforms are used in communication: (1) International perspectives on climate, (2) Victims in climate politics, (3) Sustainable transition of agriculture, (4) climate in economy and the corporate sector, (5) Sustainable transition of the transport sector, (6) Social media practice, (7) rhetoric. After developing a structured codebook, we manually do a closed coding of all our observations from our immersion journal. By following this approach, we have made a foundation for doing in-depth qualitative analyses (MacQueen et al., 1998).

4.4 Topics (9/38)

Through our immersion journal we get a sense of the topics related to the climate debate and their distributions across the political sphere, offering insights that help us answer subquestion one. For example, we observe that the right wing (Venstre, De Konservative) and the Social Democrats focus on economy and the corporate sector. We also note that most parties at some point are defining 'victims' of climate politics, and that these 'victims' are different across the

political spectrum. For example, Dansk Folkeparti expresses that car-free Sundays is politics at the expense of people in the provinces. Radikale Venstre and Alternativet emphasize the generational (in)justice regarding the consequences of climate change, while members of Dansk Folkeparti are tweeting that some climate change mitigation measures such as "rewilding" are harmful to animals and leads them to starve and suffer as visualised in Figure 3.

Figure 3: Dansk Folkeparti on Twitter

Interestingly, we see that some topics become focal points in the communication around climate, whereas other topics do not receive the same attention and are primarily addressed by a few, and mostly the same parties. For example, the proposal by the social democratic government of a new reform for Danish agriculture in April 2021 engages both actors from the left-wing and right-wing, even the smaller party Nye Borgerlige, which is not usually active in the debate around climate issues. Understood through the concept of issue competition (Green-Pedersen & Mortensen, 2010), this topic seems to be an issue that all the parties claim.

4.5 Social Media Practice (11/16)

Throughout the netnographic campaign we have had a strong focus on the form of communication used on Twitter and Facebook including social media practice of the MPs and parties. This is one of the strengths of netnography; through

immersing into deep data (Kozinets, 2020), we get a sense of *how* the content is communicated, and not just *what* the content is, which is why the method is especially suitable for answering subquestion two.

A key aspect of social media practice for Danish MPs and parties on Twitter is sharing posts from other members of one's party or from news media through retweeting. The actors often use articles from news media as context for a political statement that they want to share, or they show support to fellow party members by retweeting their posts. It seems that the left wing is retweeting more from green NGOs and different activist groups compared to the right-wing. In addition, some accounts, such as Socialistisk Folkeparti, is based almost solely on retweets from its eminent party members. The tendency related to retweeting will be further elaborated in the following section, where we present a network of retweets among the MP accounts of content related to climate.

We also note that the parties differ in their communication with their audience. Dansk Folkeparti, for example, directly addresses the audience through asking questions about opinions and encourage people to "share" their content. Socialdemokratiet rely extensively on video material in their communication with their audience on both platforms. We further observe that it is common for several of the left-wing parties to present visual material showing their support toward green transition. An example could be how Torsten Gejl, climate spokesman from Alternativet, is communicating opinions through visual means as shown in Figure 4.

A key tendency among the left-wing parties and Venstre are their use of Facebook and Twitter to bring forward their role in realizing green transition related initiatives. This way of communicating political victories on social media is especially prevalent for Socialdemokratiet, which are forming government, who continuously emphasize how their initiatives contribute to make Denmark

an international frontrunner in green transition. Common for the remaining parties is the tendency to criticize the government. In addition, we observe that the MPs seem to interact more frequently on Twitter compared to Facebook by using retweets and through commenting. This social media practice will be further explored in the next sections, where we introduce visual networks.

Figure 4: Twitter account of climate spokesman from Alternativet's Torsten Gejl

5 Visual networks (32/17)

The following section will delve deeper into the question of how Danish MPs and parties use social media to communicate about climate related issues. We will address this question by the use of visual networks. As Brooker, Barnett, and Cribbin (2016) assert, a visual analytical approach is suited to explore the construction of a social phenomenon played out on social media. Importantly, this approach acknowledges that we can only understand the phenomenon we aim to uncover when taking the socio-technical assemblage that shapes how the digital public interacts, into account (Brooker et al., 2016). Deriving from our netnographic knowledge we firstly deploy a manual network that visualises how Danish MPs and parties use Facebook and Twitter to interact with each other in

order to explore how their digital public engagement is constituted (Decuypere, 2020). Secondly, from the insights gained from the manual networks we further explore the structure in the Danish MPs and parties' communication around climate issues through an automated retweet network.

5.1 Manual network (34/28)

Through the construction of a manual network we can zoom in on specific situations where online interactional practices unfold (Decuypere, 2020; Campagnolo, 2020; Lai, Pagh, & Zeng, 2019). Our manual network is inspired by Decuypere (2020)s visual network analyses (VNA). Manual networks are dependent on structured qualitative data, which allows us to iteratively locate and record context of relations across data sites and actor types. Consequently, the manual network is not restricted to the specific limited attributes available as are automated networks (Campagnolo, 2020; Lai et al., 2019). Decuypere (2020) describes VNA as a method for visualization and interpretation of qualitative relational data and emphasizes that it is a tool that contributes to an understanding of a certain phenomenon rather than a representation of the phenomenon (ibid.). Thus, manual networks visually emphasize that data and knowledge is constructed and not simply given (Campagnolo, 2020).

5.1.1 Comparing data sites (9/38)

We use manual networks for analyzing interactions of Danish MPs in relation to a specific case, namely the agricultural reform that the government proposed in April 2021. We find that this case provides a good example of social media interactions among Danish MPs as we derive from the immersion journal that all the MPs across party affiliation participate in this discussion. Thus, all of the actors seem to claim the issue (Green-Pedersen & Mortensen, 2010).Furthermore, the

case reflects a persistent debate of making agriculture more climate-friendly and the costs and benefits of doing that. We find this a prevalent sub-issue of the climate debate. As we aim to investigate how Danish MPs and parties use social media to set their agenda, a comparison of platforms further lets us reflect on how platform-specific affordances facilitate the interaction. As suggested by Lai et al. (2019) we made a cross-platform study and followed the interaction on Facebook and Twitter to compare the MPs' use of the two platforms. Following the sociomaterial approach of Decuypere we consider both human and nonhuman entities in interactions (Decuypere, 2020). Thus, the agriculture proposal is considered an actor that the MPs can interact with through the contents of their posts on social media. The interactions between MPs are repostings, retweets, comments, posts or tweets. The interactions in the networks are qualitatively classified as either supportive or unsupportive, based on the text in the particular tweet or retweet. We interpret retweeting (twitter) or sharing a post (Facebook) without adding text as an endorsement (Housley et al., 2017).

From the manual networks depicted below, we see several interesting tendencies. Firstly, we notice that the MPs endorse members of their own party, while all oppositional parties are characterized by unsupportive interactions towards the government. From doing netnography, we have encountered such unsupportive interactions in form of critique to be very common toward the party forming government from both left-wing and the right-wing.

