Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis

Matthew Andreotta[1,2] · Robertus Nugroho[2,3] · Mark J. Hurlstone[1] · Fabio Boschetti[4] · Simon Farrell[1] · Iain Walker[5] · Cecile Paris[2]

Abstract

To qualitative researchers, social media offers a novel opportunity to harvest a massive and diverse range of content without the need for intrusive or intensive data collection procedures. However, performing a qualitative analysis across a massive social media data set is cumbersome and impractical. Instead, researchers often extract a subset of content to analyze, but a framework to facilitate this process is currently lacking. We present a four-phased framework for improving this extraction process, which blends the capacities of data science techniques to compress large data sets into smaller spaces, with the capabilities of qualitative analysis to address research questions. We demonstrate this framework by investigating the topics of Australian Twitter commentary on climate change, using quantitative (non-negative matrix inter-joint factorization; topic alignment) and qualitative (thematic analysis) techniques. Our approach is useful for researchers seeking to perform qualitative analyses of social media, or researchers wanting to supplement their quantitative work with a qualitative analysis of broader social context and meaning.

Keywords Big data · Topic modeling · Thematic analysis · Twitter · Climate change · Joint matrix factorization · Topic alignment

Introduction

Social scientists use qualitative modes of inquiry to explore the detailed descriptions of the world that people see and experience (Pistrang & Barker, 2012). To collect the voices of people, researchers can elicit textual descriptions of the world through interview or survey methodologies. However, with the popularity of the Internet and social media technologies, new avenues for data collection are possible. Social media platforms allow users to create content (e.g., Weinberg & Pehlivan, 2011), and interact with other users (e.g., Correa, Hinsley, & de Zùñiga, 2011; Kietzmann, Hermkens, McCarthy, & Silvestre, 2010), in settings where "Anyone can say Anything about Any topic" (*AAA slogan*,

Allemang & Hendler, 2011, pg. 6). Combined with the high rate of content production, social media platforms can offer researchers massive and diverse dynamic data sets (Yin & Kaynak, 2015; Gudivada et al., 2015). With technologies increasingly capable of harvesting, storing, processing, and analyzing this data, researchers can now explore data sets that would be infeasible to collect through more traditional qualitative methods.

Many social media platforms can be considered as textual corpora, willingly and spontaneously authored by millions of users. Researchers can compile a corpus using automated tools and conduct qualitative inquiries of content or focused analyses on specific users (Marwick, 2014). In this paper, we outline some of the opportunities and challenges of applying qualitative textual analyses to the big data of social media. Specifically, we present a conceptual and pragmatic justification for combining qualitative textual analyses with data science text-mining tools. This process allows us to both embrace and cope with the volume and diversity of commentary over social media. We then demonstrate this approach in a case study investigating Australian commentary on climate change, using content from the social media platform: Twitter.

✉ Matthew Andreotta
matthew.andreotta@research.uwa.edu.au

Extended author information available on the last page of the article.

Opportunities and challenges for qualitative researchers using social media data

Through social media, qualitative researchers gain access to a massive and diverse range of individuals, and the content they generate. Researchers can identify voices which may not be otherwise heard through more traditional approaches, such as semi-structured interviews and Internet surveys with open-ended questions. This can be done through diagnostic queries to capture the activity of specific peoples, places, events, times, or topics. Diagnostic queries may specify geotagged content, the time of content creation, textual content of user activity, and the online profile of users. For example, Freelon et al. (2018) identified the Twitter activity of three separate communities ('Black Twitter', 'Asian-American Twitter', 'Feminist Twitter') through the use of hashtags[1] in tweets from 2015 to 2016. A similar process can be used to capture specific events or moments (Procter et al., 2013; Denef et al., 2013), places (Lewis et al., 2013), and specific topics (Hoppe, 2009; Sharma et al., 2017).

Collecting social media data may be more scalable than traditional approaches. Once equipped with the resources to access and process data, researchers can potentially scale data harvesting without expending a great deal of resources. This differs from interviews and surveys, where collecting data can require an effortful and time-consuming contribution from participants and researchers.

Social media analyses may also be more ecologically valid than traditional approaches. Unlike approaches where responses from participants are elicited in artificial social contexts (e.g., Internet surveys, laboratory-based interviews), social media data emerges from real-world social environments encompassing a large and diverse range of people, without any prompting from researchers. Thus, in comparison with traditional methodologies (Onwuegbuzie & Leech, 2007; Lietz & Zayas, 2010; McKechnie, 2008), participant behavior is relatively unconstrained if not entirely unconstrained, by the behaviors of researchers.

These opportunities also come up with challenges, because of the following attributes (Parker et al., 2011). Firstly, social media can be *interactive*: its content involves the interactions of users with other users (e.g., conversations), or even external websites (e.g., links to news websites). The ill-defined boundaries of user interaction have implications for determining the units of analysis of qualitative study. For example, conversations can be lengthy, with multiple users, without a clear structure or end-point. Interactivity thus blurs the boundaries between users, their content, and external content (Herring, 2009;

Parker et al., 2011). Secondly, content can be *ephemeral* and *dynamic*. The users and content of their postings are transient (Parker et al., 2011; Boyd & Crawford, 2012; Weinberg & Pehlivan, 2011). This feature arises from the diversity of users, the dynamic socio-cultural context surrounding platform use, and the freedom users have to create, distribute, display, and dispose of their content (Marwick & Boyd, 2011). Lastly, social media content is *massive in volume*. The accumulated postings of users can lead to a large amount of data, and due to the diverse and dynamic content, postings may be largely unrelated and accumulate over a short period of time. Researchers hoping to harness the opportunities of social media data sets must therefore develop strategies for coping with these challenges.

A framework integrating computational and qualitative text analyses

Our framework—a mixed-method approach blending the capabilities of data science techniques with the capacities of qualitative analysis—is shown in Fig. 1. We overcome the challenges of social media data by automating some aspects of the data collection and consolidation, so that the qualitative researcher is left with a manageable volume of data to synthesize and interpret. Broadly, our framework consists of the following four phases: (1) harvest social media data and compile a corpus, (2) use data science techniques to compress the corpus along a dimension of relevance, (3) extract a subset of data from the most relevant spaces of the corpus, and (4) perform a qualitative analysis on this subset of data.

