Political Communication During a Natural Disaster: The Case of the 2019-2020 Australian Bushfires

Group Name: Fri-753377-8

Character Count (incl. spaces): 61739

June 17, 2021

Abstract

This paper examines how the framing of a natural disaster by politicians can shape the public discourse. We use the 2019-2020 Australian bushfire season, also known as the Black Summer, as a particularly interesting case due to its possible relation to climate change. By employing a mixed-methods strategy, we effectively map the political discourse in Australia using an intensive cross-platform netnography and extensive quantitative analysis of Twitter-data. We find that the Australian political landscape is highly polarised into political blocs despite the occurrence of a national crisis. These two blocs differ significantly in their framing of the bushfire, with the opposition, particularly the Labor Party and the Greens, "exploiting" the situation to push a political agenda on sustainable transition and approach climate change. The government party and its support parties on the other side, tackle the bushfire as a "regular" natural disaster with a focus on giving thanks to local authorities and showing support for communities that were hit the hardest.[1]

[1] All the relevant code used to generate the results in our exam paper can be found at our Github page under the directory "Notebooks": https://github.com/frederik-kilpinen/Exam2021-DM-ASDS2

Contents

1 Introduction 4

2 Previous Research 4

3 Research Design and Methods 5

3.1 Data . . . 9

3.1.1 A Note on Anonymisation . . . 9

4 Complementary Analysis of the Netnography, Manual Network, and Retweet Network 10

5 Content Analysis 14

5.1 Party-Hashtag Networks . . . 14

5.2 MP-Semantic Network . . . 19

6 Advanced Social Data Science 2 21

6.1 Text Preprocessing . . . 21

6.2 Topic Models: LDA and hSBM . . . 22

6.2.1 Topic Model Results . . . 22

6.3 Computer Assisted Keyword Retrieval . . . 24

6.3.1 Bushfire Time-Series . . . 26

6.4 Predicting Party on Word Usage . . . 27

6.4.1 Validity of Results . . . 28

7 Quali-Quantitative Integration 30

7.1 Using Qualitative Insights to Assess Quantitative Bias . . . 32

7.2 Platform Dependency . . . 33

8 Conclusion 34

References 35

A MP-Hashtag network degree centrality 37

B Party-Hashtag network degree centrality 38

List of Figures

1 Research Design . . . 6

2 Retweet Network . . . 10

3 Manual Network . . . 11

4 Tweet by Adam Bandt . . . 13

5 McCormack Meme . . . 13

6 Retweets by McCormack . . . 14

7 Party-Hashtag Network . . . 16

8 Party-Hashtag Network (excluding single use hashtags) . . . 17

9 Tweets by Adam Bandt & Peter Kahil . . . 18
10 Tweet by Susan Templeman . . . 19
11 MP-Semantic Network based on Word Co-occurrence . . . 20
12 hSBM Bushfire Topic Time-Series . . . 23
13 Schematic approach of computer assisted keyword retrieval . . . 24
14 Bushfire Subset Time-Series . . . 26
15 Distribution of Top 100 AMEs across model specification . . . 29
16 Top 100 AMEs Rank-order Correlation across Model Specification . . . 29
17 hSBM Climate Topic Time-Series . . . 31
18 Jason Falinski . . . 32
19 Tweet by Mark Butler . . . 33

List of Tables

1 Bushfire Subset Descriptive Statistics . . . 9
2 Topic Model Results . . . 23
3 Boolean Keywords . . . 25
4 hSBM and Bushfire Subset Correlation . . . 27
5 The Australian Greens Top 10 AMEs . . . 28
6 The Liberal Party Top 10 AMEs . . . 28
7 Rank-order Correlation between $l1$ and $l2$ Regularization . . . 30
8 Rank-order Correlation between $l1$ and No Regularization . . . 30
9 MP-Hashtag network degree centrality . . . 37
10 Party-Hashtag network degree centrality . . . 38

DM ID	ASDS 2 ID	Responsible for Sections
37	23	1, 2, 4, 6, 6.3, 7, 8
20	22	1, 3, 5, 6.1, 6.3.1, 7.1, 8
31	31	1, 3.1, 5.1, 6.2, 6.4, 7.2, 8
11	20	1, 3.1.1, 5.2, 6.2.1, 6.4.1, 8

1 Introduction

In June 2019, the Australian bushfire season took off surprisingly early in the state of New South Wales and was the beginning of what would later be known as the *Black Summer*. The 2019-2020 bushfire season was one of the most severe bushfire in recent history, claiming close to 500 lives and an estimated cost of more than 100 billion Australian Dollars (Davey & Sarre, 2020). Needless to say, the fires were at the center of political attention, especially when Prime Minister Scott Morrison went on holiday to Hawaii in November 2019 as the bushfires were about to peak. Allegations of Morrison not taking the bushfires seriously were, by some political opponents, linked to a general unwillingness by the government to undergo necessary *sustainable transitions* to tackle climate change. This illustrates an interesting dynamic in the political landscape, namely how political entrepreneurs can utilise a natural disaster to push a political agenda. Therefore, we pose the following research question:

- *How do politicians frame natural disasters and how does it relate to their party affiliation?*

To shed light on this question, we will use the Australian bushfire as a case study, focusing on the political communication of Australian MPs. We focus on the bushfire since its severity can be linked to climate change and it is therefore possible that politicians will leverage this natural disaster to push communication and perhaps even policy towards a sustainable transition. As a final remark before we dive into the literature, we emphasise that we refrain from stating whether or not the bushfires were in fact related to climate change. First of all, we do not have the competencies to draw such conclusions. Secondly, and more importantly, it is not important for our research question, as we are interested in the political response to a natural disaster, not the actual event itself.

2 Previous Research

There exists a large body of literature relating to the role that crisis events, such as accidents, natural disasters, and terror attacks, play in the shaping of public discourse and political agendas. The focus here, more generally, is on how these events act as catalysts for policy change or as tipping points for the adoption of new socio-political regimes (Pelling & Dill, 2009). While the scope of our research is less concerned with concrete policy change as a result of the bushfires, in part due to the relative recency of these events, the way in which members of parliament frame the issue, and how political parties may use the crisis to set the political agenda will be our main focus.

An important distinction to be made when researching the role that a certain event plays in contributing to public discourse and political agendas is whether it can be considered a crisis or a disaster. Faulkner (2001) attempts to clarify this distinction, claiming that while crises tend to be "induced by the actions or inactions of an organisation", disasters are seen as unanticipated and unpredictable natural phenomena, beyond the control of any given organisation. While this distinction can be a useful one, there are many instances where the line between what constitutes a crisis and what constitutes a disaster remains unclear, especially when events may be seen as direct consequences of climate change. Alongside this blurred line, there exists an incentive for politicians in a policy debate to either depict an event as a crisis, set in motion by human or organisational negligence or alternatively, to argue that the same event is rather a natural disaster and is therefore beyond the control of any responsible entity. This kind of calculated blame fixing is key to paving the foundations for certain policy directions and is hugely influential in political agenda-setting (Stone, 1989). Ultimately, the interpretation and communication of these events by politicians and the media determine whether the event in question becomes an important policy issue. Parliamentarians framing the bushfires as a result of global warming may leverage the crisis to push for climate action, while those who fail to draw this connection would regard such response futile, claiming that events are beyond both human and organisational control.

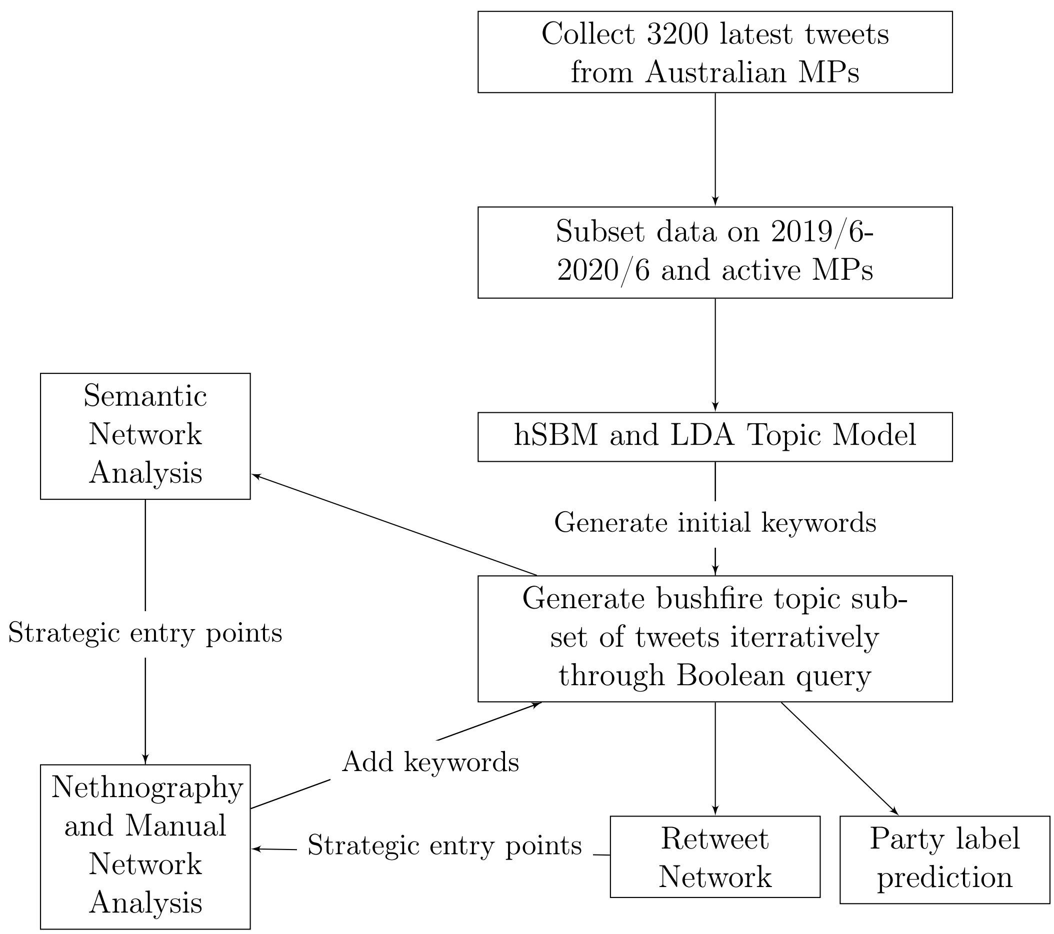
The popularity of Twitter, not only as a social network but as a news source, has led to the site becoming one of the main social mediums used by politicians (van Vliet, Törnberg, & Uitermark, 2020). As a result, the analysis of both political communication and discourse is now increasingly carried out using a mixture of quantitative and qualitative methods on Twitter data. The COVID-19 pandemic enabled many researchers to analyse tweets by politicians to investigate political communication surrounding disasters, how it circulates, and the resulting social implications. Manfredi-Sánchez, Amado-Suárez, and Waisbord (2021) adopt this approach, utilising a qualitative comparative analysis and a quantitative content analysis to determine how four recently elected presidents (of Brazil, Spain, Mexico, and Argentina) managed their political communication on Twitter during the first few months of the COVID-19 pandemic. While Kerbleski (2019) considers a set of Donald Trump's tweets for an in-depth, qualitative critical discourse analysis to investigate how Trump's discourse functions in society, particularly in response to the Californian wildfires of 2018. The selection of a similar case study, the Australian *Black Summer* of 2019/2020, affords us not only the exploration of political communication surrounding a crisis but also to explore the response to, what many consider, a pressing environmental issue and the potential sustainable transition that may succeed events.

3 Research Design and Methods

In this section, we outline the overall research design of our study. For clarity, as our study includes many steps and different methods, we present a schematic overview of the different methods and how they integrate to form a comprehensive and congruent strategy. A flow-chart describing the process is visualised in Figure 1.

Before continuing, a brief justification of the general rationale behind our design choice is in order. Our target sample consists of tweets from a well-defined actor-set of Australian MPs during the specific period of the 2019-2020 bushfires. Moreover, we are interested in tweets relating to a specific topic, namely, bushfires. Against this background, the target sample can be described as bounded in terms of actors, time, and topics (Rafail, 2018). The temporal demarcation of the sample is straightforward, using the timestamps associated with each tweet. Further subsetting into tweets that relate to the bushfire is, however, far from trivial. We anticipate that the quality of our analysis, to a large extent, relies on our ability to discriminate bushfire tweets from non-bushfire tweets. This task is further complicated by the partial overlap of the bushfires and the COVID-19 pandemic. COVID-19 related tweets likely share many words with tweets referring to the bushfires, since both are frequently referred to as crises or disasters. For these reasons, our research design to a high degree revolves around resolving these issues. To do so, we draw inspiration from Raalund and Carlsen (2021) and their framework for textual analysis *computer assisted learning and measurement* (CALM) through which computational tools like topic modeling are used to assist, but the focus is on human immersion and direct validation. In this regard, we also explicitly try to integrate methods and ideas from ASDS2 and Digital Methods.