In both Figure 5 and Figure 6 we can identify regions of actors sharing similar opinions about the central actor, Landbrugsreformen. The right-wing parties express that the reform will harm the industry, and this opinion is more represented on Facebook than on Twitter. This might suggest a political divide in platform-use, where the right wing is more active on Facebook than they are on Twitter. Our netnographic observations support this notion. Secondly,

Figure 5: Manual network of Facebook

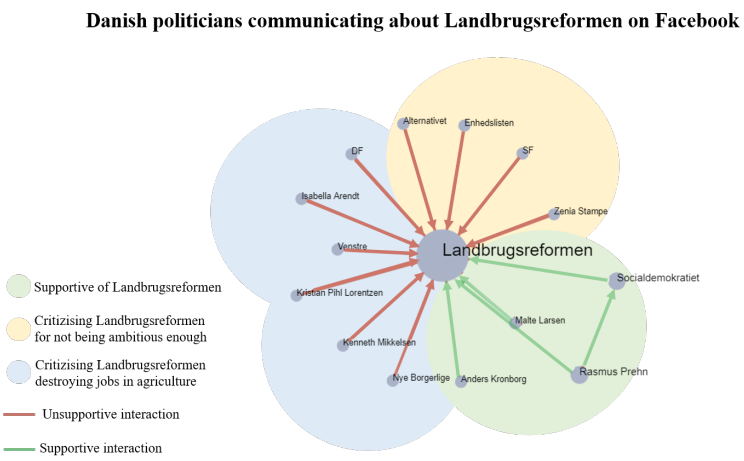
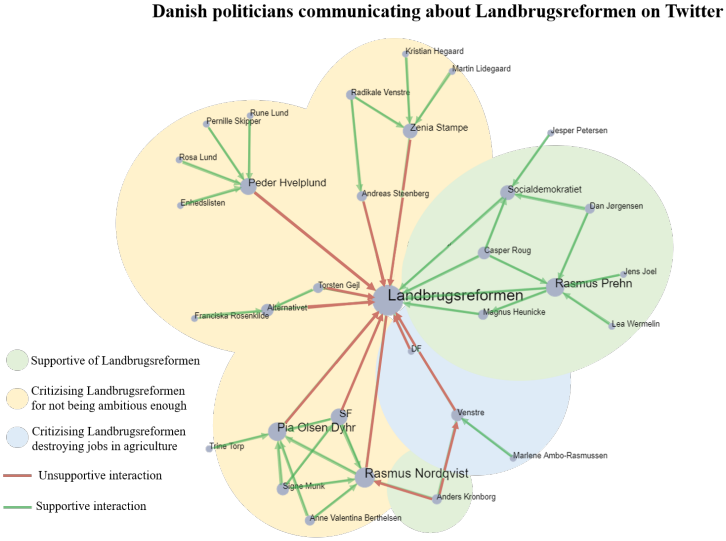


Figure 6: Manual network of Twitter



the manual networks indicate that there is more interaction among the MPs on Twitter than on Facebook. Interestingly, we notice little interaction *across* political affiliation, but a high degree of interaction between MPs of the same parties. In parties from the opposition we see that most parties have one or two central actors that serve as a focal point for the other party members. These central actors tweet about the agricultural reform, whereas the other party members retweet them to express endorsement of the tweet. As a consequence, there are not several different tweets surrounding the agricultural reform for each party, rather MPs from the same parties collectively advance the same opinion by retweeting. Hence, we see that a platform specific affordance of Twitter, such as retweeting, serves as an important factor in shaping the interaction between politicians. The manual networks have enabled us to explore such kind of platform-specific affordances, because we are able to compare with the interactional patterns of Facebook, where we see much less interaction between politicians.

5.2 Automated retweet network (11/16)

As evident from the manual network retweets serve as an important interactional tool for Danish MPs on Twitter, and are thus means of shaping the communication around a certain climate related issue. We find examining this tendency thoroughly a crucial point in understanding the agenda setting dynamics related to the key issue of the climate debate. Therefore, we use an automated retweet network that allows for more actors to understand the general interaction patterns on Twitter for our population.

The retweet network is created by scraping all Danish MPs and party profiles' retweets subsetted on tweets surrounding the climate debate (Appendix 5). Each node represents an individual MP or political party and is colored by

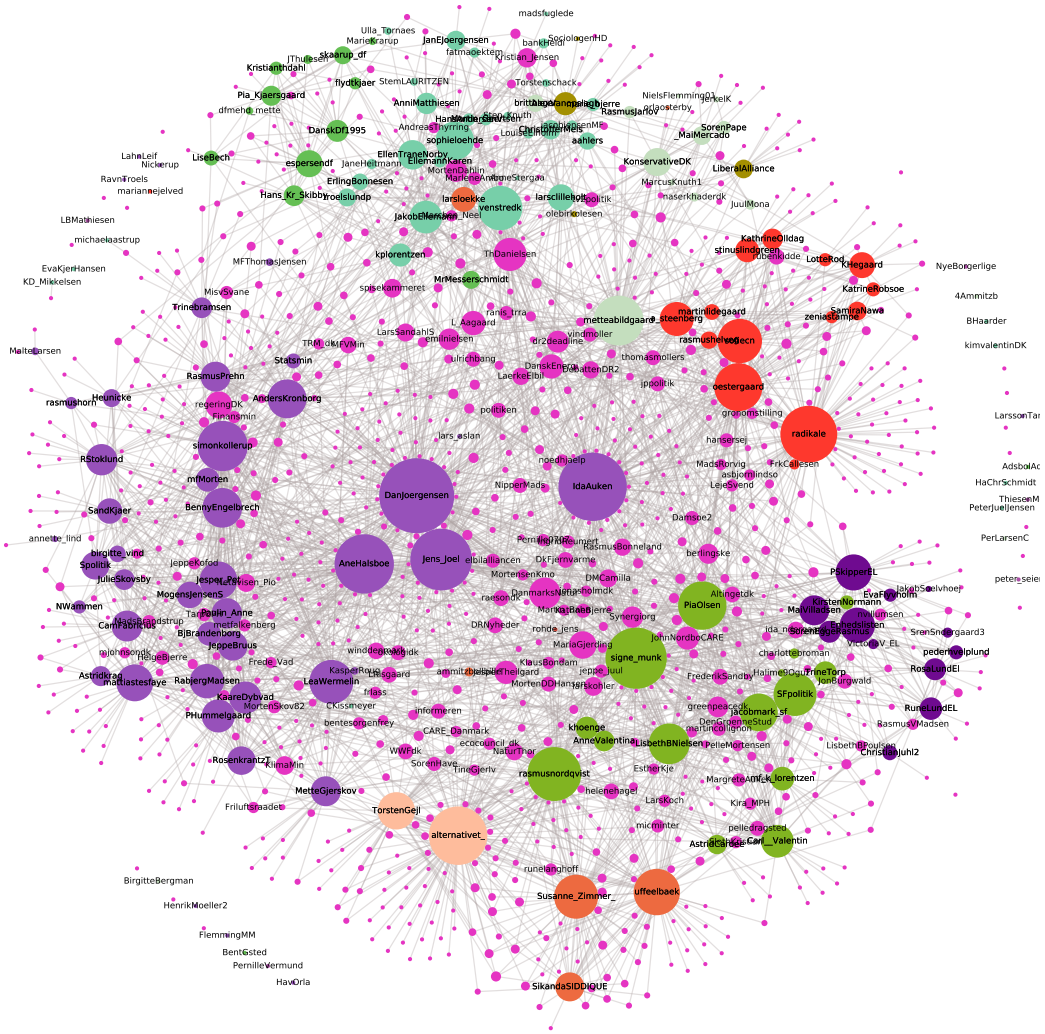
party affiliation, except for nodes colored in magenta, which represent other actors in the communication, such as NGOs, media or laymen. The size of a node is determined by the degree centrality, i.e., the number of edges the node has (Wasserman, 1994). Therefore, if a MPs node is large it indicates a high activity in the communication around climate- either by being retweeted or retweeting other actors.