Phase 1: Harvest social media data and compile a corpus

Researchers can use automated tools to query records of social media data, extract this data, and compile it into a corpus. Researchers may query for content posted in
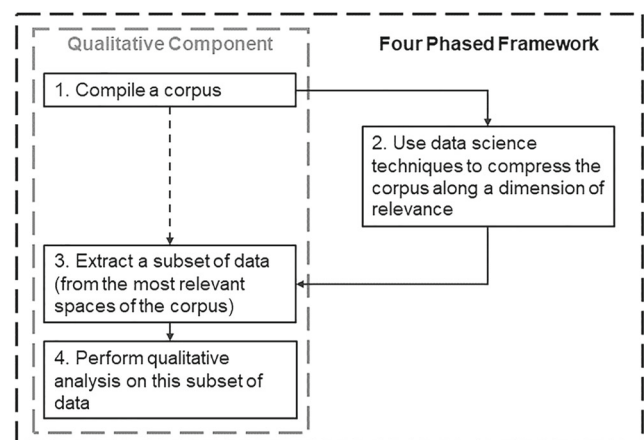
---

[1]On Twitter, users may precede a phrase with a hashtag (#). This allows users to signify and search for tweets related to a specific theme.

Fig. 1 Schematic overview of the four-phased framework

a particular time frame (Procter et al., 2013), content containing specified terms (Sharma et al., 2017), content posted by users meeting particular characteristics (Denef et al., 2013; Lewis et al., 2013), and content pertaining to a specified location (Hoppe, 2009).

Phase 2: Use data science techniques to compress the corpus along a dimension of relevance

Although researchers may be interested in examining the entire data set, it is often more practical to focus on a subsample of data (McKenna et al., 2017). Specifically, we advocate dividing the corpus along a dimension of relevance, and sampling from spaces that are more likely to be useful for addressing the research questions under consideration. By relevance, we refer to an attribute of content that is both useful for addressing the research questions and usable for the planned qualitative analysis.

To organize the corpus along a *dimension of relevance*, researchers can use automated, computational algorithms. This process provides both *formal* and *informal* advantages for the subsequent qualitative analysis. Formally, algorithms can assist researchers in privileging an aspect of the corpus most relevant for the current inquiry. For example, topic modeling clusters massive content into semantic topics—a process that would be infeasible using human coders alone. A plethora of techniques exist for separating social media corpora on the basis of useful aspects, such as sentiment (e.g., Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2010; Paris, Christensen, Batterham, & O'Dea, 2015; Pak & Paroubek, 2011) and influence (Weng et al., 2010).

Algorithms also produce an informal advantage for qualitative analysis. As mentioned, it is often infeasible for analysts to explore large data sets using qualitative techniques. Computational models of content can allow researchers to consider meaning at a corpus-level when interpreting individual datum or relationships between a subset of data. For example, in an inspection of 2.6 million tweets, Procter et al. (2013) used the output of an information flow analysis to derive rudimentary codes for inspecting individual tweets. Thus, algorithmic output can form a meaningful scaffold for qualitative analysis by providing analysts with summaries of potentially disjunct and multifaceted data (due to interactive, ephemeral, dynamic attributes of social media).

Phase 3: Extract a subset of data from the most relevant spaces of the corpus

Once the corpus is organized on the basis of relevance, researchers can extract data most relevant for answering their research questions. Researchers can extract a manageable amount of content to qualitatively analyze. For example, if the most relevant space of the corpus is too large for qualitative analysis, the researcher may choose to randomly sample from that space. If the most relevant space is small, the researcher may revisit Phase 2 and adopt a more lenient criteria of relevance.

Phase 4: Perform a qualitative analysis on this subset of data

The final phase involves performing the qualitative analysis to address the research question. As discussed above, researchers may draw on the computational models as a preliminary guide to the data.

Contextualizing the framework within previous qualitative social media studies

The proposed framework generalizes a number of previous approaches (Collins & Nerlich, 2015; McKenna et al., 2017) and individual studies (e.g., Lewis et al., 2013; Newman, 2016), in particular that of Marwick (2014). In Marwick's general description of qualitative analysis of social media textual corpora, researchers: (1) harvest and compile a corpus, (2) extract a subset of the corpus, and (3) perform a qualitative analysis on the subset. As shown in Fig. 1, our framework differs in that we introduce formal considerations of relevance, and the use of quantitative techniques to inform the extraction of a subset of data. Although researchers sometimes identify a subset of data most relevant to answering their research question, they seldom deploy data science techniques to identify it. Instead, researchers typically depend on more crude measures to isolate relevant data. For example, researchers have used the number of repostings of user content to quantify influence and recognition (e.g., Newman, 2016).

The steps in the framework may not be obvious without a concrete example. Next, we demonstrate our framework by applying it to Australian commentary regarding climate change on Twitter.

Application Example: Australian Commentary regarding Climate Change on Twitter

Social media platform of interest

We chose to explore user commentary of climate change over Twitter. Twitter activity contains information about: the textual content generated by users (i.e., content of tweets), interactions between users, and the time of content creation (Veltri & Atanasova, 2017). This allows us to examine the content of user communication, taking into account the temporal and social contexts of their behavior. Twitter data is relatively easy for researchers to access. Many tweets

reside within a public domain, and are accessible through free and accessible APIs.

The characteristics of Twitter's platform are also favorable for data analysis. An established literature describes computational techniques and considerations for interpreting Twitter data. We used the approaches and findings from other empirical investigations to inform our approach. For example, we drew on past literature to inform the process of identifying which tweets were related to climate change.

Public discussion on climate change

Climate change is one of the greatest challenges facing humanity (Schneider, 2011). Steps to prevent and mitigate the damaging consequences of climate change require changes on different political, societal, and individual levels (Lorenzoni & Pidgeon, 2006). Insights into public commentary can inform decision making and communication of climate policy and science.

Traditionally, public perceptions are investigated through survey designs and qualitative work (Lorenzoni & Pidgeon, 2006). Inquiries into social media allow researchers to explore a large and diverse range of climate change-related dialogue (Auer et al., 2014). Yet, existing inquiries of Twitter activity are few in number and typically constrained to specific events related to climate change, such as the release of the Fifth Assessment Report by the Intergovernmental Panel on Climate Change (Newman et al., 2010; O'Neill et al., 2015; Pearce, 2014) and the 2015 United Nations Climate Change Conference, held in Paris (Pathak et al., 2017).

When longer time scales are explored, most researchers rely heavily upon computational methods to derive topics of commentary. For example, Kirilenko and Stepchenkova (2014) examined the topics of climate change tweets posted in 2012, as indicated by the most prevalent hashtags. Although hashtags can mark the topics of tweets, it is a crude measure as tweets with no hashtags are omitted from analysis, and not all topics are indicated via hashtags (e.g., Nugroho, Yang, Zhao, Paris, & Nepal, 2017). In a more sophisticated approach, Veltri and Atanasova (2017) examined the co-occurrence of terms using hierarchical clustering techniques to map the semantic space of climate change tweet content from the year 2013. They identified four themes: (1) "calls for action and increasing awareness", (2) "discussions about the consequences of climate change", (3) "policy debate about climate change and energy", and (4) "local events associated with climate change" (p. 729).