Figure 1: Research Design



We now proceed with the schematic overview of our research design and introduce the methods in more detail.

1. **Data Collection:** As an initial step, we fetch the 3200 latest tweets (query limit) from each MP with an active Twitter account using the Twitter API. To ensure that only active politicians are included in our dataset, we limit our sample to be comprised solely of members of parliament that were elected representatives in the 45th and/or the 46th Parliament of Australia. If a politician was elected

for only one parliament, we only extract tweets within the period they were in parliament. Next, we limit the data to contain tweets from the time period of the bushfires between June 2019 and May 2020.

2. **Initial Keywords through Topic Models:** We perform text preprocessing steps to the corpus of tweets and fit two topic models - hierarchical Stochastic Block Model (hSBM) and Latent Dirichlet Allocation (LDA) - to find the bushfire topic (Blei, 2012; Gerlach, Peixoto, & Altmann, 2018). We use the most predictive words as an exploratory and computationally informed approach to derive initial keywords relating to the bushfire. More details about the implementation are given in Section 6.2.

3. **Bushfire Subset through Boolean Query:** Borrowing notation from King, Lam, and Roberts (2017), we create a reference query Q_R and a search query Q_S based on the keywords from the topic model. We restrict Q_R to a small amount of keywords K_R which we qualitatively asses to unambiguously classify a post as being related to the bushfire, e.g. the keyword "bushfire". Using a simple Boolean approach, we define a reference set R of tweets t as $R = \{t : Q_R\}$. The search query Q_S is instead defined more broadly as keywords which are suspected to be used in bushfire tweets, e.g. "disaster". More details on this are given in Section 6.3.

4. **Semantic Network Analysis:** We use the reference set R to generate bipartite semantic networks of hashtags and word co-occurrence. The nodes either represent individual MPs or are aggregated on political party. These networks are used to explore central actors and sub-topics and note down any relevant keywords which can be used to expand either Q_R or Q_S. More details about our implementation is given in Section 5.

5. **Retweet Network Analysis:** We again use the reference set R, this time to create a retweet network. The nodes represent MPs and edges represent retweets between MPs. Next to the characteristics already described, an advantage over the manual network is that the computational method allows us to be exhaustive, so to include every single MP in our analysis that is active on Twitter and their relations to other MPs in form of retweets. Even though retweets are frequently interpreted as a sign of endorsement (Boyd, Golder, & Lotan, 2010), combining the strengths of the computational and manual network allows us to ground the patterns found in the retweet network with observations from "the field" and increase interpretative validity (Blok et al., 2021), while at the same time validating emerging patterns from the manual network with coherent findings in computational methods. This provides an accurate and differentiated picture of the relations between political actors during the bushfire crisis.

6. **Netnography:** Using the insights about central actors gained from steps 4 and 5 as strategic entry points, we initiate our netnography. In doing so, we again make sure to note down any relevant keywords to expand Q_R and Q_S. We think that this approach can guide us to toward a more systematic and focused netnographic exploration. This is especially relevant for us as the actor-set is relatively large,

making it unattainable for us to be entirely exhaustive. In line with Lai, Pagh, and Zeng (2019), we use different strategies to demarcate the scope of our immersion. First, the context is determined in advance and only data within the context of the 2019/2020 bushfire season is regarded, in line with our research question. Secondly, we are only interested in communication that either originates from or addresses active politicians or their political parties. Therefore, we will not examine, for example, how NGOs or other organisations frame the event. On the one hand, this limits the diversity of perspectives on the fires that we encounter. On the other hand, this allows us to dive even deeper in the communication patterns that evolve from active politicians and their parties. This clear scope further facilitates the coherence between our qualitative and quantitative analysis since the dataset for our quantitative analysis only consists of text originating from active MPs. Thirdly, we aim to gain information from different platforms to mitigate Twitter-specific platform effects and detect cross-platform communication patterns. Taken together, our aim is not to get a broad picture of the general public discourse surrounding the bushfires but rather to deeply immerse into the positioning, framing, and interrelations of active MPs and their parties during this national crisis.

7. **Manual Network Analysis:** Even though computational methods for network analyses possess considerable advantages such as reproducibility and the ability to process large amounts of data, Venturini, Bounegru, Gray, and Rogers (2018) point out that these approaches also contain considerable shortcomings. Computational networks are always moderated by the affordances of the different platforms and might not capture the nuances of interpersonal relations. In a similar vein, Venturini et al. (2018) argue that these methods only allow for representations that are mathematically tractable. Moreover, dominant network tools are not able to capture multiple edges between two nodes, even though very different types of relations might exist simultaneously. Accordingly, as Caliandro (2018) outlines, the goal of online ethnography is not merely to identify communities most exhaustively but to map the practices through which actors and digital platforms structure social formations around a focal object. In this study, the focus is rather on an event, however, the same mechanisms apply and, using a positional map as used by Campagnolo (2020), we aim to understand the political situation around the bushfire topic and investigate relations between political actors, parties, and news media. Based on the observations during the netnographic immersion, the visual network was established and continuously refined throughout our analyses.

8. **Iterate Steps 3-7 and Create Final Subset:** Having done steps 3-7 and created Q_R and Q_S, we implement the algorithm proposed by King et al. (2017) in conjunction with a pre-trained word-embeddings model to mine keywords that further expand our query (Mikolov, Grave, Bojanowski, Puhrsch, & Joulin, 2017). We then repeat steps 3-7 until we believe we have exhausted all relevant keywords. From the final set of keywords, we build a last Boolean query which we use to create the final subset of bushfire tweets. Based on this data, we also re-run our networks one last time, which are then presented in the analysis. More details about the algorithm are again given in Section 6.3.

9. **Party Label Prediction:** For ASDS2, we run a multinomial logistic regression to predict the party of our bushfire tweets. This is used to understand which words are most predictive of the different parties. The implementation and results are given in Section 6.4.

As a final note, it should be mentioned that the steps above are a stylised and somewhat idealistic depiction of our design. In practice, the processes are not linear and we jump back and forth between the different steps.

3.1 Data

To ensure that we exhaustively sample all relevant actors, we rely on the manually validated list of Australian MPs and their Twitter handles in The Twitter Parliamentarian Database, gathered by the EU-funded ODYCCEUS project (van Vliet et al., 2020). After running through steps 1-8 as outlined in our research design, we end up with a subset of 2,988 bushfire tweets out of the 167,296 tweets initially collected. We present descriptive statistics of the subset in Table 1. The numbers of tweets and MPs are very unevenly distributed across parties. For example, Katter's Australian Party only has one tweet and MP whereas Labor has 1,639 and 69 MPs respectively. For this reason, most of the quantitative methods employed do not include Center Alliance or Katter's Australian Party. Sometimes they are also automatically excluded, as is the case in the retweet network due to a lack of ties.

Table 1: Bushfire Subset Descriptive Statistics

Party	**Tweet Count**	**Unique MPs**	**Top Tweeter**
Australian Greens	197	1	Adam Bandt
Australian Labor Party	1639	69	Mike Kelly
Centre Alliance	29	2	Rebekha Sharkie
Katter's Australian Party	1	1	Bob Katter
Liberal National Party of Queensland	300	10	Stuart Robert
Liberal Party of Australia	440	47	Scott Morrison
The Nationals	382	12	Darren Chester

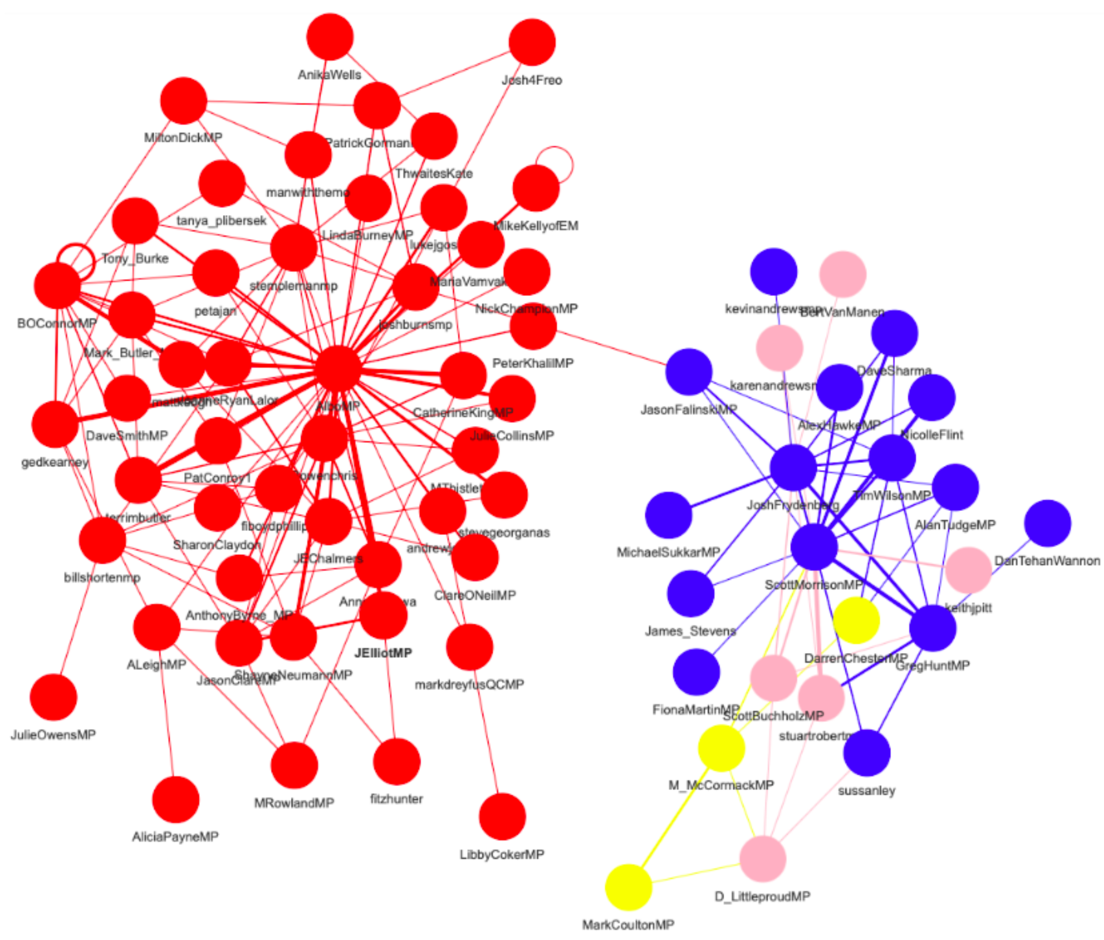
3.1.1 A Note on Anonymisation

A question, that according to Kozinets (2015) should be addressed no matter the project, is that of appropriately anonymising data. However, in our paper, we have decided *not* to anonymise any actors. The paper's actors are Australian MPs communicating on social media (primarily Twitter). Therefore, we regard the data to constitute public statements by public figures which are not revealing any private interactions. In most cases, it can be assumed that the MPs' post messages on Twitter to reach as many people as possible with their political agenda. Additionally, being able to identify the MPs can be of great importance for our analysis. In terms of highlighting and explaining political behaviour, but also in regards to validation purposes, e.g. to check whether the networks look as we would expect them to look. We also note that not anonymising MPs, especially when dealing with Twitter data, is standard practice in the political science literature. From a consequentialist point of view, we argue that the potential benefits from not anonymising the Australian MPs in our paper exceeds the harm. Needless to say, all data is gathered and stored on the University of Copenhagen's OneDrive in compliance with GDPR.