5.2.1 Findings and analysis (32/17)

From Figure 7 it is evident that MPs from the left-wing constitute a larger share of the network and have larger nodes (indicating higher activity and connection) than parties from the right-wing. The most right-oriented MPs and parties, such as Nye Borgerlige, Dansk Folkeparti and Liberal Alliance are on the periphery of the network with nodes of small sizes. This finding is further validated by our observations from the netnography and descriptive of our data (Figure 2), namely that the left-wing are more active and present in the debate around climate on Twitter. However, as shown from the manual network, this is partly explained by right-wing parties using Facebook rather than Twitter as their main platform of communication.

We will analyse the form of the network through Decuypere (2020)'s analytical concepts of regions, infrastructure and centers. Firstly, it is noticeable that politicians from the same parties tend to cluster in the same regions and additionally that there is a higher order cluster of the left-wing in the bottom and right-wing in the top. As previously noted we interpret retweets without additional text as a proxy for endorsement of what has been posted. Therefore we draw the conclusion that MPs are more likely to endorse tweets from members of their own wing, and especially members of their own party. This tendency of politicians being self-referential in parties is in line with one of the main findings in (Kaiser et al., 2016)'s study on hyperlink networks in Germany.

Figure 7: Automated retweet network of MP's and parties surrounding green transition



Secondly, turning to the infrastructure of the network we observe that the left-wing parties are more likely than the right wing to retweet actors outside the political field, where especially green NGO's, such as GreenPeace, WWF and Danmarks Naturfredningsforening are prevalent. This tendency was also evident from our netnography. Of the parties on the right-wing, Venstre retweets most actors outside the political field, but is rather retweeting industrial associations, such as DKfjernvarme (Danish district heating) and Spisekammeret (agriculture and food). Hence, the infrastructure of the network indicates that MPs integrate different stakeholders in the agenda-setting around climate change.

Thirdly, looking at the center of the network we see that especially four MPs (Dan Jørgensen, Jens Joel, Ane Halsboe, and Ida Auken) from Socialdemokratiet are situated centrally in the left-wing region of the network. This implies that they are related to several other actors in the network, either by retweeting other MPs themselves or being retweeted by other MPs. The MPs in the government are centrally located because they have a high degree-centrality, i.e. many connections. Their position as governing party (S) might explain this, as they have to be responsive to the systemic agenda of climate issues, and it is in their interest as a governing party to influence the composition and the substance of these issues Green-Pedersen and Mortensen (2010).

In sum, the visual networks have allowed us to grasp the interactional practices between the actors, but also illustrated how they act differently on different platforms. We find that the platforms are used in different ways across the political spectrum, which could help us answer subquestion two. However, a crucial point and maybe weakness in these methods is that we only get a "picture" of the behaviour; we can't really say if it is due to the platform specific affordances or the agency of the actors.

6 Content analysis (34/28)

Another way to investigate the relationships between our actors and the topics in their climate related tweets, is through content analysis. We therefore apply principle component analysis as it allows us to investigate how the actors are positioned in relation to each other, but also in relation to frequently used words surrounding the word "klima" (climate). In the following we account for our practice and findings.

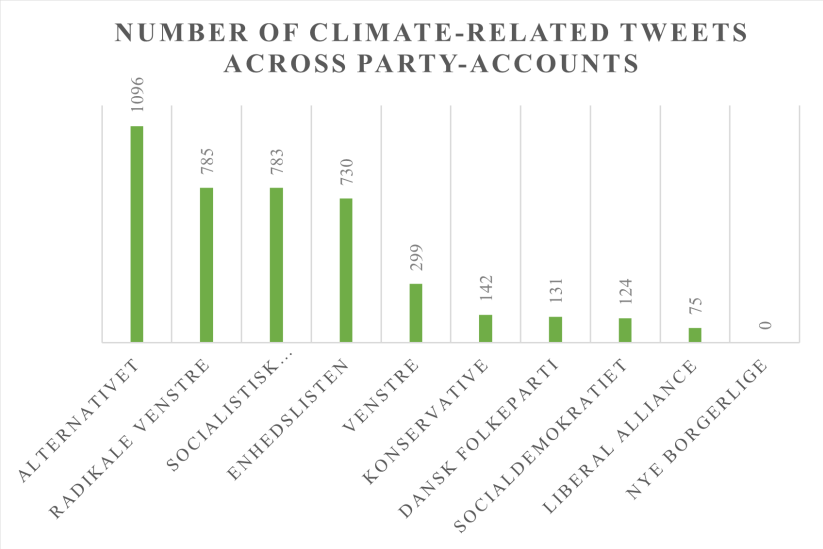
6.1 Subsampling based on party accounts (9/38)

The following principal component analysis is performed based on a subset of our initial dataset containing tweets related to the climate debate. These tweets are retrieved based on an automated Word2Vec model combined with our qualitative assessment, which is further described in section 9. Before conducting our principal component analysis, different pre-processing steps were performed as described in section 8. In relation to this specific part of the project, it was further decided to remove words with less than 4 letters, as they appeared in our initial analysis, without carrying any contextual meaning.

As we are primarily interested in investigating possible differences or similarities between Danish parties, we decide to narrow our dataset further down to a subset of the parties' main accounts on Twitter. With this strategy we aim to simplify the inputs included in the analysis. We can do this confidently because our netnography and network analysis show that the tweets of MPs align topically with the tweets by the official party accounts. During our netnography and network analysis, it was further evident that there is a significant difference in the frequency the different parties and MPs tweet about climate. While topics related to the climate debate are the main priority amongst some actors, others rarely pay attention to it in their tweets, as visualised in Figure 8. These

differences combined with our subsampling strategy imply that the party Nye Borgerlige is excluded in the PCA. Thus, our final dataset used in the content analysis consists of 4.165 tweets distributed across nine parties.