Our research builds on the existing literature in two ways. Firstly, we explore a new data set—Australian tweets over the year 2016. Secondly, in comparison to existing research of Twitter data spanning long time periods, we use qualitative techniques to provide a more nuanced understanding of the topics of climate change. By applying our mixed-methods framework, we address our research question: what are the common topics of Australian's tweets about climate change?

Method

Outline of approach

We employed our four-phased framework as shown in Fig. 2. Firstly, we harvested climate change tweets posted in Australia in 2016 and compiled a corpus (phase 1). We then utilized a topic modeling technique (Nugroho et al., 2017) to organize the diverse content of the corpus into a number of topics. We were interested in topics which commonly appeared throughout the time period of data collection, and less interested in more transitory topics. To identify enduring topics, we used a topic alignment algorithm (Chuang et al., 2015) to group similar topics occurring repeatedly throughout 2016 (phase 2). This process allowed us to identify the topics most relevant to our research question. From each of these, we extracted a manageable subset of data (phase 3). We then performed a qualitative thematic analysis (see Braun & Clarke, 2006) on this subset of data to inductively derive themes and answer our research question (phase 4).[2]

Phase 1: Compiling a corpus

To search Australian's Twitter data, we used CSIRO's Emergency Situation Awareness (ESA) platform (CSIRO, 2018). The platform was originally built to detect, track, and report on unexpected incidences related to crisis situations (e.g., fires, floods; see Cameron, Power, Robinson, & Yin 2012). To do so, the ESA platform harvests tweets based on a location search that covers most of Australia and New Zealand.

The ESA platform archives the harvested tweets, which may be used for other CSIRO research projects. From this archive, we retrieved tweets satisfying three criteria: (1) tweets must be associated with an Australian location, (2) tweets must be harvested from the year 2016, and (3) the content of tweets must be related to climate change. We tested the viability of different markers of climate change tweets used in previous empirical work (Jang & Hart, 2015; Newman, 2016; Holmberg & Hellsten, 2016;

---

[2]The analysis of this study was preregistered on the Open Science Framework: https://osf.io/mb8kh/. See the Supplementary Material for a discussion of discrepancies. Analysis scripts and interim results from computational techniques can be found at: https://github.com/AndreottaM/TopicAlignment.
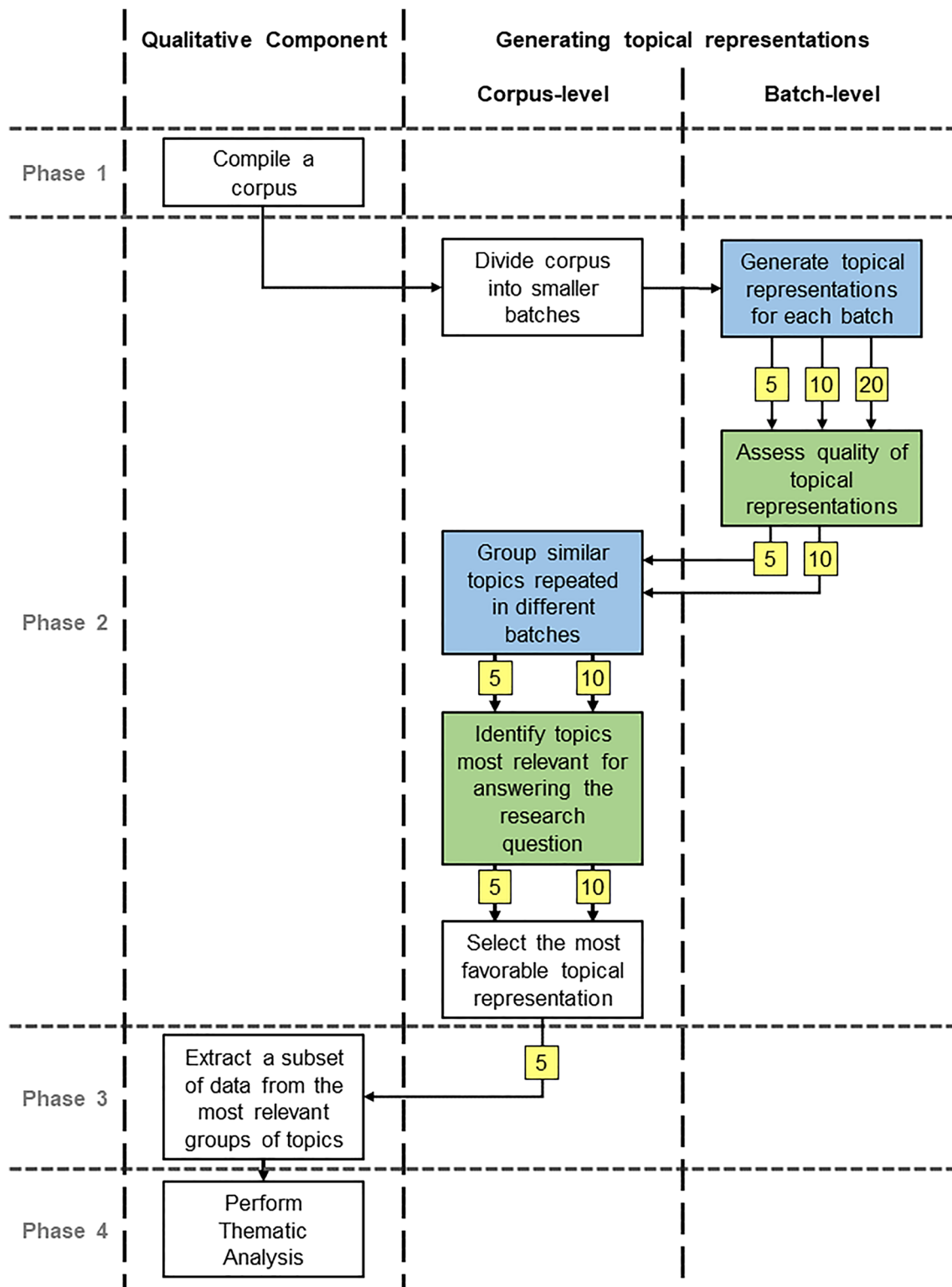
Fig. 2 Flowchart of application of a four-phased framework for conducting qualitative analyses using data science techniques. We were most interested in topics that frequently occurred throughout the period of data collection. To identify these, we organized the corpus chronologically, and divided the corpus into batches of content. Using computational techniques (shown in *blue*), we uncovered topics in each batch and identified similar topics which repeatedly occurred across batches. When identifying topics in each batch, we generated three alternative representations of topics (5, 10, and 20 topics in each batch, shown in *yellow*). In stages highlighted in *green*, we determined the quality of these representations, ultimately selecting the five topics per batch solution

O'Neill et al., 2015; Pearce et al., 2014; Sisco et al., 2017; Swain, 2017; Williams et al., 2015) by informally inspecting the content of tweets matching each criteria. Ultimately, we employed five terms (or combinations of terms) reliably associated with climate change: (1) "climate" AND "change"; (2) "#climatechange"; (3) "#climate"; (4) "global" AND "warming"; and (5) "#globalwarming". This yielded a corpus of 201,506 tweets.