4 Complementary Analysis of the Netnography, Manual Network, and Retweet Network

This section is an integrated analysis that is built upon our netnography, retweet network (Figure 2) and manual network (Figure 3). Note that in the manual network, the Liberal National Party of Queensland and the Liberal Party of Australia are merged as they represent congruent political agendas and only differ in representing different regions.[2]

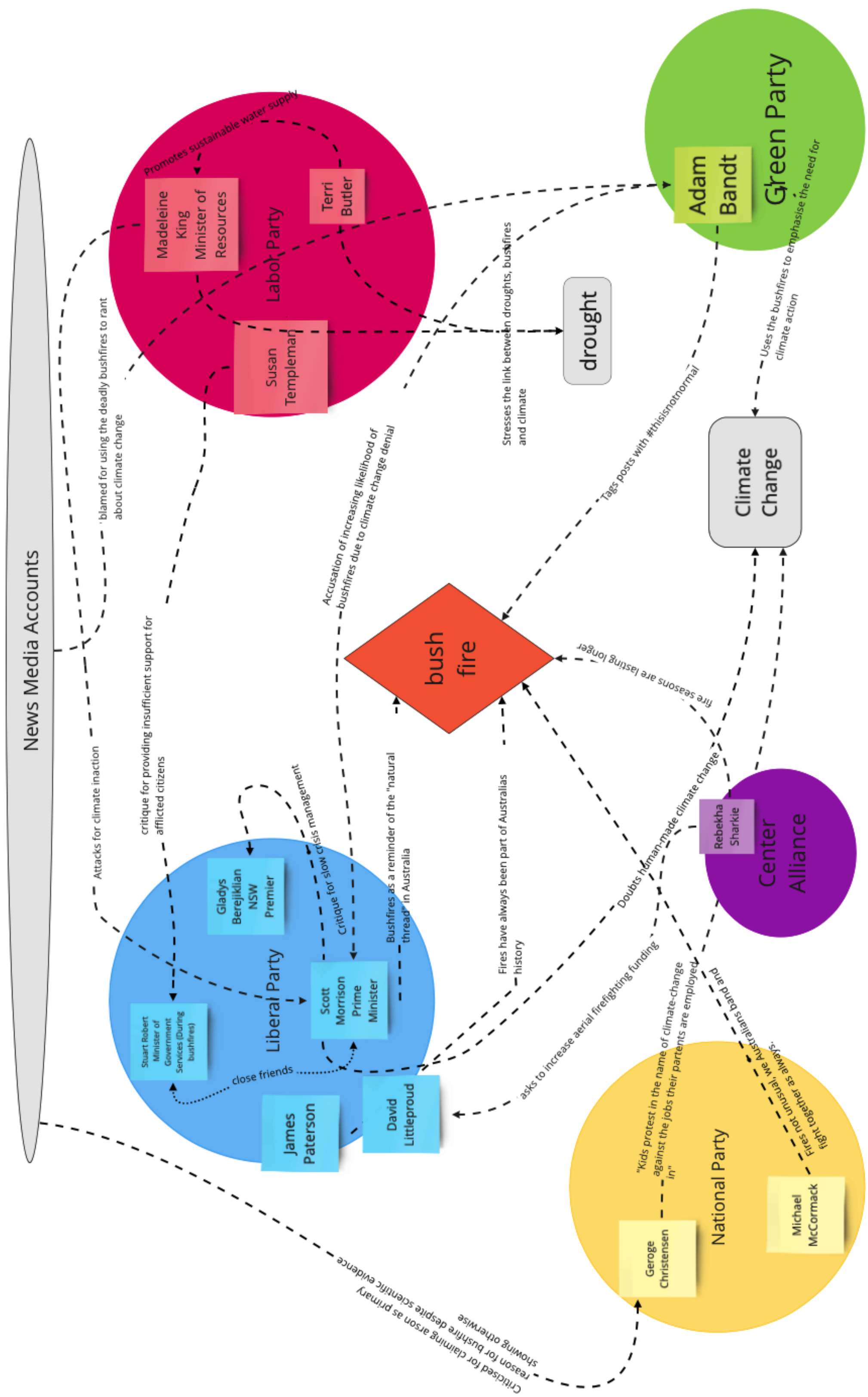
Figure 2: Retweet Network



Note: The colour of the nodes indicates party affiliation. Red: The Labor Party. Blue: The Liberal Party. Yellow: The National Party. Pink: The Liberal National Party of Queensland.

[2] The manual network can also be accessed interactively here: https://miro.com/app/board/o9J_lBlkOOY=/

Figure 3: Manual Network



Note: Edges are based on netnographic observations from various sources on the internet.

The most obvious pattern from our analysis can be found in the apparent divide between the Labor Party and the more conservative parties such as The Nationals and the two liberal parties in the retweet network. This demonstrates two circumstances in particular that we substantiate through qualitative observations from our netnographic endeavors. Firstly, the major parties in the Australian Parliament can be clearly divided into two political camps. In an interview, Anthony Albanese, leader of the Labor Party and most central node in our retweet network, criticises an increasing polarisation in the Australian Parliament and a lack of willingness to compromise. A condition that is frequently called "culture wars" in news media accounts and refers to MPs "...attacking each other with a peculiar viciousness, usually over remarkably minor issues." Accordingly, the pattern found in our retweet network is also reflected in our netnographic ground work.

A second interesting aspect in the retweet network can be found not in the presence of specific patterns but in the absence thereof. While the most represented parties in the parliament are also represented in the retweet network, the two smallest parties, The Greens and the Center Alliance, are not retweeting, or being retweeted, at all. This points towards two different conclusions. On the one hand, the absence of the Center Alliance in the retweet network is not surprising since they also did not play a dominant role in news media accounts or other social networks. Therefore, their absence rather seems to be a reflection of the parties inability to shape the public discourse on the Australian bushfires, possibly because of the few MPs of the party. On the other hand, the absence of the Green Party potentially shows a limitation of the automated network analysis which does not seem to capture one of the most dominant actors in the public discourse, simply because others do not interact in a specific way and on a specific platform.

The salience of The Greens in the bushfire discourse can be inferred from our netnography and manual network but also an exploratory automated semantic network analysis identified Adam Bandt as an MP with high degree centrality (see Appendix A), suggesting the use of similar terms as other MPs. Therefore, it seems paradoxical that none of his tweets that we categorised as bushfire-related got retweeted by any other MP. Our netnographic research depicts a possible explanation for this, as Adam Bandt's dominant demeanor also evokes criticism in the news media. He is frequently accused of taking political advantage of the deadly bushfires by pushing for climate action. Following his tweet (see Figure 4 below), a large Australian newspaper ran an article titled "We need to quit coal and cut pollution: Greens MP is blasted for using the deadly bushfires to rant about climate change".

Adam Bandt also attributes some responsibility of the bushfires directly to single actors when he says on his homepage "Scott Morrison will have to explain to anyone affected by bushfires this summer why he has increased climate pollution and made Australians' lives riskier". This is in stark contrast to the general tone of the communication surrounding the bushfires that emphasises solidarity between actors and the need to stay close together during times of crises. Not being retweeted might therefore be an indication that other politicians do not want to endorse his statements because they do not want to be seen as not empathetic through being associated with the vocal demands for climate action, which is a rather controversial and potentially dividing topic in Australian politics. On that note, David Littleproud, Emergency Management Minister, said it was not the time for climate change debates and to politicise this topic.

Figure 4: Tweet by Adam Bandt

A different angle originating from our immersion journal relates to Michael McCormack, leader of the National Party and Deputy Prime Minister of Australia. McCormack stands out with a very clear framing of the bushfires as being “not unusual or not uncommon to Australia; nor are droughts and nor are floods. But Australians get through these because we are a resilient bunch. We stick together, we band together". On the one hand, the first part of the statement illustrates clearly that he does not believe that climate change contributes to the more severe and prolonged bushfire seasons. He rather frames the fires as normal for Australia whose people historically got used to these conditions. As a reply to this, a meme on Facebook got a lot of attention, showing McCormack downplaying the increasing scope of the Australian bushfires (Figure 5).

Figure 5: McCormack Meme

On the other hand, the second part of the statement emphasises the dominant political tone that promotes togetherness and cohesion between Australians. This is further underlined by McCormack’s position in the retweet network, where his node is one of the few that has edges to members of two different parties. Grounding this information with our immersion, he retweets a tweet by PM Scott Morrison who states “Despite the many challenges we face, especially the terrible bushfires, the drought the floods, today is when we can reflect upon celebrate our great nation. Today we can all come together as one give thanks for being Australian". He also endorses actions taken by David Littleproud by retweeting his tweet about additional recovery grants for Australian farmers. Despite his seemingly inter-party interwovenness, it must be noted that his retweets are bound by the two very apparent clusters in the retweet network. He does not interact with MPs from the Green or Labor Party on Twitter and restrains his activities to politically close actors and content-wise rather descriptive and uncontroversial statements.

Figure 6: Retweets by McCormack

Finally, another actor who is particularly salient in our manual network is Madeleine King, Shadow Minister of Resources and member of the Labor Party. Scouting the communication related to her reveals a different interesting topic, as her fellow party member Terri Butler calls on her to prioritise sustainable and high-quality water supply. Diving deeper into this topic shows how water supply and wildfires are interwoven. Firstly, increasingly long periods of drought not only lead to increased risks for wildfires but also leads to communities running short on drinking water simultaneously (and consequently also on water for firefighting). Secondly, especially New South Wales suffered from low quality drinking water since rainfalls carried high quantities of ashes that diminishes its quality. Other factors, such as destroyed ground cover and altered structure, behavior, and erosion of soil lead to further problems. Accordingly, we have decided to expand our list of key terms to find tweets that relate to communication surrounding the bushfire by the word "drought".

5 Content Analysis

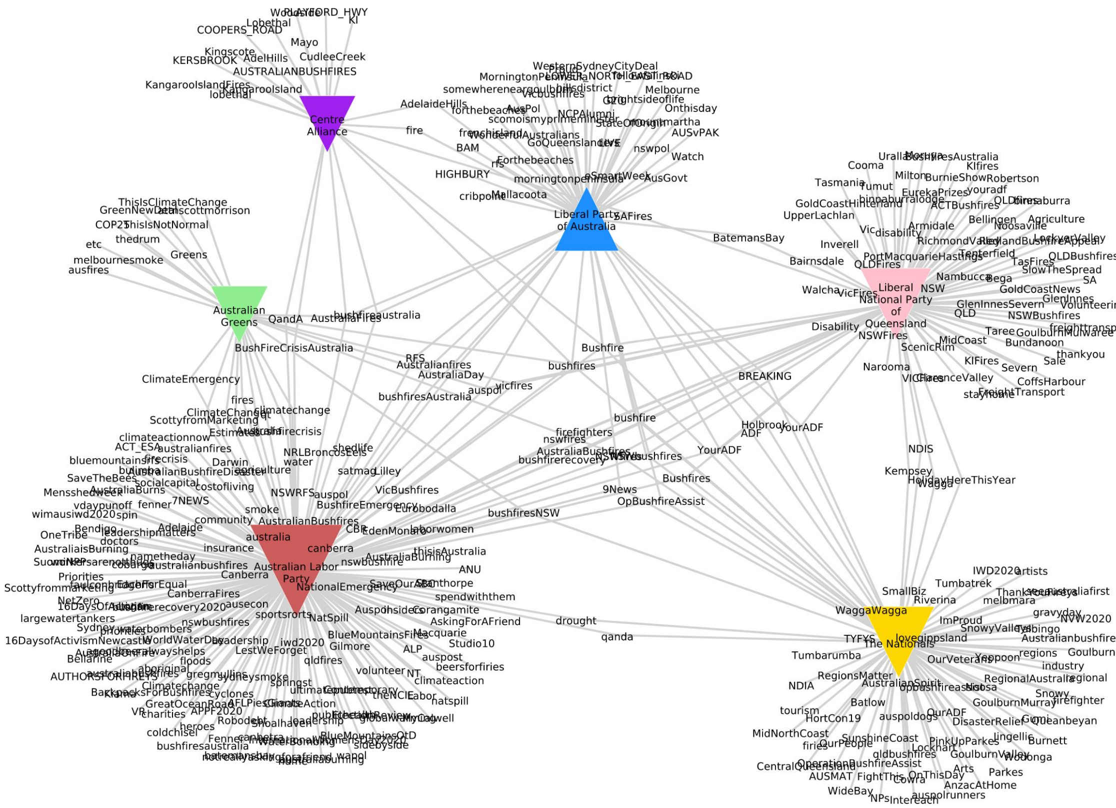
In the following section, we analyse the subset of tweets relating to the bushfires. We use techniques and methods from the field of quantitative text analysis, with a focus on semantic hashtag networks and word co-occurrences as indicators of political framing during crisis events. We supplement our quantitative analysis with extensive netnographic research, spanning social media platforms and news media sites, to ground and qualify the interpretation of quantitative patterns and observations from our Twitter dataset.

5.1 Party-Hashtag Networks

On many social media platforms, hashtags serve as a convenient indexing system, enabling users to sift through high volumes of text data to easily identify topics and conversations that they find interesting or relevant. Research shows that another key function of social media hashtags is that they act as vehicles to raise awareness of certain issues, to encourage public debate, and to facilitate the diffusion of ideas (Bonilla & Rosa, 2015). Wang, Liu, and Gao (2016) break down the spread of information by hashtags into two mechanisms: bottom-up mechanisms and top-down mechanisms. While bottom-up mechanisms are considered an autonomous process, driven by individuals and organisations from the ground, top-down mechanisms are driven by significant figures in the media or politics, with the aim of pushing certain information and agendas. The purpose of our socio-symbolic party-hashtag network is to explore how political parties in Australia leverage hashtags as a means to contribute to the formation of public attitudes and to frame the political agenda surrounding the bushfires.