Figure 8



6.2 Principle component analysis (11/16)

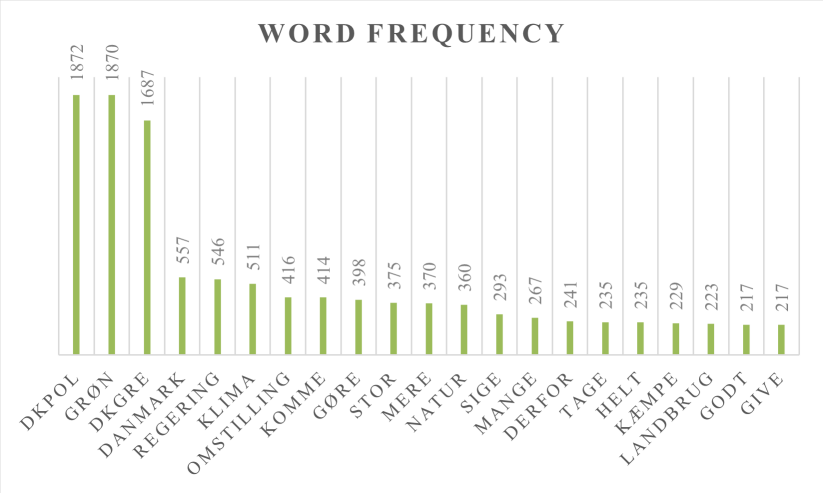
Through this part of our analysis, we aim to investigate and understand how political parties are related to each other through their use of words in tweets. We draw on the methodology from Jan Fuhse and colleagues and their definition of socio-symbolic constellations (Fuhse, Stuhler, Riebling, & Martin, 2020). They define socio-symbolic constellations as the way actors are related to each other through the symbols they use. In the context of our project, the political parties are seen as actors, while their use of words in tweets are seen as symbolic practices. In order to position the parties according to their use of words, we decide to utilise a PCA. PCA is an unsupervised method useful when work-

ing with complex data: by converting the data into new components ordered by their explained variance, it becomes possible to choose the relevant components, and thereby work with a summarised and more intuitive dataset (Wouter van Atteveldt, 2020).

In line with prior research, we decide to 'zoom in' on more local relations in our PCA instead of including all words from our corpus (Fuhse et al., 2020; Blok, 2020-working paper). In order to do so, we considered different words central in the context of the climate debate, such as "grøn" (green) and "bæredygtighed" (sustainability), but ended up agreeing on the word "klima" (climate) as our key term. "Klima" was chosen based on several reasons. First, we counted the words in our corpus and found that "klima" is one of the more frequently occurring words, only exceeded by "Danmark", "regering" (government) and "grøn" (green), as visualised in Figure 9. Secondly, we draw a random sample of 100 tweets containing either "klima" (climate) or "grøn", and find tweets containing "klima" to capture a broader context of our field of interest. While green is used as an adjective, the word "klima" is a noun that can be transformed into several nouns such as klimakrise (climate crisis), klimatopmøde (climate summit) and klimamålsætning (climate goals), which was further evident when identifying words that include "klima" as substrings in our PCA.

After defining the key term in our analysis, we determine a local neighbourhood consisting of the 400 most frequent words used in relation to klima and compound terms including klima such as "klimakrise" (climate crisis) or "klimamålsætning" (climate goal). These words are used to create an actor-term matrix, indicating how often the politicians are using each of the words. Each word frequency is afterwards standardised to create features with a mean equal to zero and a one-unit variance (Blok, 2020-working paper). We finally

Figure 9

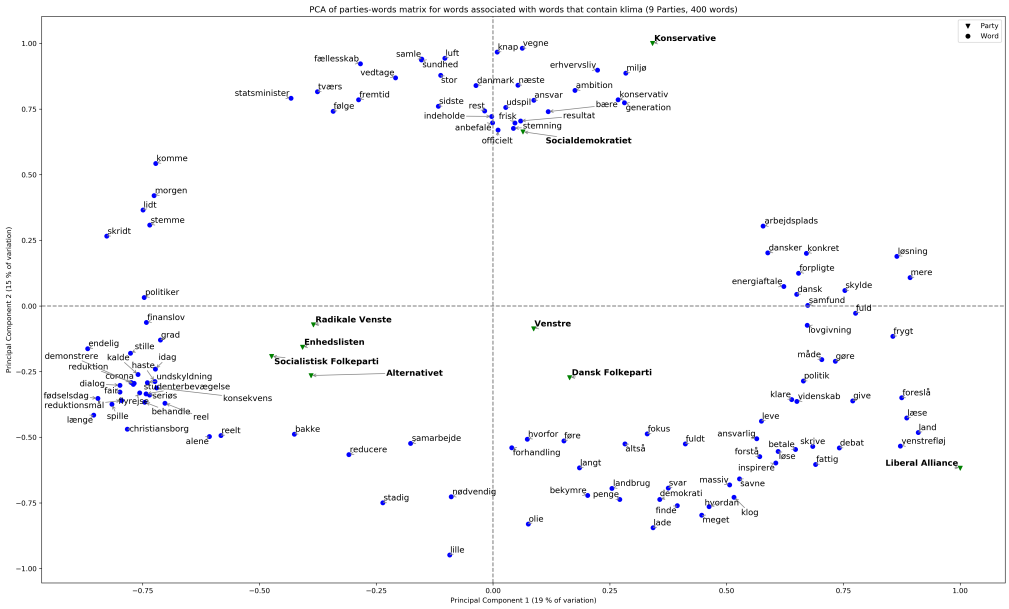


factorise the matrix into components through the PCA, leading us to nine different components (corresponding to our number of parties), with different levels of explained variance (Fuhse et al., 2020). We decide to include the first two principal components in our analysis, which together account for 34% of the explained variance in our original dataset. Even though this is a relatively low explained variance, and it can be discussed if these components are representing our original data well enough, we argue that the PCA presented in Figure 10, together with our netnographic campaign, provide us with valuable insights into how the parties, and the words they use in relation to klima, are positioned in relation to each other.

6.3 Main findings (32/17)

Based on the plot presented in Figure 10, we interpret the first axis to express different positions or attitudes regarding the climate debate. Amongst terms contributing the most to the axis, we find "løsning" (solution), "frygt" (fear), "videnskab" (science), and "reduktion" (reduction). We further see a significant

Figure 10: PCA of words associated with the word 'klima'



difference in the words used across the axis. On the left side, the terms indicate a need for immediate change and action through words such as "seriøs" (serious), haste (haste), "konsekvenser" (consequences), "undskyldning" (apology), and "studenterbevægelse" (student movement). These terms are positioned close to the left-wing parties (Enhedslisten, SF, Alternativet and Radikale Venstre). This corresponds to the tendencies found through our netnographic campaign, which show a stronger focus on the consequences of climate change and the generational injustice across the left-wing parties. The right side of the first axis is, by contrast, characterised by other terms such as "videnskab" (science) and "energy", which are positioned close to parties from the right-wing (Venstre, Dansk Folkeparti and Liberal Alliance). Nevertheless, terms such as "frygt" (fear) and "venstrefløjen"(left-wing) are also emerging, which we, based on the knowledge gained through our netnography, interpret as a criticism of the left-

wing parties' rhetoric regarding the climate crisis. An example could be how Liberal Alliance, several times accused left-wing parties of frightening children with what they define as grown-up problems, as shown in figure 11. Another essential finding is the positioning of Socialdemokratiet. Even though the party is known as a left-wing party, they are positioned on the right side of the first axis, closer to terms used by right-wing parties. Through our netnography we found some of the same tendencies, where Socialdemokratiet tweeted about the economy and the corporate sector as the only left-wing party.

Figure 11: Tweet from Liberal Alliance disagreeing with left-wing

The second axis also carries properties that could be identified as attitudinal, but we interpret it to identifying sentiments or connotations related to the word "klima". While the top part of the axis is expressing a higher degree of optimism, with words such as "fællesskab" (unity), "ambition", "sundhed" (health), "frisk" (fresh) and "fremtid" (future), the bottom part seems to be characterized by words that carry a more sceptical connotation. Here we see terms such as, "nødvendig" (neccessary), "olie" (oil), and "bekymre" (worry).