Phase 2: Using data science techniques to compress the corpus along a dimension of relevance

The next step was to organize the collection of tweets into distinct topics. A topic is an abstract representation of semantically related words and concepts. Each tweet belongs to a topic, and each topic may be represented as a list of keywords (i.e., prominent words of tweets belonging to the topic).

A vast literature surrounds the computational derivation of topics within textual corpora, and specifically within Twitter corpora (Ramage et al., 2010; Nugroho et al., 2017; Fang et al., 2016a; Chuang et al., 2014). Popular methods for deriving topics include: probabilistic latent semantic analysis (Hofmann, 1999), non-negative matrix factorization (Lee & Seung, 2000), and latent Dirichlet allocation (Blei et al., 2003). These approaches use patterns of co-occurrence of terms within documents to derive topics. They work best on long documents. Tweets, however, are short, and thus only a few unique terms may co-occur between tweets. Consequently, approaches which rely upon patterns of term co-occurrence suffer within the Twitter environment. Moreover, these approaches ignore valuable social and temporal information (Nugroho et al., 2017). For example, consider a tweet t_1 and its reply t_2. The reply feature of Twitter allows users to react to tweets and enter conversations. Therefore, it is likely t_1 and t_2 are related in topic, by virtue of the reply interaction.

To address sparsity concerns, we adopt the non-negative matrix inter-joint factorization (NMijF) of Nugroho et al. (2017). This process uses both tweet content (i.e., the patterns of co-occurrence of terms amongst tweets) and socio-temporal relationship between tweets (i.e., similarities in the users mentioned in tweets, whether the tweet is a reply to another tweet, whether tweets are posted at a similar time) to derive topics (see Supplementary Material). The NMijF method has been demonstrated to outperform other topic modeling techniques on Twitter data (Nugroho et al., 2017).

Dividing the corpus into batches

Deriving many topics across a data set of thousands of tweets is prohibitively expensive in computational terms. Therefore, we divided the corpus into smaller batches and derived the topics of each batch. To keep the temporal relationships amongst tweets (e.g., timestamps of the tweets) the batches were organized chronologically. The data was partitioned into 41 disjoint batches (40 batches of 5000 tweets; one batch of 1506 tweets).

Generating topical representations for each batch

Following standard topic modeling practice, we removed features from each tweet which may compromise the quality of the topic derivation process. These features include: emoticons, punctuation, terms with fewer than three characters, stop-words (for list of stop-words, see MySQL, 2018), and phrases used to harvest the data (e.g., "#climatechange").[3] Following this, the terms remaining in tweets were stemmed using the Natural Language Toolkit for Python (Bird et al., 2009). All stemmed terms were then tokenized for processing.

The NMijF topic derivation process requires three parameters (see Supplementary Material for more details). We set two of these parameters to the recommendations of Nugroho et al. (2017), based on empirical analysis. The final parameter—the number of topics derived from each batch—is difficult to estimate *a priori*, and must be made with some care. If k is too small, keywords and tweets belonging to a topic may be difficult to conceptualize as a singular, coherent, and meaningful topic. If k is too large, keywords and tweets belonging to a topic may be too specific and obscure. To determine a reasonable value of k, we ran the NMijF process on each batch with three different levels of the parameter—5, 10, and 20 topics per batch. This process generated three different representations of the corpus: 205, 410, and 820 topics. For each of these representations, each tweet was classified into one (and only one) topic. We represented each topic as a list of ten keywords most prevalent within the tweets of that topic.

Assessing the quality of topical representations

To select a topical representation for further analysis, we inspected the quality of each. Initially, we considered the use of a completely automatic process to assess or produce high quality topic derivations. However, our attempts to use completely automated techniques on tweets with a known topic structure failed to produce correct or reasonable solutions. Thus, we assessed quality using human assessment (see Table 1). The first stage involved inspecting each topical representation of the corpus (205, 410, and 820 topics), and manually flagging any topics that were clearly problematic. Specifically, we examined each topical representation to determine whether topics represented as separate were in fact distinguishable from

---

[3] 83 tweets were rendered empty and discarded from the corpus.

Table 1 Two-staged assessment of the quality of topic derivations

Stage	Level of inspection	Quality metric	Definition
1	Topical representation of each batch	Distinctiveness	Degree to which topics within a batch can be distinguished from other topics of the same batch
2	Individual topics	Coherency	Degree to which the topic contains keywords that are semantically similar
		Meaning	Degree to which the topic contains keywords that reference fewer discussions and events
		Interpretability	Degree to which the keywords convey specific information about a topic
		Related tweets	Degree to which tweets in the topic reflect the keywords and meaning of the topic

one another. We discovered that the 820 topic representation (20 topics per batch) contained many closely related topics.

To quantify the distinctiveness between topics, we compared each topic to each other topic in the same batch in an automated process. If two topics shared three or more (of ten) keywords, these topics were deemed similar. We adopted this threshold from existing topic modeling work (Fang et al. 2016a, b), and verified it through an informal inspection. We found that pairs of topics below this threshold were less similar than those equal to or above it. Using this threshold, the 820 topic representation was identified as less distinctive than other representations. Of the 41 batches, nine contained at least two similar topics for the 820 topic representation (cf., 0 batches for the 205 topic representation, two batches for the 410 topic representation). As a result, we chose to exclude the representation from further analysis.

The second stage of quality assessment involved inspecting the quality of individual topics. To achieve this, we adopted the pairwise topic preference task outlined by Fang et al. (2016a, b). In this task, raters were shown pairs of two similar topics (represented as ten keywords), one from the 205 topic representation and the other from the 410 topic representation. To assist in their interpretation of topics, raters could also view three tweets belonging to each topic. For each pair of topics, raters indicated which topic they believed was superior, on the basis of coherency, meaning, interpretability, and the related tweets (see Table 1). Through aggregating responses, a relative measure of quality could be derived.