Having explored several different compositions for the semantic hashtag network, we aggregate tweets at the party level, as this benefits the readability and therefore the interpretability of the network. As a result, instead of individual MPs representing actors in the bipartite network, it is political parties. This provides insights into the hashtags favoured by members of each party when tweeting about the bushfires. Some interesting observations can be made regarding the clustering of the hashtags around each political party, indicating that MPs from different parties tend to use hashtags to label themes and topics quite differently. Hashtags with the highest network degree centrality, and therefore with the highest use case between parties, tend to specify the bushfires directly, the top three hashtags by degree centrality being #bushfire, #NSWfires and #AustraliaFires (see Appendix B). Also frequently used by all parties is #YourADF, paying homage to the Australian Defence Force who were often the first response to bushfire outbreaks. Political communication, represented via hashtag usage, appears vastly heterogeneous between parties, the few hashtags that are homogeneous in the political sphere all relate directly to the crisis, presumably to raise awareness of the events as they unfold.

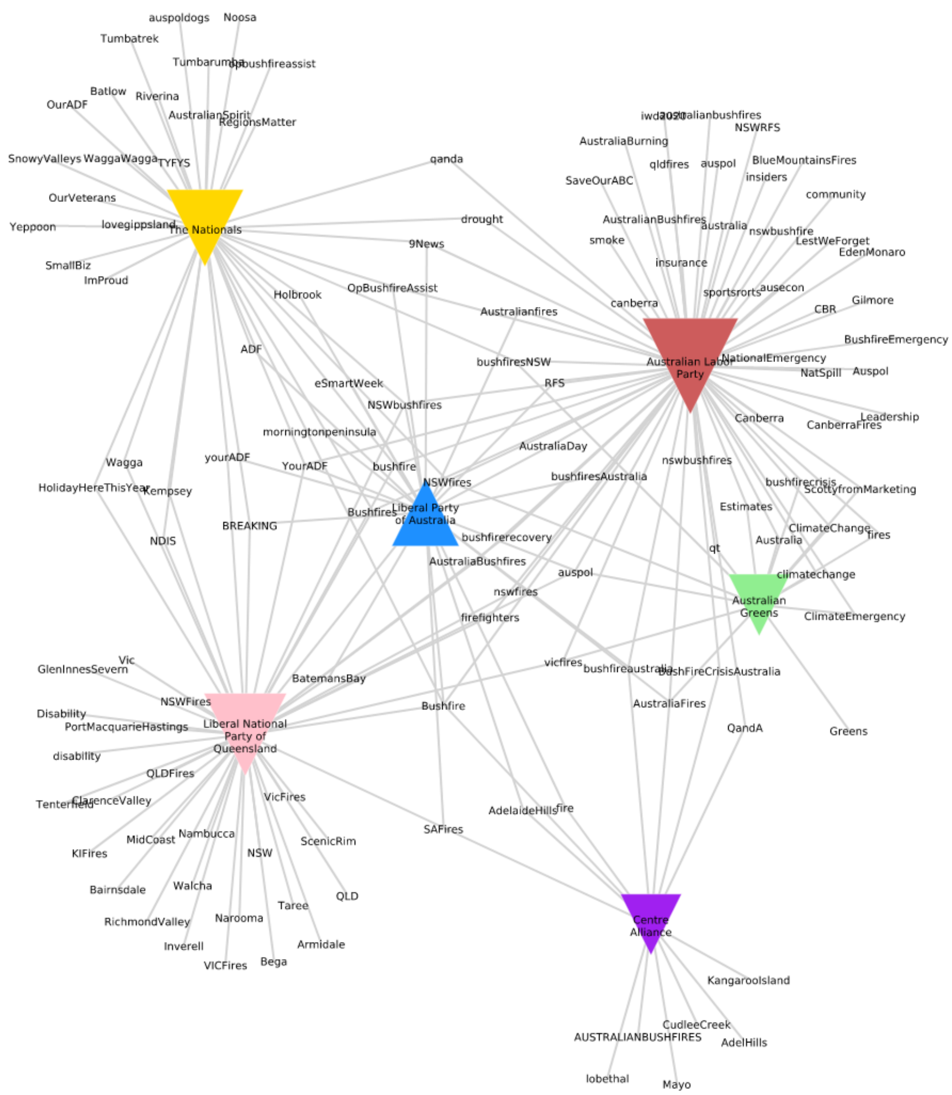
Figure 7: Party-Hashtag Network



Note: Upside-down triangles represent the opposition and upward pointing triangles the incumbent party. The Fruchterman-Reingold algorithm has been used to visualise the network.

To facilitate our qualitative exploration of the network and to help with the identification of emergent frames, we re-run the party-hashtag network, only keeping hashtags that are used more than once. The resulting network not only appears more readable, but the removal of single use hashtags also aids our enquiry, since prevalent themes and agendas that political parties wish to push will likely result in the repeat usage of hashtags. With this comprehensive mapping of actor-content based relationships within the domain of the bushfires, we employ netnographic exploration to gain a more deeply grounded interpretation of the observable patterns within the domain through the manual reading of tweets. The symbiotic relationship between computational analysis and netnography that we implement was inspired by the work of Krieg, Berning, and Hardon (2017), who use the vast body of informant-based semantics, afforded to them by keyword co-occurrence networks, to guide their ethnographic field work to qualify, contextualise and therefore properly understand their visualisations.

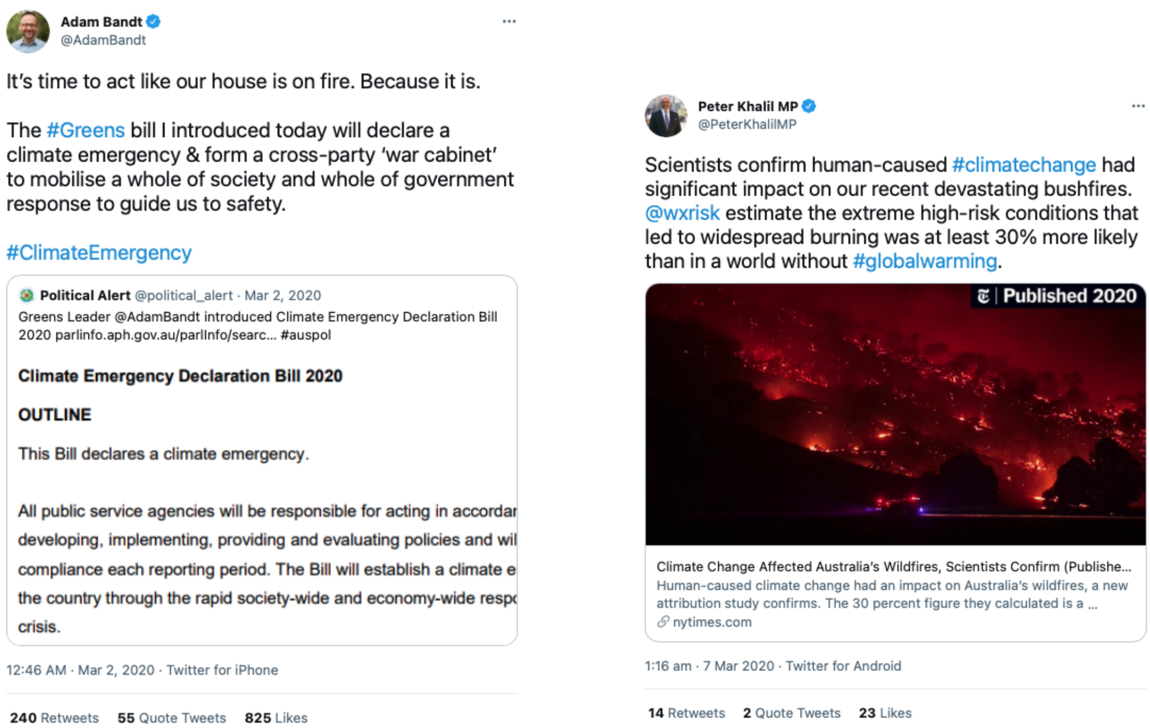
Figure 8: Party-Hashtag Network (excluding single use hashtags)



Note: Upside-down triangles represent opposition and triangle the incumbent party. The Fruchterman-Reingold algorithm has been used to visualise the Network.

Using the party-hashtag network above to navigate the qualitative reading of tweets, the most prevalent theme that we observe relates to the use of hashtags referring to the environment by members of both the Australian Greens and the Australian Labor Party, who frame the bushfire as a direct result of global warming. The hashtags #ClimateEmergency and #ClimateChange appear in a number of tweets by members of both parties, contributing to the formation of public attitudes on climate issues, while Green MP Adam Bandt even uses the climate framing to push the Green's own political agenda through the introduction of the Climate Emergency Declaration Bill to parliament (Figure 9). Members of the Australian Labor Party also use climate related hashtags to push the climate agenda, in the below example (Figure 9) using scientific evidence to back up their relational claim between global warming and the bushfires. Interestingly, none of the other political parties in Australia frame the bushfires as a climate issue, at least not through the use of hashtags. Yet, this is partly in line with expectations from our earlier netnographic endeavours, with climate change a widely contested topic at the very heart of Australian politics.

Figure 9: Tweets by Adam Bandt & Peter Kahil



Another way that both the Australian Greens and the Labor Party leveraged the bushfire for political gain was through unifying behind the derogatory #ScottyFromMarketing, to launch attacks on the Prime Minister who they feel has been underwhelming in his response to the crisis. This hashtag seems to suggest that Morrison is not qualified for the position as prime minister based on his previous employment as the managing director of Tourism Australia and Tourism New Zealand prior to that. Growing frustrations over the government's handling of the bushfire caused the use of the hashtag to circulate widely in the public sphere, eventually drawing comments from the prime minister himself, who accused the Labor Party of running a smear campaign and criticised what he deemed taking political advantage of a crisis.

The Nationals also leverage the crisis to promote their own political interests, adopting a fiercely patriotic tone in their Twitter communication in response to the bushfire. Hashtags such as #AustralianSpirit, #ImProud and #TYFYS (thank you for your service) express a strong national identity and seem to rally-round the flag as well as the service men and women on the front lines. Through deeper reading of communication by the Nationals, a trend quickly emerges, with an onus on getting National MPs out speaking to people in badly affected areas and commending the bravery, resilience and generosity of rural Australians in response to the bushfires. This kind of framing could be interpreted as an attempt to gain popularity amongst their target audience, since historically the party represents farmers, graziers and other rural communities.

Two potential shortcomings associated with the party-hashtag networks are formulated by Campagnolo (2020), when exploring the benefits of situational analogue