What we find interesting about this axis is that almost all parties, independent of affiliation, are positioned at the bottom of the axis, while only Socialdemokratiet and Konservative are positioned at the top. We find one possible explanation in our immersion journal, which shows observations of all parties expressing a disapproval towards the climate policies of Socialdemokratiet. According to our netnographic observations, this can be explained through a dynamic in which the left-wing parties don't find it ambitious enough, while the right-wing parties express scepticism if the policies are regarded as intrusive and radical.

7 Advanced Social Data Science 2 (34/28)

Through this part of the project, we aim to back up our insights gained through digital methods by deploying more automated means. We therefore use Word2Vec to define a subsample containing words related to the climate agenda of Danish parties and MPs which we then use as search queries to find relevant tweets. We then estimate a hierarchical stochastic block model (HSBM) to determine which topics the different parties talk about, and how the topics are distributed across party-lines. This part of the project will thus aim to answer the following research question:

What are the topics in the Danish climate debate on Twitter, and how do the parties differ topic-wise?

We seek to answer the question through data acquired from Twitter. As our population of interest is Danish parties and MPs, we initiate our analysis by scraping all actors with an active account (Appendix 7). As described in section 2.1, we decide to narrow our sample further down by focusing on tweets from the 1st of January 2019 until now. We choose this time interval to capture tweets in the period around the election in 2019, often referred to as the climate election

(Davidsen-Nielsen, 2019). We believe that this sampling strategy provides us with a comprehensive dataset containing a wide range of topics related to the climate debate. Our final dataset used in this part of the project is thereby containing 230.315 tweets distributed across 166 political accounts on Twitter.

8 Preprocessing (9/38)

After extracting a subset from our initial sample, we performed several preprocessing steps to make the inputs to our models less complex (Denny & Spirling, 2018) (Appendix 7). The steps include:

1. **Removing punctuation and noise:** We remove what we consider uninformative characters such as punctuation, separators, extra white spaces and special characters. Even though Twitter-specific annotation like mentions, hashtags, and URLs can be considered informative in some cases, we decide to remove them as they are beyond the interest of this project.

2. **Lowercasing:** We further lowercase all letters in the tweets. The rationale behind this step is to ensure that words are only assessed on the exact spelling and not their casing (Denny & Spirling, 2018).

3. **Removing words:** As the next step, we remove stop words, which are regarded as not carrying meaning to the content in the tweets. Before removing the stop words, we inspect the words included in the NLTK package to ensure that we do not exclude words of potential relevance (Wouter van Atteveldt, 2020). Additionally, we only keep nouns, verbs, adjectives, and adverbs. These word classes were chosen due to our interest in specific things, actions, and their characteristics (Wouter van Atteveldt, 2020).

4. **Linguistic preprocessing:** After removing superfluous material from the tweets, we subtract the tokens through a TweetTokenizer, thereby ending up with vectors containing the most relevant tokens from each document (Denny & Spirling, 2018). Accordingly, we lemmatize the tokens through Stanza, thereby reducing each word to its root. By lemmatizing the words, it is possible to avoid all conjugations in a dictionary in our models, making it easier to detect and subtract meaningful words through Word2Vec and thereby relevant topics through HSBM.

Apart from the steps described above, other considerations were made based on our prior knowledge of our field of interest. An example could be that we decide to keep numbers, as we expect them to carry essential information concerning how political parties and MPs communicate about climate. This was especially evident through our netnography, where words and phenomena such as *70-målet*, *CO2*, and *2030-målet* were central in how the parties and MPs framed their ambitions and goals.

9 Using Word2Vec to subset the data (11/16)

From our dataset of 230.315 tweets from Danish MPs and party accounts, we define a subset containing tweets related to the political discussion about climate. By using a Word2Vec model we extract a list of keywords related to the public debate surrounding climate and create a subset of tweets containing at least one of these words. The intuition of Word2Vec rests on the distributional hypothesis; that words that occur in similar distributions tend to have similar meanings (Jurafsky & Martin, n.d.). The model is trained on a large text corpus to use unsupervised learning to assess similarity of words from word-co-occurrence (Jurafsky & Martin, n.d.).

To perform computational methods on our data, the text must be transformed into numerical values. This is done through word embeddings which are representations of words through vectors (Jurafsky & Martin, n.d.). A logistic regression classifier is trained to test whether the word is likely to occur close to another given word by calculating its cosine similarity metric based on the dot product of their vectors (Jurafsky & Martin, n.d.). In practice, it means that words such as 'green' and 'climate' will be close to each other because they share contextual and thus distributional features in our tweets, while the word 'cookies' will be far from both.

9.1 Discussion of analytical choices (32/17)

We use the Word2Vec algorithm from the gensim package to run the model in Python and set the key parameters (Appendix 7). The window parameter defines how many neighbouring words to count when determining word similarity. Through a heuristic process of interchangeably adjusting the window, we find that a window of four leaves us with the best quality of words.

For our study, we are interested in words similar to "klima" (climate), and we use the Word2Vec model to find words that have similar vector-representations and thus share contextual similarity with this word in our twitter corpus. Our approach to keyword generation involves the human faculties of decision-making. While human beings recall words from their memory poorly, they are good at assessing whether words that are presented to them are representative for a given concept (King, Lam, & Roberts, 2017). By giving the trained model a seed word, "klima", it generates a list of words with matching values indicating the similarity to the seed word. For each word, we draw on our knowledge from our qualitative inquiry to manually assess whether each word suggested by the model is related to the climate debate or not. Each word is added to a list of

keywords and to a list of additional seeds for the model, so it keeps suggesting new similar words. We continue this process until we asses that the model does no longer generate meaningful words. We leave out words that we estimate will generate too much noise such as “future” (fremtid) even though the weight of the word indicates a high similarity to our initial seed “klima”. We end up with a list of 167 keywords (Appendix 8). By only selecting tweets containing minimum one of the keywords from the defined list, we end up with a climate sample containing 31.243 tweets. We assess the precision of this approach, by individually coding a random sample of 100 tweets from the sub-sample. We thereby assess the precision to be 87.5% and the intercoder reliability to be 97%. Even though our precision is relatively high, we notice that the model for example suggests the word ”miljø” (environment), which contributes to a considerable amount of noise. This poses a challenge, as it is a highly interesting word for our case, but at the same time a word that can be used in different contexts. For example, it leads to us subsetting tweets that include the word ”arbejdsmiljø” (work environment). Nevertheless we deem the word important for our analysis and choose to include it.

10 Hierarchical Stochastic Block Model (34/28)

We choose to deploy a computer-assisted exploration strategy to get an overview of which topics emerge from our sub-sample of climate related tweets, and to examine the parties’ topical distribution around the climate debate. More specifically, we estimate a hierarchical stochastic block model (HSBM) on our subset of climate-related tweets.

The HSBM relies on an integration of the methods from community detection and topic models (Gerlach, Peixoto, & Altmann, 2018). Both techniques can be deployed to extract useful information from unstructured data. Topic

models are applied to discover the main themes pervading in a large text corpus (Blei, 2012) whereas the idea of community detection is to find groups in data, that exhibit similar connectivity patterns (Gerlach et al., 2018). By unifying these two methods, HSBM infers topical structures by representing text corpora as bipartite networks, where the nodes consist of words and documents, and the strength of the edge between two nodes are determined by the number of occurrences of each word in the document. The method yields a hierarchy of nested stochastic block models (SBMs), meaning that clusters of nodes are grouped into larger clusters of nodes. In other words, the HSBM works as a generative process that forms a hierarchical order, where higher levels are formed by the levels below (ibid.).