Initially, members of the research team assessed 24 pairs of topics. Results from the task did not indicate a marked preference for either topical representation. To confirm this impression more objectively, we recruited participants from the Australian community as raters. We used Qualtrics—an online survey platform and recruitment service—to recruit 154 Australian participants, matched with the general Australian population on age and gender. Each participant completed judgments on 12 pairs of similar topics (see Supplementary Material for further information).

Participants generally preferred the 410 topic representation over the 205 topic representation ($M = 6.45$ of 12 judgments, $SD = 1.87$). Of 154 participants, 35 were classified as indifferent (selected both topic representations an equal number of times), 74 preferred the 410 topic representation (i.e., selected the 410 topic representation more often than the 205 topic representation), and 45 preferred the 205 topic representation (i.e., selected the 205 topic representation more often that the 410 topic representation). We conducted binomial tests to determine whether the proportion of participants of the three just described types differed reliably from chance levels (0.33). The proportion of indifferent participants (0.23) was reliably lower than chance ($p = 0.005$), whereas the proportion of participants preferring the 205 topic solution (0.29) did not differ reliably from chance levels ($p = 0.305$). Critically, the proportion of participants preferring the 410 topic solution (0.48) was reliably higher than expected by chance ($p < 0.001$). Overall, this pattern indicates a participant preference for the 410 topic representation over the 205 topic representation.

In summary, no topical representation was unequivocally superior. On a batch level, the 410 topic representation contained more batches of non-distinct topic solutions than the 205 topic representation, indicating that the 205 topic representation contained topics which were more distinct. In contrast, on the level of individual topics, the 410 topic representation was preferred by human raters. We use this information, in conjunction with the utility of corresponding aligned topics (see below), to decide which representation is most suitable for our research purposes.

Grouping similar topics repeated in different batches

We were most interested in topics which occurred throughout the year (i.e., in multiple batches) to identify the most stable components of climate change commentary (phase 3). We grouped similar topics from different batches using a topical alignment algorithm (see Chuang et al. 2015). This process requires a similarity metric and a similarity threshold. The similarity metric represents the similarity between two topics, which we specified as the proportion of shared keywords (from 0, no keywords shared, to 1, all ten keywords shared). The similarity threshold is a value below which two topics were deemed dissimilar. As above, we set the threshold to 0.3 (three of ten keywords shared)—if two topics shared two or fewer keywords, the

Table 2 Glossary of critical terms

Concept	Definition	Process for derivation
Topic	An abstract representation of semantically related words and concepts	Topic Modeling
Group of topics	A collection of similar topics from different batches	Topic Alignment
Prevalent topic groupings	Groups of topics which contain at least three topics	Topic Alignment
Theme	A patterned meaning of tweets that distinctly answers our research question: what are the common topics of Australian's tweets about climate change?	Thematic Analysis

topics could not be justifiably classified as similar. To delineate important topics, groups of topics, and other concepts we have provided a glossary of terms in Table 2.

The topic alignment algorithm is initialized by assigning each topic to its own group. The alignment algorithm iteratively merges the two most similar groups, where the similarity between groups is the maximum similarity between a topic belonging to one group and another topic belonging to the other. Only topics from different groups (by definition, topics from the same group are already grouped as similar) and different batches (by definition, topics from the same batch cannot be similar) can be grouped. This process continues, merging similar groups until no compatible groups remain. We found our initial implementation generated groups of largely dissimilar topics. To address this, we introduced an additional constraint—groups could only be merged if the mean similarity between pairs of topics (each belonging to the two groups in question) was greater than the similarity threshold. This process produced groups of similar topics. Functionally, this allowed us to detect topics repeated throughout the year.

We ran the topical alignment algorithm across both the 205 and 410 topic representations. For the 205 and 410 topic representation respectively, 22.47 and 31.60% of tweets were not associated with topics that aligned with others. This exemplifies the ephemeral and dynamic attributes of Twitter activity: over time, the content of tweets shifts, with some topics appearing only once throughout the year (i.e., in only one batch). In contrast, we identified 42 groups (69.77% of topics) and 101 groups (62.93% of topics) of related topics for the 205 and 410 topic representations respectively, occurring across different time periods (i.e., in more than one batch). Thus, both representations contained transient topics (isolated to one batch) and recurrent topics (present in more than one batch, belonging to a group of two or more topics).

Identifying topics most relevant for answering our research question

For the subsequent qualitative analyses, we were primarily interested in topics prevalent throughout the corpus. We operationalized *prevalent topic groupings* as any grouping

of topics that spanned three or more batches. On this basis, 22 (57.50% of tweets) and 36 (35.14% of tweets) groupings of topics were identified as prevalent for the 205 and 410 topic representations, respectively (see Table 3). As an example, consider the prevalent topic groupings from the 205 topic representation, shown in Table 3. Ten topics are united by commentary on the Great Barrier Reef (Group 2)—indicating this facet of climate change commentary was prevalent throughout the year. In contrast, some topics rarely occurred, such as a topic concerning a climate change comic (indicated by the keywords "xkcd" and "comic") occurring once and twice in the 205 and 410 topic representation, respectively. Although such topics are meaningful and interesting, they are transient aspects of climate change commentary and less relevant to our research question. In sum, topic modeling and grouping algorithms have allowed us to collate massive amounts of information, and identify components of the corpus most relevant to our qualitative inquiry.

Selecting the most favorable topical representation

At this stage, we have two complete and coherent representations of the corpus topics, and indications of which topics are most relevant to our research question. Although some evidence indicated that the 410 topic representation contains topics of higher quality, the 205 topic representation was more parsimonious on both the level of topics and groups of topics. Thus, we selected the 205 topic representation for further analysis.

Phase 3. Extract a subset of data

Extracting a subset of data from the selected topical representation

Before qualitative analysis, researchers must extract a subset of data manageable in size. For this process, we concerned ourselves with only the content of prevalent topic groupings, seen in Table 3. From each of the 22 prevalent topic groupings, we randomly sampled ten tweets. We selected ten tweets as a trade-off between comprehensiveness and feasibility. This thus reduced our data space for qualitative analysis from 201,423 tweets to 220.