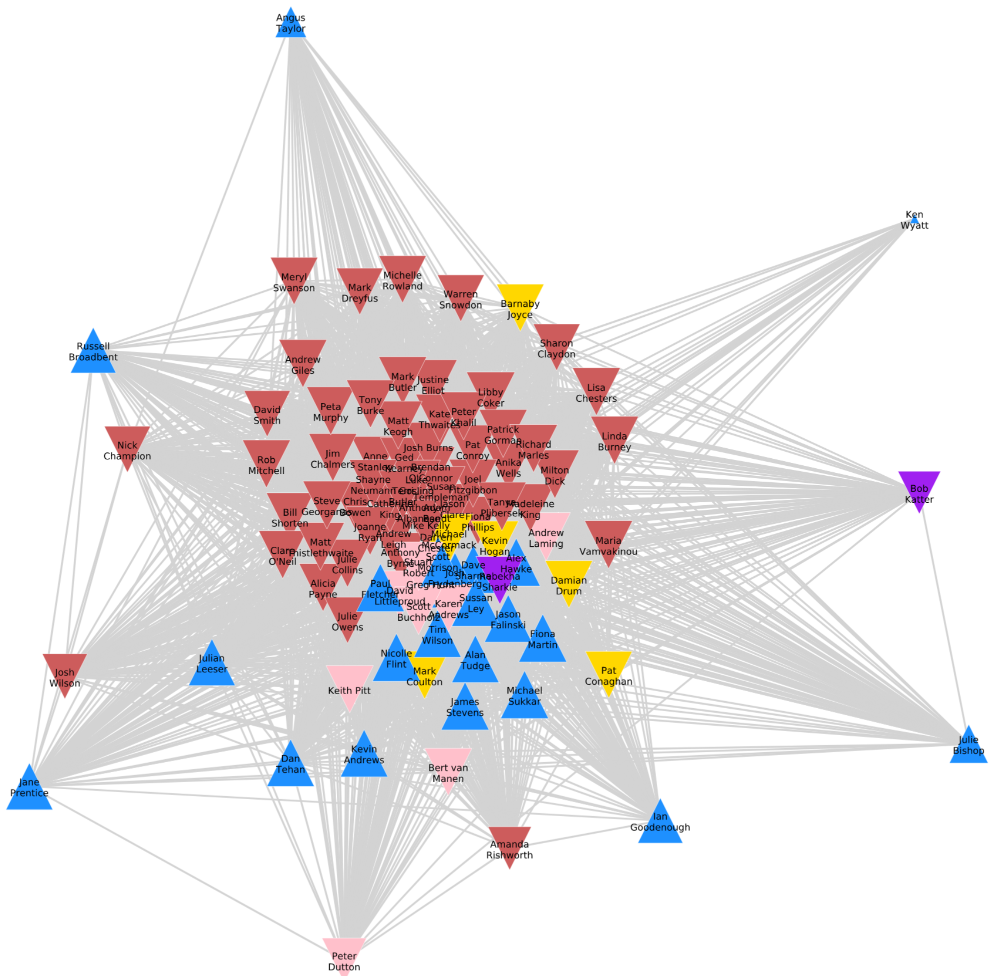
Figure 10: Tweet by Susan Templeman

mapping. Firstly, when using hashtags as a unit of analysis, a sizable limitation pertains to the asymmetric use of hashtags among actors. Politicians that favour using hashtags to promote certain agendas will be represented in our quantitative networks, yet those who do not use them will not be captured in our analysis. This will likely create a one-sided picture of socio-semantic relations and create a discourse map with sizable gaps in. When interpreting the semantic networks at face value, there is no way to sense the areas where a lack of data makes the networks incomplete. Secondly, questions may arise regarding the "naturalness" of data generated by social media due to platform effects. It is difficult to untangle whether politicians adopt certain hashtags deliberately or whether they are influenced by hashtags that are currently trending. This is a prime example of *algorithmic confounding*, whereby digital media does not simply facilitate the spread of social phenomena, but actively participates in its production (Salganik, 2019). These shortcomings essentially reinforce the need for the combination of qualitative readings, immersion, and manual network mapping to partially plug the gaps caused by a lack of data and to create a more complete, and thus more meaningful mapping of political communication.

5.2 MP-Semantic Network

Another way of studying the MPs' socio-symbolic constellations (as per Fuhse, Stuhler, Riebling, and Martin (2020) is through their similarity in word usage. This is what we visualise in Figure 11, where we once again study political communication at micro-level, where MPs represent nodes and edges represent shared symbolic practices. In practice, this means that if two politicians use the same word, an edge is drawn between them in the network.

Figure 11: MP-Semantic Network based on Word Co-occurrence



Note: Upside-down triangles represent opposition and triangle the incumbent party. Parties are colored as Labor(red), Liberal Party (Blue), The Nationals(Yellow), The National Party of Queensland(Pink), The Greens(Green). The Fruchterman-Reingold algorithm has been used to visualise the Network.

The figure shows that political party and symbolic practices are linked together in clusters of communities. If we use the big cluster of MPs in the middle of the network as a starting point for our analysis, we see that MPs from the Labor Party cluster together with only few MPs from other parties.

The most distinctive exception is - paradoxically - the far right politician Barnaby Joyce. Joyce is a notorious hard-hitter and former leader of the National Party, which is why his entanglement with the Labor Party is unexpected. During the time of the bushfire Joyce was center of attention, as he caused a scandal during the bushfires by publicly blaming the Green party for the death of several bushfire victims in New South Wales due to the failure to support hazard-reduction burning during winter. Reading his tweets, it is clear that Joyce uses negative campaigning as a communication strategy to ridicule the policy of opposing parties.

We believe that this frequent interaction with his political opponents is responsible for his positioning in the network, since he is likely to dwell on the same issues and topics relating to the bushfire as them. This increases the propensity to use the same terms, although with a completely different meaning. Naturally, this is not caught by the underlying statistical model which relies solely on the co-occurrence of words, thus highlighting the need for qualitative assessments of anomalies.

The co-occurrence network, however, underpins what we already have shown in the two previous networks: There is considerable heterogeneity in political communication between blocs but homogeneity within the blocs. Again, this is hardly surprising if it was not for the fact that these patterns seem robust, even during a natural disaster. This suggests that these practices are deeply embedded into all aspects of Australian politics.

Furthermore, there are some extreme cases which at first glance might seem interesting but that one should refrain from trying to explain. The placement of outliers such as Amanda Risworth, Russel Broadbent, Angus Taylor, Ken Wyatt and Bob Katter is simply due to statistical noise as they have not tweeted more than one or two times about the bushfire. To remain transparent, we have kept them in the analysis although their position is without substantial meaning.

6 Advanced Social Data Science 2

In this section, we present the analysis conducted for ASDS2. We begin by briefly describing text-preprocessing steps. Next, we present our results from topic modelling. Following that, we create a subset of bushfire tweets using a computer assisted approach. Lastly, we fit a multinomial logistic model to predict party labels to analyse the words that are most predictive of the different parties.

6.1 Text Preprocessing

Denny and Spirling (2018) show that different preprocessing specifications have extensive effects on the final analysis and is therefore not to be neglected[3]. We follow the preprocessing steps from (Slapin & Proksch, 2008) (extracted from the `Wordfish` documentation as in Denny and Spirling (2018)):

- Lowercasing all text
- Removing all punctuation (including emojis and special characters)
- Stemming text with the `PorterStemmer` from `nltk`
- Removing all numbers
- Removing all English stopwords from the `nltk` dictionary

In addition to the steps above, we also transform the tweets to a numerical representation in a *document-feature-matrix* (DFM) and remove very frequent and infrequent words and group tweets by MP and date. For all unigrams, we also include corresponding bigrams.

[3] The authors also propose using the software `preText` to assess the sensitivity of different preprocessing procedures. However, as there is no viable `Python` implementation (that we are aware of), this is beyond the scope of this assignment.

6.2 Topic Models: LDA and hSBM

We employ two different variants of topic models, namely a simple Latent Dirichlet Allocation (LDA) and a hierarchical Stochastic Block Model (hSBM), as proposed in Gerlach et al. (2018). The LDA is a probabilistic model aimed to compute the hidden structure of topics from which a set of documents is assumed to be generated from (Blei, 2012). However, it relies on a Dirichlet distribution in the estimation of topics which is known not to be aligned with Zipf's law for the frequency of words (Gerlach et al., 2018). Furthermore, there is no theoretical solution to set the right numbers of topics for the hyperparameter k (van Atteveldt, Trilling, & Arcíla, 2021). The hSBM is a network based approach to topic modelling, in which the DFM is defined as a bipartite network and has been shown to perform better than the LDA on both real and artificial data (Gerlach et al., 2018). Topics are treated as a latent structure in the network and are inferred through probabilistic community detection. Furthermore, the hSBM is non-parametric and, therefore, circumvents LDA's problem with handling topics from a non-Dirichlet distribution (Gerlach et al., 2018). To determine k for the LDA, we rely on the number of topics automatically detected by the hSBM through its posterior distribution.

6.2.1 Topic Model Results

As we are dealing with an unsupervised model, human validation is essential before we move forward. We establish face validity by inspecting words from the topics of interest. Fitting the hSBM on the 50,081 tweets from the period between June 2019 and June 2020, we find a total of 232 topics in the first level hierarchy and 42 in the second level. After reading all the topics, we asses that one topic found in the second level hierarchy is useful, as the most important tokens rather unambiguously seem to be about the bushfire. The 20 most important tokens from this topic are displayed in Table 2. We also include a topic related to climate change which we later touch on in Section 7. Next, we fit the LDA model to the same data and set the amount of topics to $k = 42$ based on the results from the hSBM. We regard this as reasonable value for k because the hSBM second level hierarchy topics seem to display a rather distinct set of interpretable topics.

The results from the LDA model are at face-value slightly more difficult to interpret. We find that three topics relate to the bushfire but contain tokens that are less obvious and need to be qualitatively assessed through our netnography. As we are mainly using the topic models as an initial exploratory tool, we anticipate, that the hSBM will be more useful and consequently, we choose to focus on these results. There are two important things to note here. First of all, it could be that the LDA actually picks up more heterogeneous discussions revolving the bushfires, e.g. relating to differences in communication influenced by party affiliation. The hSBM results could, in this sense, be problematic as they would not reveal certain aspects of the public discussion. Secondly, this also highlights a more general concern relating to model dependency, i.e. that the results are highly dependent on somewhat arbitrary modelling choices. Against this background, we ensure to not base conclusions solely on the topic models but rather through the triangulation possible by the complementarity of quali-quantitative methods used in our exam.

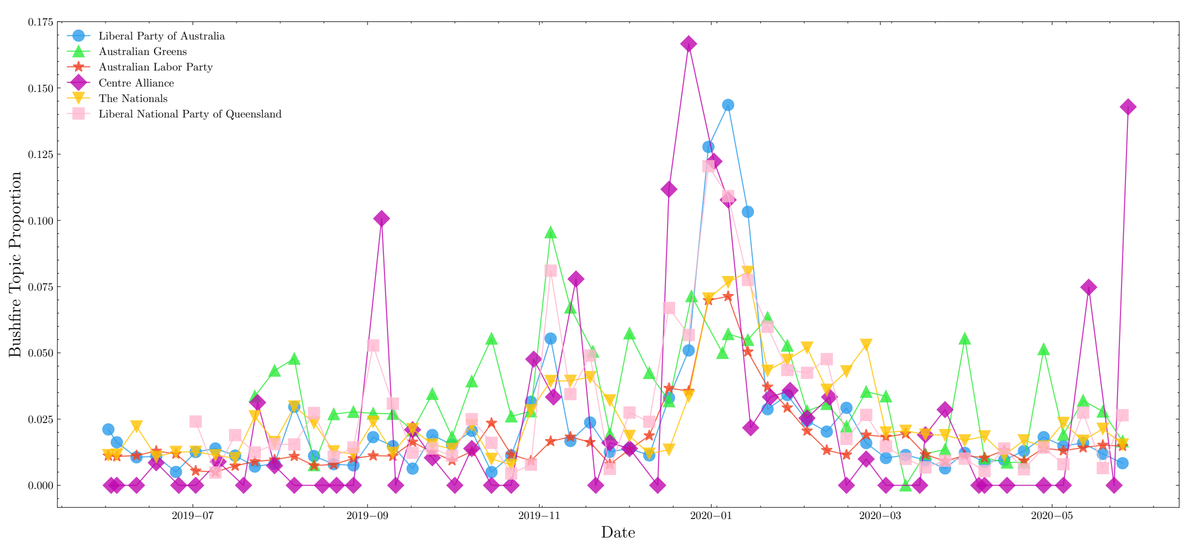
Table 2: Topic Model Results

Model	Topic	Tokens
hSBM	Bushfire	bushfir, fire, emerg, impact, affect, area, recoveri, recent, base, devast, relief, condit, immedi, danger, burn, toward, declar, environ, climat_emerg, summer, spot, bushfir_crisi, bushfir_affect
hSBM	Climate	chang, climat, action, plici, climat_chang, emiss, green, energi, global, coal, gov, target, price, renew, zero, au, net, clean, foreign, warm, ineq, action_climat, reduct, suffer
LDA	Bushfire 1	fire, volunt, bushfir, catch, tank, firefight, commun, to, town, resid, local, great, near, area, today, servic, australia, help, your evacu
LDA	Bushfire 2	busi, via, small, bushfir, fire, small_busi, canberratim, via_canberratim, impact, interview, warn, local, confer, govern, climatechangg, area, sever, pm, smoke, press
LDA	Bushfire 3	drought, bushfire, farmer, help, sport, affect, save, fire, farm devast, disast, commun, money, rort, fund, grant, relief, natur, flood, assist

Note: Table shows top tokens from topics found in the second-level hierarchy of hSBM and LDA with $k = 42$. Both models are fitted with the same DFM of shape $N = 50081$ and vocabulary $V = 10970$. N are all tweets from MPs in the period 2019/06/01 – 2020/05/31. The dimensionality of V was reduced by removing words occurring less than 10 times.

As a second step in evaluating the hSBM results, we plot the weekly average proportion of the bushfire topic, aggregated on party and in relation to all other 41 topics. Because of very few observations, we leave out Katter's Australian Party from the analysis. As shown in Figure 12, a clear spike in the time-series occurs during December 2019 until beginning of February 2020. This helps to substantiate that the model has indeed been able to pick up the topic, as these where the most intense months of the bushfire. Surprisingly, however, is the proportionally low amount of bushfire tweets by the Green party during this period. This contradicts our understanding about the Greens high Twitter activity relating to the topic. One possible explanation is that the Greens use different words when talking about the bushfires and this is not picked up by this topic. Lastly, we can also see from the plot that some parties are more consistent than others. This is likely due the Greens and Center Alliance having comparably fewer MPs and consequently fewer observations that are being aggregated and, therefore, have more pronounced spikes.