10.1 Why HSBM? (9/38)

Another unsupervised method useful for exploring the topical distribution in unstructured data is latent Dirichlet allocation (LDA) (Blei, 2012). However, we have chosen HSBM rather than the LDA topic model for several reasons. Firstly, the HSBM better matches the statistical properties of real texts. Since HSBM makes fewer assumptions about the underlying structure of the data, the model can admit heterogenous group-mixtures at multiple scales. This also means that the HSBM is better in finding small topics than LDA (Gerlach et al., 2018). Secondly, HSBM is non-parametric and automatically extracts the number of topics based on the posterior distribution. In LDA, however, the user has to rely on a heuristic approach and manually choose the number of topics that the model should estimate, making it hard to know whether the chosen number of topics is adequate (ibid.). These fundamental downsides of the LDA topic model increase the risk of the LDA either over- or underfitting (ibid.). This point is further supported by Gerlach et al. (2018), who demonstrate how

HSBM outperforms LDA significantly both on real and artificial data and thus constitutes a better topic model overall. Furthermore, we tried running the LDA model on our data, but did not obtain any meaningful topics compared to the topics found by HSBM, which in this case reaffirms the notion of Gerlach et al. (2018) that HSBM is better at finding structures in the data than LDA.

10.2 Presentation of results (11/16)

We choose to estimate the HSBM on our sample of climate related tweets, i.e., one document corresponds to one green tweet from a politician or party as the model yields meaningful topics without aggregating by date, politician, or party. By keeping the tweets non-aggregated, we can get a higher level of document-level variation. We choose to remove the most infrequent words used (less than five) and most frequently used (more than 25000) from the text corpus, as we consider these words uninformative for the content of the topics. As previously mentioned, HSBM yields a hierarchical structure. This means that topics of the lowest level of the stochastic block model contains more fundamental information and high granularity of the data, whereas higher levels are based on clustering of the levels below, thus displaying the topics in a more aggregated form. We choose to display the topics on level 0 since this yields a high granularity of the topics, which we deem relevant for answering our research question. That is, in order to explore what topics MPs communicate about, we want to gain a basic knowledge of the composition and substance of these topics. Hence, exploring the topics of the lowest level of the HSBM enables us to focus on more specific topics, such as electrical vehicles, fossil fuels, and agriculture, which can guide our exploration of the climate related topics MPs and political parties' address in their communication on Twitter.

10.2.1 Deciding on topics (32/17)

The HSBM model yields 94 different topics on level 0, where we explore the 10 most probable words for each topic, which can be noted as $P(w|Topic = i)$. We draw on a heuristic approach when selecting topics. As a guiding principle we focus on the topics that are cohesive, i.e., include words that are semantically coherent and topics that are exclusive. All group members have qualitatively coded the topics. In the coding process we have relied on the knowledge gained of the topics from our previously deployed digital methods. Hereafter, we have aligned our codes and formulated suitable titles for the selected topics. Through this process, we end up with 23 topics in total (Appendix 9).

Based on our findings from digital methods we deem it relevant to delve deeper into the topics 'agriculture', 'vehicles' and 'European cooperation'. However, before proceeding with the further analysis we choose to validate the topics by taking a sample of 100 documents that have the highest probability of belonging to one of the topics and code these as true/false positives. Two group members have independently coded the topics, which yields a precision score and inter rater reliability score as displayed in Figure 12. We consider the high precision and inter rater reliability score as indicating that the topics are reliable and that we have selected proper titles for the topics. However, it is important to keep in mind that we are not able to display the true/false negatives and therefore cannot determine the model's accuracy or recall.

Figure 12: Validation matrix of topics

Topic	**Coder 1**		**Coder 2**		Inter-rater reliability
	True positive	False positive	True positive	False positive	
Agriculture	89	11	96	4	89%
Vehicles	100	0	100	8	100%
European community	90	10	92	8	98%

Figure 13 displays the words that have the highest probability of contributing

to the topics, respectively. The figures provide us with a sense of the substance and composition of the climate issues the Danish parties address on Twitter. As displayed in Figure 13 we see that the topic "agriculture" revolves around ecology and food, while the topic "vehicles" surrounds both electrical vehicles and fossil vehicles.

Furthermore, we are interested in the topical distribution of the parties, in order to gain a sense of differences and similarities party-wise. Since the topical probability of each document is on the level of individual tweets, we aggregate the documents by party and take the mean. By taking the mean value of the aggregated document probabilities we take into account that parties differ in frequency of tweeting around climate. We do this as we are interested in the topical focus of the parties, rather than how often they address the topics.

By aggregating tweets party-wise, we implicitly assume that the aggregation of tweets of politicians affiliated to the same party can be indicative of a party's general topical distribution. Drawing on the knowledge gained from our visual networks and PCA we think this is a reasonable assumption as these methods illustrate that members of the same party tend to appear in the same topical field and retweet each other and from similar accounts.

From Figure 14 we see that parties generally address most of the topics related to climate to some degree. Denmark's largest party, Socialdemokratiet (S), address all the included topics and are not standing out with extreme values in any of the topical distributions, which indicate that their topical focus is quite evenly distributed over all topics relating to climate. As a contrast, right wing parties such as Nye Borgerlige (NB), Dansk Folkeparti (DF), and Liberal Alliance (LA) are dominant in some climate topics, and not present in others. As we see from Figure 2, they do not have a high frequency of tweets related to the debate surrounding climate, thus indicating that they claim certain issues