Table 3 Prevalent topic groupings (205 topic representation) and associated keywords

Group	Total batches	Proportion of corpus (%)	Common keywords
1	13	9.83	action need now
2	10	4.92	#greatbarrierreef **barrier** great **reef**
3	7	4.53	coal new
4	6	3.00	action pai plan **real**
5	6	2.35	denial malcolm nation new one robe senat
6	5	3.54	hottest new record year
7	5	3.47	#qanda energi need renew
8	5	2.93	fight govt one **peopl**
9	5	2.45	#parisagr **agreement pari** ratifi time world
10	4	2.23	#qanda emiss health impact **need** polici risk talk
11	4	1.68	**extrem** link make now power renew scientist **weather**
12	3	1.84	**debat nation one** senat
13	3	1.75	action **believ** malcolm **peopl** real world
14	3	1.62	**impact** iss **malcolm peopl**
15	3	1.58	#qanda **need real reef**
16	3	1.54	#qldpol **#scienc** #wapol denier need **year**
17	3	1.53	flood iss now **polici** scientist
18	3	1.45	**action** impact klein naomi world
19	3	1.39	malcolm repo **risk scientist warn**
20	3	1.38	**latest** stop **thank world**
21	3	1.26	**latest peopl** planet **thank** year
22	3	1.21	**level rise sea**

Note. Only keywords shared between at least half of all topics within a group are included. Keywords in bold are shared between all topics of that group. All keywords are presented in stemmed form

Phase 4: Perform qualitative analysis

Perform thematic analysis

In the final phase of our analysis, we performed a qualitative thematic analysis (TA; Braun & Clarke, 2006) on the subset of tweets sampled in phase 3. This analysis generated distinct themes, each of which answers our research question: what are the common topics of Australian's tweets about climate change? As such, the themes generated through TA are topics. However, unlike the topics derived from the preceding computational approaches, these themes are informed by the human coder's interpretation of content and are oriented towards our specific research question. This allows the incorporation of important diagnostic information, including the broader socio-political context of discussed events or terms, and an understanding (albeit, sometimes ambiguous) of the underlying latent meaning of tweets.

We selected TA as the approach allows for flexibility in assumptions and philosophical approaches to qualitative inquiries. Moreover, the approach is used to emphasize similarities and differences *between* units of analysis (i.e.,

between tweets) and is therefore useful for generating topics. However, TA is typically applied to lengthy interview transcripts or responses to open survey questions, rather than small units of analysis produced through Twitter activity. To ease the application of TA to small units of analysis, we modified the typical TA process (shown in Table 4) as follows.

Firstly, when performing phases 1 and 2 of TA, we initially read through each prevalent topic grouping's tweets sequentially. By doing this, we took advantage of the relative homogeneity of content within topics. That is, tweets sharing the same topic will be more similar in content than tweets belonging to separate topics. When reading ambiguous tweets, we could use the tweet's topic (and other related topics from the same group) to aid comprehension. Through the scaffold of topic representations, we facilitated the process of interpreting the data, generating initial codes, and deriving themes.

Secondly, the prevalent topic groupings were used to create initial codes and search for themes (TA phase 2 and 3). For example, the groups of topics indicate content of climate change action (group 1), the Great Barrier Reef (group 2), climate change deniers (group 3), and

Table 4 Phases of thematic analysis

Phase	Description of the process
1. Familiarizing yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Note. Reprinted from "Using thematic analysis in psychology," by V. Braun and V. Clarke, 2006, *Qualitative Research in Psychology*, *3*(2), p. 87. Copyright 2006 by the Edward Arnold (Publishers) Ltd. Reprinted with permission

extreme weather (group 5). The keywords characterizing these topics were used as initial codes (e.g., "action", "Great Barrier Reef", "Paris Agreement", "denial"). In sum, the algorithmic output provided us with an initial set of codes and an understanding of the topic structure that can indicate important features of the corpus.

A member of the research team performed this augmented TA to generate themes. A second rater outside of the research team applied the generated themes to the data, and inter-rater agreement was assessed. Following this, the two raters reached a consensus on the theme of each tweet.

Results

Through TA, we inductively generated five distinct themes. We assigned each tweet to one (and only one) theme. A degree of ambiguity is involved in designating themes for tweets, and seven tweets were too ambiguous to subsume into our thematic framework. The remaining 213 tweets were assigned to one of five themes shown in Table 5.

In an initial application of the coding scheme, the two raters agreed upon 161 (73.181%) of 220 tweets. Inter-rater reliability was satisfactory, Cohen's $\kappa = 0.648$, $p < 0.05$. An assessment of agreement for each theme is presented in Table 5. The *proportion of agreement* is the total proportion of observations where the two coders both agreed: (1) a tweet belonged to the theme, or (2) a tweet did not belong to the theme. The *proportion of specific agreement* is the conditional probability that a randomly selected rater will assign the theme to a tweet, given that the other rater did (see Supplementary Material for more information). Theme 3, theme 5, and the N/A categorization had lower levels of agreement than the remaining themes, possibly as tweets

belonging to themes 3 and 5 often make references to content relevant to other themes.

Theme 1. Climate change action The theme occurring most often was climate change action, whereby tweets were related to coping with, preparing for, or preventing climate change. Tweets comment on the action (and inaction) of politicians, political parties, and international cooperation between government, and to a lesser degree, industry, media, and the public. The theme encapsulated commentary on: prioritizing climate change action ("*Let's start working together for real solutions on climate change*");[4] relevant strategies and policies to provide such action ("*#OurOcean is absorbing the majority of #climatechange heat. We need #marinereserves to help build resilience.*"); and the undertaking ("*Labor will take action on climate change, cut pollution, secure investment & jobs in a growing renewables industry*") or disregarding ("*act on Paris not just sign*") of action.

Often, users were critical of current or anticipated action (or inaction) towards climate change, criticizing approaches by politicians and governments as ineffective ("*Malcolm Turnbull will never have a credible climate change policy*"),[5] and undesirable ("*Govt: how can we solve this vexed problem of climate change? Helpful bystander: u could not allow a gigantic coal mine. Govt: but srsly how?*"). Predominately, users characterized the government as unjustifiably paralyzed ("*If a foreign country did half the damage to our country as #climatechange we would declare war.*"), without a leadership focused on addressing

---

[4] The content of tweet are reported verbatim. Sensitive information is redacted.

[5] Malcolm Turnbull was the Prime Minister of Australia during the year 2016.