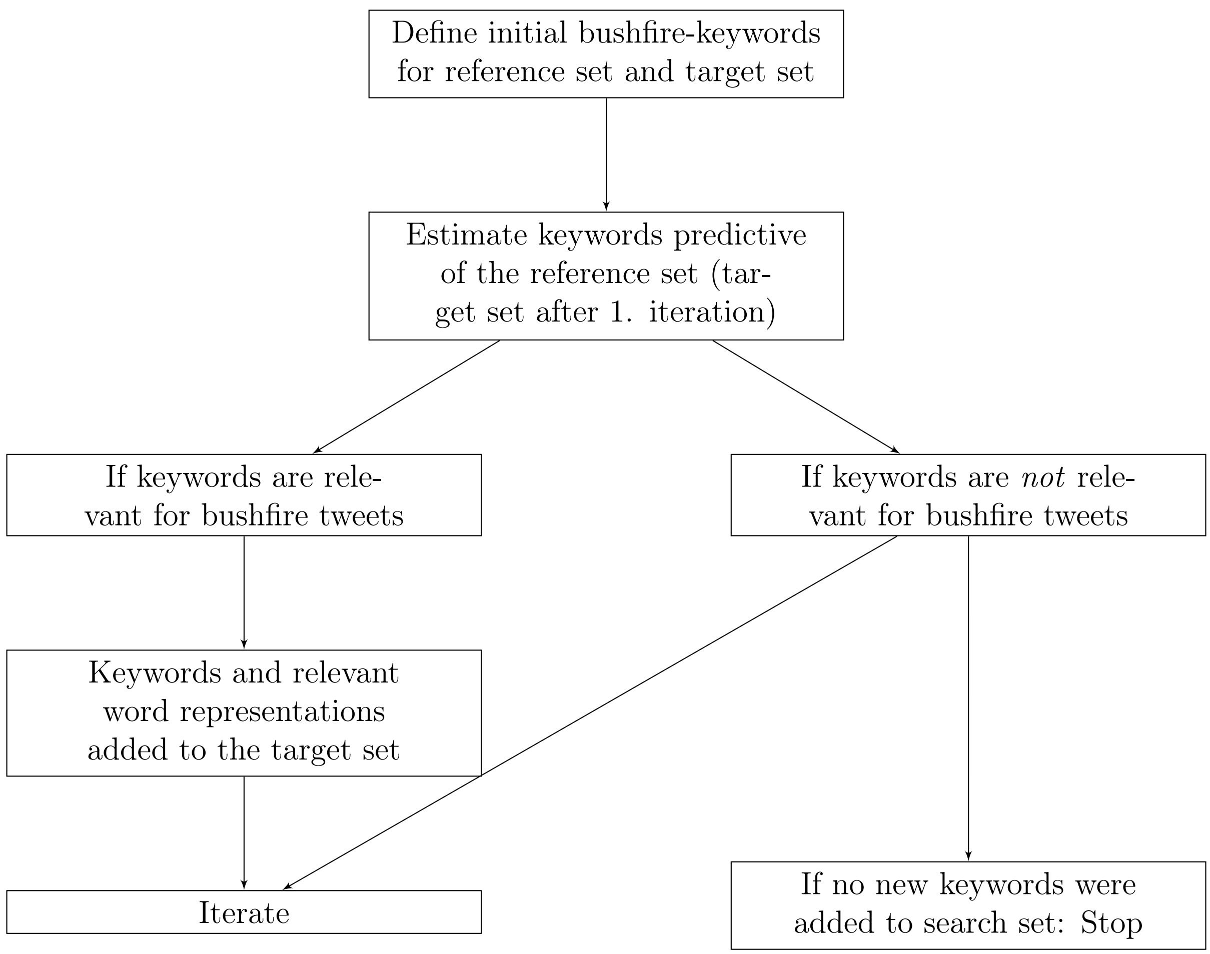
Figure 12: hSBM Bushfire Topic Time-Series



6.3 Computer Assisted Keyword Retrieval

To subset our dataset into bushfire-related tweets, we employ a computer-assisted keywords model as in King et al. (2017). Therefore, we split our dataset in a *reference set* and a *search set*. For the initial reference set, we decide on the following words from the topic models as rather unambiguous keywords relating to bushfires: bushfir, bushfir_crisi, bushfir_affect, and firefight. The rest of the dataset serves as the search set from which our goal is to detect the *target set*, which is a new subset of bushfire-tweets. For the initial target set, we decide on the following words from the topic models: fire, disast, recov, emerg, wildlif, and nsw. Terms like disast, recov, and emerg are used for the search and not the reference set, as they potentially refer to COVID-19 related tweets as well. From our netnography, we know that "nsw" refers to *New South Wales*, a region hit particularly hard by the bushfire.

Figure 13: Schematic approach of computer assisted keyword retrieval



We then use an ensemble of two models, näive bayes and logistic regression, to find other keywords that are predictive of documents in the reference set. Presented with these keywords, we make a qualitative assessment of whether or not they are likely to describe a tweet about bushfires from the target set. In an extension to this, we have implemented the word embeddings model `FastText`, developed by Bojanowski, Grave, Joulin, and Mikolov (2017) and pre-trained on English Wikipedia articles, which is used in cases where we believe a suggested keyword is likely to describe a tweet from the target set. This model is trained on 16 billion tokens from various text sources and used to outline relevant word representations of the keywords which we, depending on their relevance, accept. We iterate this entire process five times, resulting in an exhaustive list of keywords. As earlier mentioned, King et al. (2017) show that it is difficult for humans to recall topic-related keywords and computers might be better at such task. Following the outlined process, we also experienced this, as the model ended up suggesting terms such as *conflagration* and *aerial* which are very relevant but very unlikely for us to come up with.

On the basis of the final list of keywords resulting from our model, we made a qualitative assessment of each term and subsetted the list further to keep a relatively long list of keywords to detect bushfire-tweets. This list was the used for a simple Boolean query that is subsetting our documents to tweets containing one or more of the terms. Following the Boolean query, we evaluated the performance by manually coding a sample of 400 random tweets as either being about the bushfires (1) or not (0). This resulted in a precision score of 0.75 ($ci : 0.71 - 0.79$). As stated in King et al. (2017), one of the benefits of the Boolean query is that it is much faster than statistical classifiers and can, therefore, be substantially improved with little effort by continued refinement of the used keywords. As we were not satisfied with the performance of the initial results, we made an additional query with even fewer keywords (see Table 3). Furthermore, we conditioned on a list of "negative" keywords, making sure that tweets using COVID-19 terms were removed. 2976 tweets contained one or more of the keywords.

Table 3: Boolean Keywords

Boolean Query Type	**Keywords**
Positive (True)	fire, firey, firefight, firefighters, firemen, fireman, bushfir, bushfir_crisi, bushfir_affect, blaze, flames, conflagration, ash, smoke, aerial, burn opbushfireassist, nswfires, nsw, blacksummer, two-alarm, three-alarm, four-alarm, five-alarm, habitats, wild-life
Negative (False)	covid, coron, rona, pandemic, epidemic, flu, virus, vaccin

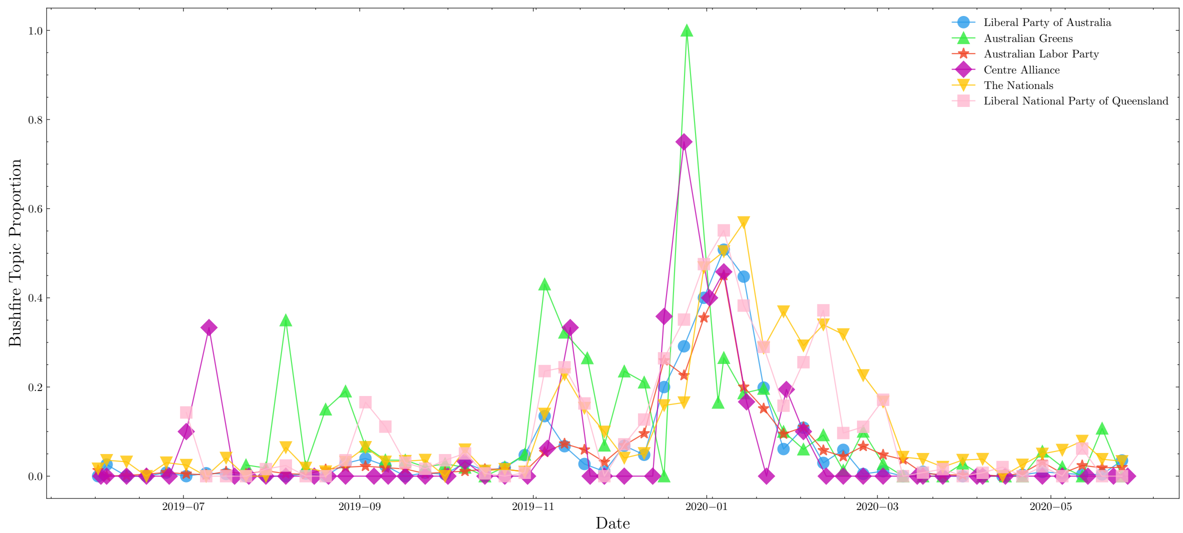
Note: A "positive" Boolean query refers to a query where tweets are extracted if they contain the corresponding keywords, whereas tweets containing the keywords in the "negative" Boolean query are removed.

Following the same procedure from before - but with a new random sample - resulted in a high precision score of 0.92 (ci: 0.89 - 0.94). However, this level of precision comes with a price. We do not know the true number of positives but as the query gets simpler, the recall will correspondingly decrease. In other words, we are expecting that the vast majority of our final subsample of tweets is about the bushfire, although an unknown amount of actual bushfire-tweets are omitted. If some type of tweets are systematically omitted, our analysis will potentially suffer from various kinds of biases which we discuss further in Section 7.1. However, we still believe that it is worth focusing on high precision despite the lower recall, as most tweets regarding the bushfires explicitly mention the bushfires.

6.3.1 Bushfire Time-Series

To compare the topic model results to the bushfire subset, we create a dummy variable for if a tweet in the full dataset during the bushfire period is about the bushfire, as defined by our Boolean query. Next, we aggregate the data in the same way as our topic model in Figure 12 and calculate the average for the dummy, which then represents a proportion of bushfire tweets per party and week.

Figure 14: Bushfire Subset Time-Series



Interestingly, the overall patterns are fairly similar to the topic model but with the Greens being a clear exception. The proportions, however, especially during the most intense period around January, are substantially larger than what the topic model suggested. The largest proportion given by the hSBM was about 17% for the Center Alliance in January while in our bushfire subset, all parties have a proportion of around 50% or higher. Most notably, all the tweets from the Greens are estimated to be about the bushfire for the first week of January. As we have validated the accuracy of the bushfire subset using random sampling and manual reading, we trust that the bushfire subset is represents a more accurate description of reality than the topic model.

To more formally test the similarity between the hSBM output and the bushfire subset, we also computed the within party correlation of the proportions, presented in Table 4. All coefficients are statistically significant at $p < 0.001$ and unsurprisingly, the party showing the weakest correlation is the Greens.

Table 4: hSBM and Bushfire Subset Correlation

Party	**Pearsons-r**
Australian Greens	0.658^{***}
Australian Labor Party	0.950^{***}
Centre Alliance	0.704^{***}
Liberal National Party of Queensland	0.915^{***}
Liberal Party of Australia	0.935^{***}
The Nationals	0.918^{***}

Note: $***p < 0.01$,$**p < 0.05$, $*p < 0.1$.

6.4 Predicting Party on Word Usage

As a last step in our analysis for ASDS2, we implement a classification model where we predict the party of a tweet based on the text. The purpose is to determine the words that are predictive of a given party. Our approach builds on a paper by Wu, where she uncovers gender biases on a professional forum for economist by looking at the words in a logistic regression that are predictive of a forum post being about women (Wu, 2018). Since we face a multiclass problem, we instead fit a multinomial logistic regression to our data and calculate the *average marginal effect (AME)* of each token k and party j by:

$$AME_{jk} = \frac{\partial Pr[y_i = j|\mathbf{w}_i]}{\partial w_{jk}} = N^{-1}\sum_{i=1}^{N} p_{ij}\left(\theta_{jk} - \sum_{m=1}^{J} p_{im}\theta_{mk}\right)$$

where $\mathbf{w}_i$ is the vector of token frequencies for each observation and $\boldsymbol{\theta}_j$ is the estimated coefficient vector for each party of length equal to the vocabulary. We base our calculations of the AME on Cameron and Trivedi (2005) in depth overview and derivations of the Multinomial Logit and its marginal effects. The benefit of using AMEs as a measure of feature importance rather than, say, simply ranking the estimated coefficients by size, is that they at least in theory have a potentially substantive and intuitive interpretation as opposed to only relative one. The AME tells us, on average, how the probability of a party j changes as we add an additional count of the token k when the feature space is defined as the *document-feature-matrix* (DFM). Although other vectorial representations of text like the *term frequency–inverse document frequency* sometimes perform better in prediction tasks, for the purpose of interpreting θ and as pointed out by Gentzkow, Kelly, and Taddy (2019), the DFM should be preferred for interpretability.