Figure 13: Words with the highest probability in chosen topics

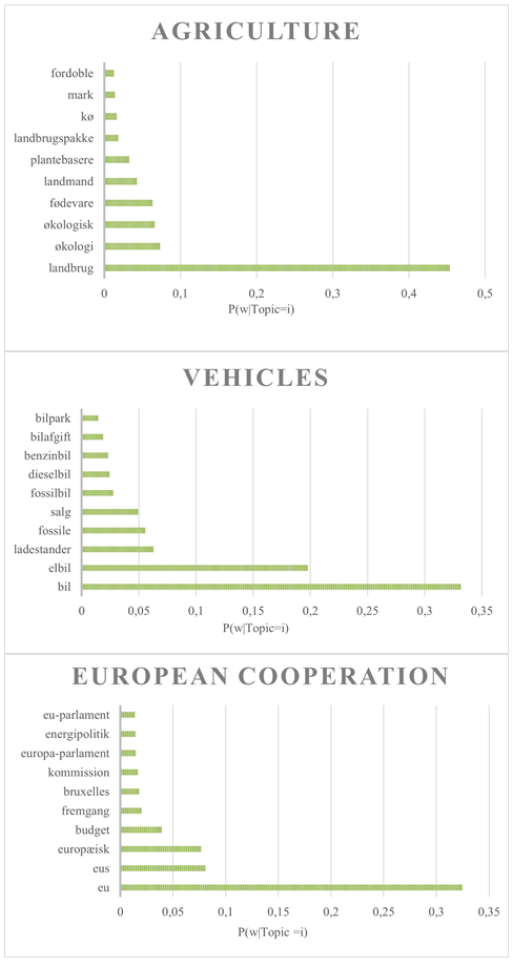
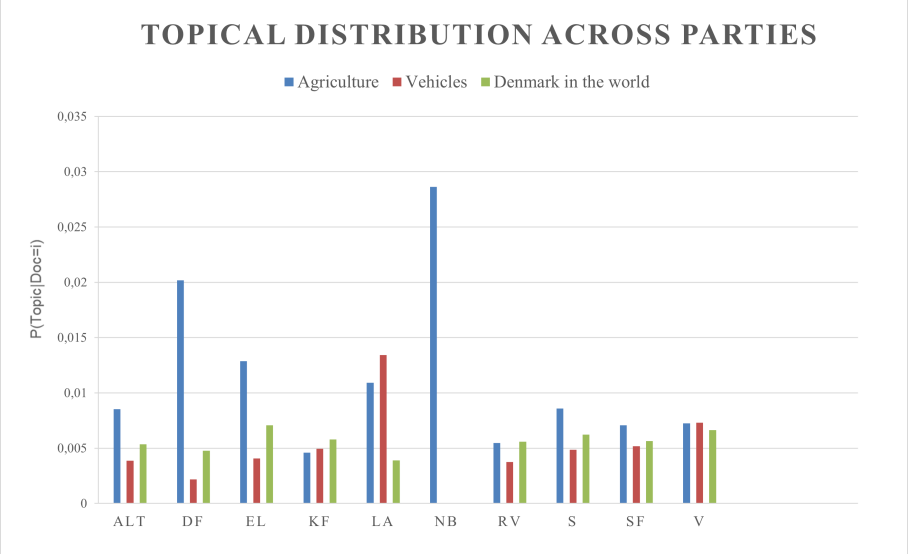


Figure 14



over others.

In summary, the HSBM has enabled us to explore the prevalent climate related topics in the political debate, provided us with a better sense of the content of the topics, and indicated that parties differ in the topics they address. However, we are aware of the inherent biases of ML-systems, and understand that they do not always provide accurate representation. Nevertheless, combined with the other quantitative and qualitative methods used in Digital Methods, we argue that the findings from our topic model align with other observations, and are therefore to some extent robust.

11 Quali-quantitative integration across methods (34/28)

The methods that we apply throughout the project represent both qualitative and quantitative approaches to data collection and analysis. We argue that we

have successfully integrated the methods as they have been complementing each other through validation and contextualization. With this quali-quant integration, we have aimed to mitigate a concern raised by Munk, namely that the field of digital sociology is struggling with meaning making, and when mapping digital traces there's always doubt about how these were intended, what they meant, and what the context was (Munk, 2019).

Overall, our immersion journal was diligently used to provide context to our quantitative findings. We recognize netnography as advantageous in grasping the meaning of actions, by providing exhaustive contextualization and thus a basis for interpretation. Oppositely, the quantitative methods allow us to extend our analysis to include more data, and thus move from a micro perspective to phenomena on a macro level, and help us expand our territory of interest. For example, when exploring our manual network, we noticed tendencies that we wanted to explore at a larger scale, to rule out that this was a specific phenomenon attributed to a specific case. Thus we automated a network of retweets that included a higher amount of actors to see if the tendency was evident for a bigger sample, too.

Conversely, the quantitative analysis could suggest interesting points of further examination that we had not thought about, as was the case with our immersion journal and PCA. As we saw words such as "frygt" (fear) and "venstrefløjen" (left-wing) emerging in the PCA, it sparked our curiosity. We thus went back to scouting (Kozinets, 2020), and discovered tweets in which Liberal Alliance accused the left wing for using fear-inducing rhetorics, especially towards children. This is how the relationship between the PCA and our netnography became a mutually reinforcing one; we used our immersion journal to create context around the relationships that the PCA showed, and then we used the PCA to find new and interesting data in an iterative and exploratory

process.

We experienced the same kind of quali-quant synergy between the codes in our immersion journal and our topic model used for ASDS2. As previously stated, we structured our observations through manual coding, however aware of the fact that the practice has been under scrutiny (Lee & Martin, 2014): The criticism entails that it simplifies to the extent that it does "violence" to the material and disturbs the interpretative aspect of hermeneutics (Lee & Martin, 2014). Furthermore, coding seemingly leads to drawing conclusions not based on the actual patterns that emerge in the data, but on *"how we happen to label the material"* (Lee & Martin, 2014). We take into consideration that we might have simplified and biased the findings in the process, despite having assured inter-coder reliability (MacQueen et al., 1998). While this is possibly an inherent problem in qualitative research, we argue that the integration of quantitative methods can help researchers in assessing whether their codes are arbitrary labels, or if they actually reflect the patterns in the data. We ran an unsupervised pattern recognizing model (HSBM) that learns from the data itself on our tweets. With this, one could say that we are performing the exact opposite of manual coding on our data, as we remove the human interpretive aspect from the pattern detection. Yet, our model finds some categories that coincide with our manual codes. For example, while we had manually assigned the codes "Transport", "International perspective - Denmark in the world", and "Agriculture" to our material, the HSBM produced distinctive topics in which words associated with these topics occurred (Appendix 9).

Furthermore, the quantitative unsupervised methods are useful to identify patterns and word similarity in large text corpora that are not possible to qualitatively assess (Carlsen & Raulund, 2021-working paper). In line with Carlsen and Raulund (2021-working paper)'s alternative approach to computa-

tional grounded theory (CALM), we argue that while computational methods are widely beneficial, keeping the humans in the loop remain an important aspect in this kind of research. As a conclusive remark, we argue that the various qualitative and quantitative methods applied to a large extent are able to fill out each other's weaknesses and providing relevant data for answering our research question.

12 Conclusion (All)

Through this project we used a range of both quantitative and qualitative methods to investigate how Danish parties and MPs frame their political agenda in relation to climate on social media sites such as Twitter and Facebook. All methods have contributed to different insights, and have, in combination, been valuable tools for understanding and exploring our field of interest. Through our project it has become evident that Danish parties and MPs make use of different strategies when framing their agenda. While the left-wing is tweeting more frequently about climate related topics, they are also more oriented towards the consequences of climate change, and demand more immediate action. In contrast, the right-wing has a stronger focus on economic aspects and the consequences for involved parts such as farmers. Common for both parts is that they share a general dissatisfaction with the government's climate policies.