Table 5 Summary of themes

Theme	Number of occurrences	Example tweets	Examples of events users mention	Proportion of agreement	Proportion of specific agreement	Related prevalent topic groupings	Related topic
1. Climate change action	87 (39.55%)	"Yes! Let's start working together for real solutions on climate change #QandA"	Paris Climate Change Agreement. Australian Federal Election.	0.845	0.795	1, 9	need action make chang save leader sta elect vital brave
2. Consequences of climate change	43 (19.55%)	"RT @user: Reefs of the future could look like this if we continue to ignore #climatechange..... https://..."		0.923	0.832	2, 6, 22	mammal first wipe speci extinct rodent case declar brambl cay
3. Conversations on climate change	35 (15.91%)	"RT @user: @user not so gripping from No Principles Malcolm. Not one mention of climate change in his pitch."	Leader debates preceding elections in Australia and the United States. UNESCO removes content on Australia from a report of the risks of climate change to world heritage sites.	0.877	0.571	12	#greatbarrierreef world censor impact #reefgat reef great press barrier mention
4. Climate change deniers	28 (12.73%)	"Don't worry! According to Senator elect Malcolm Roberts, NASA fiddles the figures on Climate Change. https://..."	Malcolm Roberts elected in the Australian Senate. Donald Trump is elected as President of the United States. Malcolm Roberts debates Brian Cox on the panel discussion television program Q & A.	0.936	0.741	5	malcolm robe one senat evid demand empir mislead denial fact
5. The legitimacy of climate change and climate science	20 (9.09%)	"Do we have an international convention on 'Cloud Seeding'? Or it comes under United Nation's climate change agreement?"	CSIRO decides to cut jobs of climate researchers	0.918	0.571	19	scienc scientist csiro believe time great real risk eah make
N/A	7 (3.18%)	"#QandA.Climate change and jobs and growth"		0.964	0.429		

Note. Content was anonymized by replacing references to usernames with the term "user", and replacing links to websites with "https://...". Related prevalent topic groupings and topics are selected on the basis of sharing similar summary meaning with each theme. Prevalent topic groupings are indicated as numbers corresponding to groups shown in Table 3. Each topic is presented as the ten stemmed keywords belonging to that topic

climate change ("*an election that leaves Australia with no leadership on #climatechange - the issue of our time!*").

Theme 2. Consequences of climate change Users commented on the consequences and risks attributed to climate change. This theme may be further categorized into commentary of: physical systems, such as changes in climate, weather, sea ice, and ocean currents ("*Australia experiencing more extreme fire weather, hotter days as climate changes*"); biological systems, such as marine life (particularly, the Great Barrier Reef) and biodiversity ("*Reefs of the future could look like this if we continue to ignore #climatechange*"); human systems ("*You and your friends will die of old age & I'm going to die from climate change*"); and other miscellaneous consequences ("*The reality is, no matter who you supported, or who wins, climate change is going to destroy everything you love*"). Users specified a wide range of risks and impacts on human systems, such as health, cultural diversity, and insurance. Generally, the consequences of climate change were perceived as negative.

Theme 3. Conversations on climate change Some commentary centered around discussions of climate change communication, debates, art, media, and podcasts. Frequently, these pertained to debates between politicians ("*not so gripping from No Principles Malcolm. Not one mention of climate change in his pitch.*") and television panel discussions ("*Yes let's all debate whether climate change is happening... #qanda*").[6] Users condemned the climate change discussions of federal government ("*Turnbull gov echoes Stalinist Russia? Australia scrubbed from UN climate change report after government intervention*"), those skeptical of climate change ("*Trouble is climate change deniers use weather info to muddy debate. Careful?????????????????*"), and media ("*Will politicians & MSM hacks ever work out that they cannot spin our way out of the #climatechange crisis?*"). The term "climate change" was critiqued, both by users skeptical of the legitimacy of climate change ("*Weren't we supposed to call it 'climate change' now? Are we back to 'global warming' again? What happened? Apart from summer?*") and by users seeking action ("*Maybe governments will actually listen if we stop saying "extreme weather" & "climate change" & just say the atmosphere is being radicalized*").

Theme 4. Climate change deniers The fourth theme involved commentary on individuals or groups who were perceived to deny climate change. Generally, these were politicians and associated political parties, such as: Malcolm Roberts (a climate change skeptic, elected as an Australian Senator in 2016), Malcolm Turnbull, and Donald Trump. Commentary

focused on the beliefs and legitimacy of those who deny the science of climate change ("*One Nation's Malcolm Roberts is in denial about the facts of climate change*") or support the denial of climate change science ("*Meanwhile in Australia... Malcolm Roberts, funded by climate change skeptic global groups loses the plot when nobody believes his findings*"). Some users advocated attempts to change the beliefs of those who deny climate change science ("*We have a president-elect who doesn't believe in climate change. Millions of people are going to have to say: Mr. Trump, you are dead wrong*"), whereas others advocated disengaging from conversation entirely ("*You know I just don't see any point engaging with climate change deniers like Roberts. Ignore him*"). In comparison to other themes, commentary revolved around individuals and their beliefs, rather than the phenomenon of climate change itself.

Theme 5. The legitimacy of climate change and climate science This theme concerns the reality of climate change ("*How do we know this climate change thing is real - not a natural cycle, not an elaborate hoax?*") and the associated practice of climate science ("*#CSIROcuts will damage Aus ability to understand, respond to & plan for #climatechange*").[7] Compared to other themes, content collated under this theme contained a wide variety of sentiment. Whereas some tweets endorse anthropogenic causes of climate change, others question the contribution of humans to climate change ("*COWS FARTS CAUSE MORE THAN WE DO*") and question its existence entirely ("*The effects of Climate Change ?? OK ██, lets talk facts.....which effects are those ??*").

Discussion

Using our four-phased framework, we aimed to identify and qualitatively inspect the most enduring aspects of climate change commentary from Australian posts on Twitter in 2016. We achieved this by using computational techniques to model 205 topics of the corpus, and identify and group similar topics that repeatedly occurred throughout the year. From the most relevant topic groupings, we extracted a subsample of tweets and identified five themes with a thematic analysis: climate change action, consequences of climate change, conversations on climate change, climate change deniers, and the legitimacy of climate change and climate science. Overall, we demonstrated the process of using a mixed-methodology that blends qualitative analyses with data science methods to explore social media data.

---

[6]"*#qanda*" is a hashtag used to refer to Q & A, an Australian panel discussion television program.

[7]Commonwealth Scientific and Industrial Research Organisation (CSIRO) is the national scientific research agency of Australia.

Our workflow draws on the advantages of both quantitative and qualitative techniques. Without quantitative techniques, it would be impossible to derive topics that apply to the entire corpus. The derived topics are a preliminary map for understanding the corpus, serving as a scaffold upon which we could derive meaningful themes contextualized within the wider socio-political context of Australia in 2016. By incorporating quantitatively-derived topics into the qualitative process, we attempted to construct themes that would generalize to a larger, relevant component of the corpus. The robustness of these themes is corroborated by their association with computationally-derived topics, which repeatedly occurred throughout the year (i.e., prevalent topic groupings). Moreover, four of the five themes have been observed in existing data science analyses of Twitter climate change commentary. Within the literature, the themes of climate change action and consequences of climate change are common (Newman, 2016; O'Neill et al., 2015; Pathak et al., 2017; Pearce, 2014; Jang & Hart, 2015; Veltri & Atanasova, 2017). The themes of the legitimacy of climate change and climate science (Jang & Hart, 2015; Newman, 2016; O'Neill et al., 2015; Pearce, 2014) and climate change deniers (Pathak et al., 2017) have also been observed. The replication of these themes demonstrates the validity of our findings.