Similar to Wu, we fit the model using $l1$-regularization to penalise non predictive words and reduce the risk of overfitting. Because of limited space, we restrict the results here to display the most predictive words for the Greens and the Liberal Party. First looking at the Greens in Table 5, several words are coherent with what we have seen in our netnography and semantic networks, while others are less obvious or clear. For example, that the token "climate emergency" is one of the most predictive words is in line with the Greens' framing of the bushfire. Although not certain, the high placement of the word "normal" could be explained by the phrase "This is not normal", commonly used by MP Adam Bandt.

Table 5: The Australian Greens Top 10 AMEs

Token	Australian Greens	Australian Labor Party	Liberal Party of Australia	The Nationals
green	0.19	-0.28	0.07	0.03
climateemerg	0.17	-0.12	-0.03	-0.02
heatwav	0.16	-0.11	-0.03	-0.02
letter	0.16	-0.12	-0.02	-0.01
normal	0.14	-0.10	-0.03	-0.02
case_bushfir	0.14	-0.10	-0.03	-0.02
arctic	0.14	-0.10	-0.03	-0.02
global	0.13	-0.09	-0.02	-0.02
incom_support	0.13	-0.09	-0.02	-0.01
climat	0.13	-0.03	-0.07	-0.02

Note: The table is sorted by the 10 largest average marginal effect (AME) estimates for the Australian Greens, derived from a $l1$-regularized multinomial logit model, fitted with 4 Australian parties as the response variable.

Turning the attention to the Australian Liberal Party, we find a quite different set of most predictive words, several of which correspond to insights gained through our qualitative work. As an example, the words "courage" and "spoke" could both be related to the Liberals Party's focus on cohesion by praising first responders, firefighters and volunteers. It is also interesting that words like "ensure" are highly predictive of the Liberals, as this might point to differences in communication related to being the incumbent party who might want to explicate or highlight which actions have been taken.

Table 6: The Liberal Party Top 10 AMEs

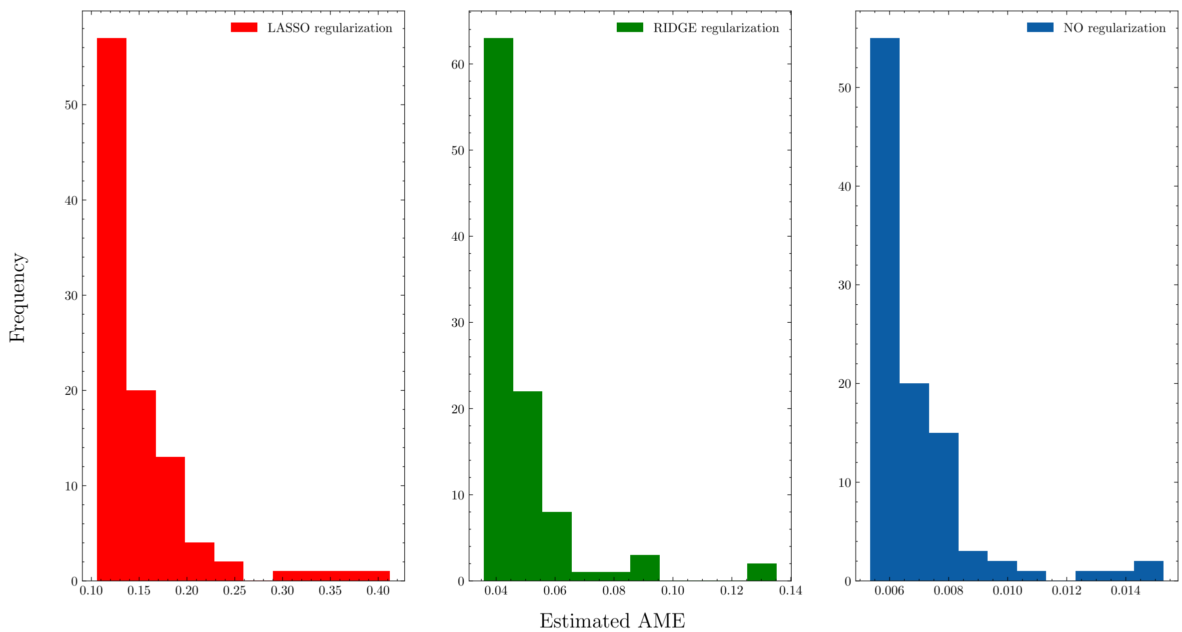
Token	Australian Greens	Australian Labor Party	Liberal Party of Australia	The Nationals
network	-0.03	-0.16	0.23	-0.05
ensur	-0.03	-0.16	0.23	-0.05
courag	-0.03	-0.15	0.22	-0.05
join	-0.02	-0.12	0.18	-0.04
spoke	-0.01	-0.17	0.18	0.00
addit	-0.02	-0.12	0.18	-0.04
good_idea	-0.02	-0.11	0.17	-0.04
death	-0.02	-0.11	0.17	-0.04
per	-0.02	-0.11	0.16	-0.03
new_year	-0.02	-0.11	0.16	-0.03

Note: The table is sorted by the 10 largest average marginal effect (AME) estimates for the Liberal Party of Australia, derived from a $l1$-regularized multinomial logit model, fitted with 4 Australian parties as the response variable.

6.4.1 Validity of Results

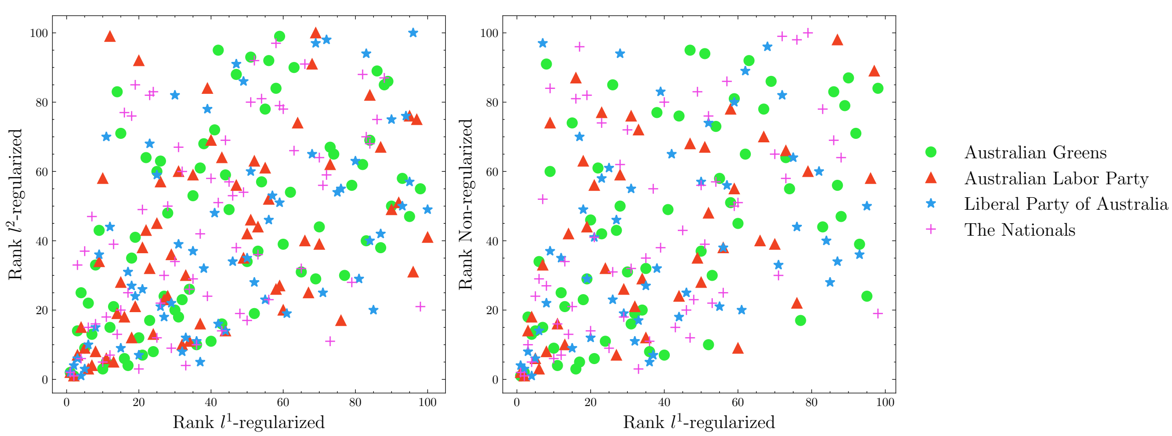
There are a couple of things to note about the results presented here. First of all, we could have measured the statistical uncertainty of these estimates either using the delta method or bootstrapping. However, when testing different model specifications, we find that making inferences about effect sizes is likely to be both uninformative and probably ill advised with the kind of high dimensional data that does not satisfy normality assumptions. In particular, introducing bias in the estimates by regularization dramatically changes the magnitude of our estimated AMEs. In order to highlight this issue, we plot the distribution of the 100 most important AMEs, estimated by a model with $l1$-regularization, $l2$-regularization, and no regularization. As displayed in Figure 15, the distributions are similarly shaped but the difference in magnitude is important.

Figure 15: Distribution of Top 100 AMEs across model specification



To further analyse this issue, we perform a rank-order correlation between the top 100 AMEs, comparing the lasso-regularized model to the ridge-regularized model and non-regularized model. The results are shown in Figure 16 and Table 7 and 8. What becomes clear is that not only the magnitude is heavily influenced but also the rank-order and even the words which correspond to the top 100 AMEs.

Figure 16: Top 100 AMEs Rank-order Correlation across Model Specification



Although the correlations are moderately strong between 0.4-0.53 for the first comparison and 0.46-0.57 for the second, we need to keep in mind that this is a correlation within the intersection of the top 100 AMEs from the models being compared. In Tables 7 and 8, we also show the amount of tokens that are in the intersection. In the comparison of the non-regularized and lasso-regularized, we find for example that only 45 token are retained. Reassuringly, however, the correlation is significantly stronger for the absolute top tokens, i.e. roughly around rank 1-10.

Table 7: Rank-order Correlation between $l1$ and $l2$ Regularization

Party	Spearmans-r	Token Count
Australian Greens	0.57***	73
Australian Labor Party	0.56***	65
Liberal Party of Australia	0.57***	62
The Nationals	0.47***	68

Note: $***p < 0.01$,$**p < 0.05$, $*p < 0.1$. Token count refers to the amount of tokens retained from the intersection of the sets of top 100 AMEs from both models

Table 8: Rank-order Correlation between $l1$ and No Regularization

Party	Spearmans-r	Token Count
Australian Greens	0.53***	63
Australian Labor Party	0.48***	45
Liberal Party of Australia	0.40***	52
The Nationals	0.46***	58

Note: $***p < 0.01$,$**p < 0.05$, $*p < 0.1$. Token count refers to the amount of tokens retained from the intersection of the sets of top 100 AMEs from both models

In summary, this suggests that although we should be careful when interpreting the effect size of the AMEs, the relative importance for the top 10 tokens is robust across model specifications. Moreover, $l1$ and potentially $l2$ regularization is likely preferred for additional reasons. First of all, it is common practice when working with high-dimensional text data (Jurafsky & Martin, 2020) and has theoretical validity in the sense that it penalises tokens that are unimportant for the prediction. Secondly, in our case, the length of $\mathbf{w}_i$ exceeds the amount of observations N, which is problematic as the estimating equation for the Multinomial logit is undetermined with no unique solution.

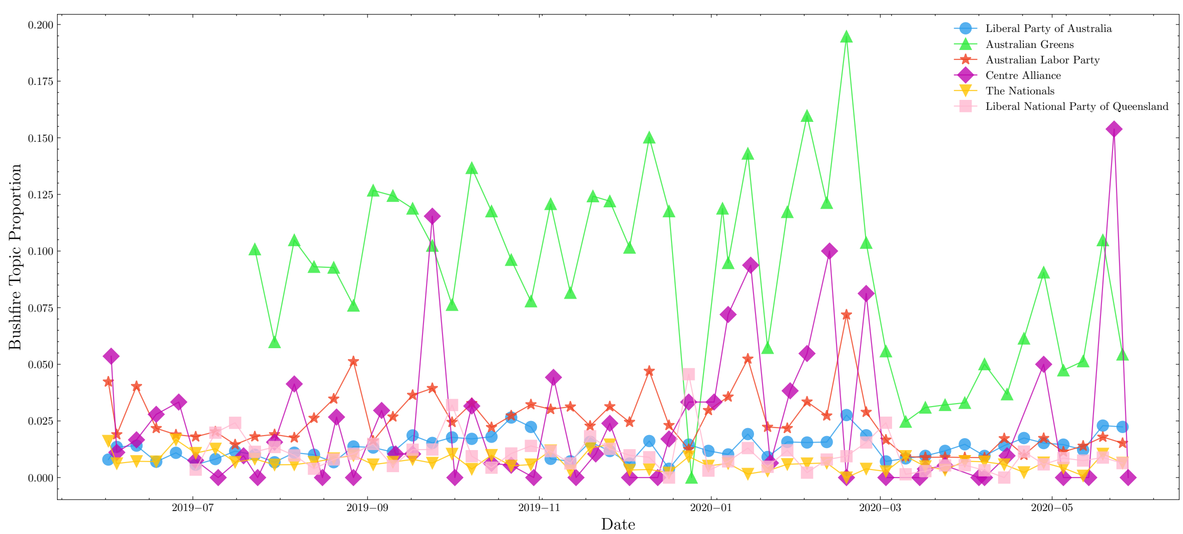
7 Quali-Quantitative Integration

In this section, we discuss the integration of qualitative and quantitative methods used in this study. We start by outlining complementary and diverging patterns observed in the data which subsequently lead to a discussion on biases inherent to different methods.