However, these findings are, naturally, tainted by the socio-technical affordances that shape the communication on Twitter and Facebook. This sparks a reflection of how the data is 'given' to us , in which we bring the awareness of the fact that the data comes from a private company with its own logics and business model, and is not created for research purposes (Venturini, Bounegru, Gray, & Rogers, 2018). Consequently, one concluding remark is that using 'Digital Methods' as a tool for social-scientific inquiry requires an intellectual plasticity

in terms of determining how digital objects can be re-purposed for social and cultural research. Simultaneously, a comprehension of how the digital objects themselves shape the conditions of the objects of study is necessary to account for this bias (Rogers, 2009; Marres & Gerlitz, 2016). This entails studying both the medium itself and the digital traces as proxies for the broader political communication surrounding climate. Having argued that the quali-quant integration ensures robust findings from the communication on Twitter, we claim to have examined the behaviour on Twitter quite thoroughly. However, if our endeavour was to gain an understanding of how politicians set their agendas while accounting for the affordances and constraints of the platform(s), we would ideally conduct additional in-depth interviews with the actors, as has been done by (Ladrech & Little, 2019), in an attempt to uncover the deliberative motivation behind the agenda-setting. This could provide more robust findings, instead of trusting data "given" from Twitter as reflecting the behaviour of interest. This, of course, touches upon the more fundamental question of how "real" the digital traces we leave behind are, i.e. to what extent they reflect actual behaviour. As such, we argue that unravelling the self-assessed motivations behind the agenda setting by the individual politician offers a contribution to Couldrys effort of re-connecting hermeneutics with big data analysis; an additional in-depth interview might unravel the (self-assessed) agency of the actor and how they navigate in and interpret their actions in Web 2.0. Such an approach might serve insights into how the actors are shaped by the medium, but also how the actors interpret the medium and use it to modify their behaviour. At its most ambitious, such efforts can provide an interpretive, human-oriented approach to big data analytics.

References

Blach-Ørsten, M., Eberholst, M. K., & Burkal, R. (2017, October). From hybrid media system to hybrid-media politicians: Danish politicians and their cross-media presence in the 2015 national election campaign. *Journal of Information Technology & Politics*, *14*(4), 334–347. Retrieved from `https://doi.org/10.1080/19331681.2017.1369917` doi: 10.1080/19331681.2017.1369917

Blau, J. (2017). *The paris agreement*. Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-53541-8` doi: 10.1007/978-3-319-53541-8

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blok, A. e. a. (2020-working paper). Inter-risk framing contests'. under review for american journal of cultural. *American Journal of Cultural Sociology*.

Brooker, P., Barnett, J., & Cribbin, T. (2016, July). Doing social media analytics. *Big Data & Society*, *3*(2), 205395171665806. Retrieved from `https://doi.org/10.1177/2053951716658060` doi: 10.1177/2053951716658060

Brundtland, G. H. (1987). Report of the world comission on environment and development: Our common future.

Campagnolo, G. M. (2020). The analogue mapping. In *Social data science xennials* (pp. 35–50). Springer International Publishing. doi: 10.1007/978-3-030-60358-8_3

Carlsen, H. B., & Raulund, S. (2021-working paper). Computational grounded theoryrevisited: from computer lead tocomputer assisted text analysis. *Big Data and Society*.

Couldry, N. (2014, November). Inaugural: A necessary disenchantment: Myth, agency and injustice in a digital world. *The Sociological Review*, *62*(4),

880–897. Retrieved from https://doi.org/10.1111/1467-954x.12158 doi: 10.1111/1467-954x.12158

Davidsen-Nielsen, A. H. . H. (2019). Folket skubbede politikerne mod danmarks første klimavalg. *Politiken*.

Decuypere, M. (2020). Visual network analysis: a qualitative method for researching sociomaterial practice. *Qualitative Research*, *20*(1), 73-90.

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168–189. doi: 10.1017/pan.2017.44

Fuhse, J., Stuhler, O., Riebling, J., & Martin, J. L. (2020, February). Relating social and symbolic relations in quantitative text analysis. a study of parliamentary discourse in the weimar republic. *Poetics*, *78*, 101363. Retrieved from https://doi.org/10.1016/j.poetic.2019.04.004 doi: 10.1016/j.poetic.2019.04.004

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*, *4*(7).

Green-Pedersen, C., & Mortensen, P. B. (2010, March). Who sets the agenda and who responds to it in the danish parliament? a new model of issue competition and agenda-setting. *European Journal of Political Research*, *49*(2), 257–281. Retrieved from https://doi.org/10.1111/j.1475-6765.2009.01897.x doi: 10.1111/j.1475-6765.2009.01897.x

Housley, W., Webb, H., Edwards, A., Procter, R., & Jirotka, M. (2017). Digitizing sacks? approaching social media as data. *Qualitative Research*, *17*(6), 627-644.

Jurafsky, D., & Martin, J. H. (n.d.). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.*

Kaiser, J., Rhomberg, M., Maireder, A., & Schlögl, S. (2016, 08). Energiewende's lone warriors: A hyperlink network analysis of the german energy transition discourse. *Media and Communication*, *4*.

King, G., Lam, P., & Roberts, M. E. (2017, April). Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science*, *61*(4), 971–988. Retrieved from `https://doi.org/10.1111/ajps.12291` doi: 10.1111/ajps.12291

Kozinets, R. (2020). *Netnography : the essential guide to qualitative social media research*. Los Angeles: SAGE.

Ladrech, R., & Little, C. (2019, June). Drivers of political parties' climate policy preferences: lessons from denmark and ireland. *Environmental Politics*, *28*(6), 1017–1038. Retrieved from `https://doi.org/10.1080/09644016.2019.1625157` doi: 10.1080/09644016.2019.1625157

Lai, S. S., Pagh, J., & Zeng, F. H. (2019). Tracing communicative patterns: A comparative ethnography across platforms, media and contexts. *Nordicom Review*, *40*(s1), 141–157. Retrieved from `https://doi.org/10.2478/nor-2019-0019` doi: doi:10.2478/nor-2019-0019

Lee, M., & Martin, J. L. (2014, December). Coding, counting and cultural cartography. *American Journal of Cultural Sociology*, *3*(1), 1–33. Retrieved from `https://doi.org/10.1057/ajcs.2014.13` doi: 10.1057/ajcs.2014.13

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998, May). Codebook development for team-based qualitative analysis. *CAM Journal*, *10*(2), 31–36. Retrieved from `https://doi.org/10.1177/1525822x980100020301` doi: 10.1177/1525822x980100020301

Marres, N., & Gerlitz, C. (2016, February). Interface methods: Renegotiating relations between digital social research, STS and sociology. *The Sociolog-*

ical Review, *64*(1), 21–46. Retrieved from https://doi.org/10.1111/1467-954x.12314 doi: 10.1111/1467-954x.12314

Munk, A. K. (2019, June). Four styles of quali-quantitative analysis. *Nordicom Review*, *40*(s1), 159–176. Retrieved from https://doi.org/10.2478/nor-2019-0020 doi: 10.2478/nor-2019-0020

Rafail, P. (2017, May). Nonprobability sampling and twitter. *Social Science Computer Review*, *36*(2), 195–211. Retrieved from https://doi.org/10.1177/0894439317709431 doi: 10.1177/0894439317709431

Rogers, R. (2009). *End of the virtual: Digital methods*. Amsterdam University Press.

Venturini, T., Bounegru, L., Gray, J., & Rogers, R. (2018, April). A reality check(list) for digital methods. *New Media & Society*, *20*(11), 4195–4217. Retrieved from https://doi.org/10.1177/1461444818769236 doi: 10.1177/1461444818769236

Wasserman, S. (1994). *Social network analysis : methods and applications*. Cambridge New York: Cambridge University Press.

Wouter van Atteveldt, D. T. C. A. C. (2020). *Computational analysis of communication*. Wiley-Blackwell; to be published October 2021.

Øyen, M. (2019). På få måneder blev alle politikere klimakrigere. hvad var det egentlig, der skete? *Altinget*.