As described throughout the paper, a major challenge was the creation of an accurate subset of tweets relating to the bushfires. Considering the created timelines in Figure 12 and 14, we can infer that the prevalence of the bushfire topic and the number of classified bushfire tweets are coherent with the severity of the fires. This increases confidence in our dataset which is further grounded in the results of the topic models which clearly capture bushfire-related terms (Table 2). Moreover, the topic models capture different communication patterns that we also observed in our netnographic and network approaches. As clearly depicted in Bushfire topic 3, the words "help", "save", "money", "grant", or "relief", display a very dominant topic around social security; whereby MPs wish to assure those affected by the fires that the government cares about their losses and is willing to spend considerable resources on the recovery of the damages. This is also reflected in the tweets in Figure 6 and in several observations during the netnographic campaign, with different politicians tying to outdo each other in making generous promises regarding the amount of resources to raise for the recovery.

A diverging pattern we found regards the role of the Center Alliance in the public discourse on the bushfires. While our netnography, retweet network, and content analysis suggests that the party only played a minor role in shaping the agenda, the time-series analyses show that the proportion of tweets by the Center Alliance related to the bushfires increased significantly when the bushfires worsened. It even seems as though they played a role in setting the bushfire agenda since the peaks occur earlier than those of the other parties (see Figure 8). One possible explanation is that Rebekha Sharkie, who is responsible for the bulk of the communication on Twitter by the Center Alliance, is representing the Division of Mayo in south Australia, a rather small area badly hit by the fires, which might not have been covered extensively by national news media or by other MPs representing the state government. This can also be seen in the most-used hashtags in tweets relating to the bushfires, such as "#Kangorooisland", "#AdelHills", "#Mayo", or #CudleeCreek, which all refer to rather small regions in south Australia.

Figure 17: hSBM Climate Topic Time-Series



Lastly, controversies within and between the parties evolved when the framing of the bushfires as a climate-related crisis was at stake. From the topic models, it can be inferred that climate change was a dominant issue within the bushfire discourse. Its interwovenness with the bushfires is further highlighted by the presence of the word "climate emergency" in the bushfire topic of the hSBM model. Considering Figure

Figure 18: Jason Falinski

17 and the most predictive words of The Greens in Table 5, the climate topic seems significantly shaped by The Greens. However, as the Party-Hashtag network shows, some climate-related hashtags seem to bridge different parties in the network. Even though our netnography suggests that using the same words does not hint towards a similar attitude towards the topic, as illustrated by Barnaby Joyce's position in the co-occurrence network, there is also evidence that the otherwise greatly divided parties might have a tendency to converge in their discourse on climate change. Not only is the only bridge that connects the two political camps in the retweet network a tweet by a Liberal Party MP about the contribution of climate change to the bushfires that was retweeted by a Labor Party MP, Prime Minister Scott Morrison also states in an interview that "...there are many contributing factors that relate to these fires. The drought is obviously one, and the dryness of the bush is the biggest factor. And we all know... that climate change, along with many other factors, contribute to what is occurring today".

7.1 Using Qualitative Insights to Assess Quantitative Bias

A crucial part of both the quantitative and qualitative analyses is our final subset of tweets. The subset of tweets related to the bushfire was based on a simple Boolean query with the keywords listed in Table 3. These terms are carefully selected in order to maximise precision at the expense of lower recall. However, this also induces bias in our analysis as some tweets might systematically by omitted if they do not explicitly mention the bushfire. During our netnography we found that some tweets written during the bushfire did not have to state that they tweets were referring to the fires as it was implied by the context at the time. The following tweet by MP Mark Butler might be seen as such an example:

Figure 19: Tweet by Mark Butler

This was tweeted when the bushfire was around its peak (13/01/2020) and should be seen as a response to the fires and their link to climate change (this can also be seen in the linked article). However, this tweet is *not* included in our final Boolean query due to its subtle reference to the bushfire.

A consistent pattern in our analysis is the framing by the opposition of the bushfires as a matter of climate change. However, we believe that there is a risk that a lot of tweets linking the bushfires to climate change did it in a subtle way, e.g. by posting the newest report on the effects of climate change on drought (and thereby implicitly bushfires). If this is truly the case, our analysis of the climate framing potentially suffers from attenuation bias. That is we underestimate the severity of this framing strategy due to low recall rate. Although this problem is not to be neglected, we still believe that our approach of subsetting tweets in relation to the bushfire is the most suitable, as the majority of relevant tweets still seemingly mentions the bushfires explicitly.

7.2 Platform Dependency

Our entire quantitative analysis and large parts of our network analysis is based on Twitter-data. As Airoldi (2018) notes, digital experiences are partly formed by the social media sites different algorithmic logics, which gives rise to a relevant question: whether or not our results are platform dependent? A relevant comparison here is the MPs activity on Facebook which we have monitored during our netnography. At large, we experience that their is a high degree of homogeneity in the content that MPs post on Twitter and Facebook. Oftentimes, the MPs are making sure to post the same update on Twitter and Facebook.

On the other hand, it seems that some algorithmic confounding does take place on the two platforms when MPs are not creating the content themselves. Most notably, the

MPs do not retweet the same content on Twitter as they share on Facebook, as this is of course highly platform dependent. Therefore, it would be interesting for future research to compare retweet-networks with Facebook sharing-networks to map the differences in the political landscapes on Twitter and Facebook. Although the posts differ in their content, we still believe that the overall network structures are fairly robust across the two platforms.

Lastly, and perhaps most surprisingly, we find that Facebook seems to better support cross-bloc political interaction. Although the content that the MPs post themselves are similar on Twitter and Facebook, the latter appears to generate more interaction from political opponents, with other MPs often commentating on posts and receiving endorsement in the form of likes. One thing to notice in this regard is that rather than responding constructively to the initial post, the commentating MPs seems to worry more about making themselves visible.

8 Conclusion

In conclusion, the triangulation achieved through the combination of computational methods for text analysis and online ethnography yielded some interesting overarching results regarding the political framing of the Australian bushfires. Most salient across methods was a general tendency of homogeneous communication within the parties and heterogeneous communication between the parties. Furthermore, the bushfires seemed to raise two particularly prominent themes. Firstly, and especially prevalent among the conservative parties, MPs emphasised solidarity and resilience to increase cohesion and a feeling of togetherness. Secondly, climate change seems to be a prominent topic related to the bushfires. It remains to be seen whether the dominant rationale of climate change skepticism in the conservative parties undergoes a lasting change in light of these extreme disasters and whether this will be traceable with computational methods in the future or even reflected in policy-making.

Future investigations could build upon our findings by honing in on the climate debate in Australian politics, comparing the climate framing practices of politicians before and after the bushfires, thereby making it possible to analyse whether a sustainable transition has been initiated in the form of public communication practices.

References

Airoldi, M. (2018). Ethnography and the digital fields of social media. *International Journal of Social Research Methodology*, *21*(6), 661–673.

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

Blok, A., Albris, K., Carlsen, H., Møller, A., Isfeldt, A., & Enggaard, T. (2021). Inter-risk framing contests: the politics of issue attention among scandinavian climate ngos during the coronavirus pandemic..

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Bonilla, Y., & Rosa, J. (2015, January). #ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the united states. *American Ethnologist*, *42*(1), 4–17. Retrieved from https://doi.org/10.1111/amet.12112 doi: 10.1111/amet.12112

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *2010 43rd hawaii international conference on system sciences* (pp. 1–10).

Caliandro, A. (2018). Digital methods for ethnography: Analytical concepts for ethnographers exploring social media environments. *Journal of Contemporary Ethnography*, *47*(5), 551–578.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Campagnolo, G. M. (2020). Social data science xennials. In *Social data science xennials* (pp. 1–18). Springer.

Davey, S., & Sarre, A. (2020). *the 2019/20 black summer bushfires.* Taylor & Francis.

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.

Faulkner, B. (2001, April). Towards a framework for tourism disaster management. *Tourism Management*, *22*(2), 135–147. Retrieved from https://doi.org/10.1016/s0261-5177(00)00048-0 doi: 10.1016/s0261-5177(00)00048-0

Fuhse, J., Stuhler, O., Riebling, J., & Martin, J. L. (2020). Relating social and symbolic relations in quantitative text analysis. a study of parliamentary discourse in the weimar republic. *Poetics*, *78*, 101363.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–74.

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science advances*, *4*(7), eaaq1360.

Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (draft)* (3rd ed.).

Kerbleski, E. (2019). Trump, twitter, and the trees: A critical discourse analysis of donald trump's tweets surrounding the november 2018 california wildfires.

King, G., Lam, P., & Roberts, M. E. (2017). Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science*,

61(4), 971–988.
Kozinets, R. V. (2015). Netnography. *The international encyclopedia of digital communication and society*, 1–8.
Krieg, L. J., Berning, M., & Hardon, A. (2017, September). Anthropology with algorithms? *Medicine Anthropology Theory*, *4*(3). Retrieved from https://doi.org/10.17157/mat.4.3.458 doi: 10.17157/mat.4.3.458
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.
Manfredi-Sánchez, J.-L., Amado-Suárez, A., & Waisbord, S. (2021, January). Presidential twitter in the face of COVID-19: Between populism and pop politics. *Comunicar*, *29*(66), 83–94. Retrieved from https://doi.org/10.3916/c66-2021-07 doi: 10.3916/c66-2021-07
Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*.
Pelling, M., & Dill, K. (2009, May). Disaster politics: tipping points for change in the adaptation of sociopolitical regimes. *Progress in Human Geography*, *34*(1), 21–37. Retrieved from https://doi.org/10.1177/0309132509105004 doi: 10.1177/0309132509105004
Raalund, S., & Carlsen, H. (2021). Computational grounded theory revisited: from computer lead to computer assisted text analysis. (under review)..
Rafail, P. (2018). Nonprobability sampling and twitter: Strategies for semibounded and bounded populations. *Social Science Computer Review*, *36*(2), 195–211.
Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.
Slapin, J. B., & Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, *52*(3), 705–722.
Stone, D. A. (1989). Causal stories and the formation of policy agendas. *Political Science Quarterly*, *104*(2), 281. Retrieved from https://doi.org/10.2307/2151585 doi: 10.2307/2151585
van Atteveldt, W., Trilling, D., & Arcíla, C. (2021). *Computational analysis of communication: A practical introduction to the analysis of texts, networks, and images with code examples in python and r*. John Wiley & Sons. Retrieved from https://books.google.se/books?id=dDk9zgEACAAJ
van Vliet, L., Törnberg, P., & Uitermark, J. (2020). The twitter parliamentarian database: Analyzing twitter politics across 26 countries. *PloS one*, *15*(9), e0237073.
Venturini, T., Bounegru, L., Gray, J., & Rogers, R. (2018). A reality check (list) for digital methods. *New media and society*, *20*(11), 4195–4217.
Wang, R., Liu, W., & Gao, S. (2016, November). Hashtags and information virality in networked social movement. *Online Information Review*, *40*(7), 850–866. Retrieved from https://doi.org/10.1108/oir-12-2015-0378 doi: 10.1108/oir-12-2015-0378
Wu, A. H. (2018). Gendered language on the economics job market rumors forum. In *Aea papers and proceedings* (Vol. 108, pp. 175–79).

A MP-Hashtag network degree centrality

Table 9: MP-Hashtag network degree centrality

MP/hashtag	**Degree**
Stuart Robert	66
Michael McCormack	44
#auspol	36
Jason Clare	27
#bushfires	27
Adam Bandt	25
Susan Templeman	25
Rebekha Sharkie	23
#bushfire	19
Darren Chester	18
Madeleine King	18
Anthony Albanese	18
Greg Hunt	17
Terri Butler	16
Luke Gosling	16
Scott Buchholz	15
Andrew Leigh	14
Mark Coulton	13
Josh Burns	12
Damian Drum	11
Alex Hawke	11
Shayne Neumann	11
#BREAKING	10
Jason Falinski	9
James Stevens	9
Ged Kearney	9
Scott Morrison	8
Anika Wells	8
Joanne Ryan	8
#Estimates	8

B Party-Hashtag network degree centrality

Table 10: Party-Hashtag network degree centrality

Party/Hashtag	**Degree**
Australian Labor Party	177
Liberal National Party of Queensland	86
The Nationals	85
Liberal Party of Australia	56
Australian Greens	25
Centre Alliance	23
#bushfires	6
#Bushfire	5
#NSWfires	4
#NSWbushfires	4
#YourADF	4
#bushfire	4
#AustraliaFires	4
#auspol	4
#bushfireaustralia	3
#OpBushfireAssist	3
#BREAKING	3
#Bushfires	3
#SAFires	3
#Katter's Australian Party	3
#9News	3
#bushfiresNSW	3
#vicfires	3
#yourADF	3
#NDIS	2
#qanda	2
#fires	2
#Australianfires	2
#qt	2
#BatemansBay	2