One of the five themes—conversations on climate change—has not been explicitly identified in existing data science analyses of tweets on climate change. Although not explicitly identifying the theme, Kirilenko and Stepchenkova (2014) found hashtags related to public conversations (e.g., "#qanda", "#Debates") were used frequently throughout the year 2012. Similar to the literature, few (if any) topics in our 205 topic solution could be construed as solely relating to the theme of "conversation". However, as we progressed through the different phases of the framework, the theme became increasingly apparent. By the grouping stage, we identified a collection of topics unified by a keyword relating to debate. The subsequent thematic analysis clearly discerned this theme. The derivation of a theme previously undetected by other data science studies lends credence to the conclusions of Guetterman et al. (2018), who deduced that supplementing a quantitative approach with a qualitative technique can lead to the generation of more themes than a quantitative approach alone.

The uniqueness of a conversational theme can be accounted for by three potentially contributing factors. Firstly, tweets related to conversations on climate change often contained material pertinent to other themes. The overlap between this theme and others may hinder the capabilities of computational techniques to uniquely cluster these tweets, and undermine the ability of humans to reach agreement when coding content for this theme (indicated by the relatively low proportion of specific agreement in our thematic analysis). Secondly, a conversational theme may only be relevant in election years. Unlike other studies spanning long time periods (Jang & Hart, 2015; Veltri & Atanasova, 2017), Kirilenko and Stepchenkova (2014) and our study harvested data from US presidential election years (2012 and 2016, respectively). Moreover, an Australian federal election occurred in our year of observation. The occurrence of national elections and associated political debates may generate more discussion and criticisms of conversations on climate change. Alternatively, the emergence of a conversational theme may be attributable to the Australian panel discussion television program Q & A. The program regularly hosts politicians and other public figures to discuss political issues. Viewers are encouraged to participate by publishing tweets using the hashtag "#qanda", perhaps prompting viewers to generate uniquely tagged content not otherwise observed in other countries. Importantly, in 2016, Q & A featured a debate on climate change between science communicator Professor Brian Cox and Senator Malcolm Roberts, a prominent climate science skeptic.

Although our four-phased framework capitalizes on both quantitative and qualitative techniques, it still has limitations. Namely, the sparse content relationships between data points (in our case, tweets) can jeopardize the quality and reproducibility of algorithmic results (e.g., Chuang et al., 2015). Moreover, computational techniques can require large computing resources. To a degree, our application mitigated these limitations. We adopted a topic modeling algorithm which uses additional dimensions of tweets (social and temporal) to address the influence of term-to-term sparsity (Nugroho et al., 2017). To circumvent concerns of computing resources, we partitioned the corpus into batches, modeled the topics in each batch, and grouped similar topics together using another computational technique (Chuang et al., 2015).

As a demonstration of our four-phased framework, our application is limited to a single example. For data collection, we were able to draw from the procedures of existing studies which had successfully used keywords to identify climate change tweets. Without an existing literature, identifying diagnostic terms can be difficult. Nevertheless, this demonstration of our four-phased framework exemplifies some of the critical decisions analysts must make when utilizing a mixed-method approach to social media data.

Both qualitative and quantitative researchers can benefit from our four-phased framework. For qualitative researchers, we provide a novel vehicle for addressing their research questions. The diversity and volume of content of social media data may be overwhelming for both the researcher and their method. Through computational techniques, the diversity and scale of data can be managed,

allowing researchers to obtain a large volume of data and extract from it a relevant sample to conduct qualitative analyses. Additionally, computational techniques can help researchers explore and comprehend the nature of their data. For the quantitative researcher, our four-phased framework provides a strategy for formally documenting the qualitative interpretations. When applying algorithms, analysts must ultimately make qualitative assessments of the quality and meaning of output. In comparison to the mathematical machinery underpinning these techniques, the qualitative interpretations of algorithmic output are not well-documented. As these qualitative judgments are inseparable from data science, researchers should strive to formalize and document their decisions—our framework provides one means of achieving this goal.

Through the application of our four-phased framework, we contribute to an emerging literature on public perceptions of climate change by providing an in-depth examination of the structure of Australian social media discourse. This insight is useful for communicators and policy makers hoping to understand and engage the Australian online public. Our findings indicate that, within Australian commentary on climate change, a wide variety of messages and sentiment are present. A positive aspect of the commentary is that many users want action on climate change. The time is ripe it would seem for communicators to discuss Australia's policy response to climate change—the public are listening and they want to be involved in the discussion. Consistent with this, we find some users discussing conversations about climate change as a topic. Yet, in some quarters there is still skepticism about the legitimacy of climate change and climate science, and so there remains a pressing need to implement strategies to persuade members of the Australian public of the reality and urgency of the climate change problem. At the same time, our analyses suggest that climate communicators must counter the sometimes held belief, expressed in our second theme on climate change consequences, that it is already too late to solve the climate problem. Members of the public need to be aware of the gravity of the climate change problem, but they also need powerful self efficacy promoting messages that convince them that we still have time to solve the problem, and that their individual actions matter.

References

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In *Proceedings of the Workshop on Languages in Social Media*, (pp. 30–38). Stroudsburg: Association for Computational Linguistics.

Allemang, D., & Hendler, J. (2011). *Semantic web for the working ontologist: Effective modelling in RDFS and OWL*, (2nd ed.). United States of America: Elsevier Inc.

Auer, M. R., Zhang, Y., & Lee, P. (2014). The potential of microblogs for the study of public perceptions of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 5(3), 291–296. https://doi.org/10.1002/wcc.273

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. United States of America: O'Reilly Media, Inc.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cameron, M. A., Power, R., Robinson, B., & Yin, J. (2012). Emergency situation awareness from Twitter for crisis management. In *Proceedings of the 21st international conference on World Wide Web*, (pp. 695–698). New York: ACM, https://doi.org/10.1145/2187980.2188183

Chuang, J., Wilkerson, J. D., Weiss, R., Tingley, D., Stewart, B. M., Roberts, M. E., & et al. (2014). Computer-assisted content analysis: Topic models for exploring multiple subjective interpretations. In *Advances in Neural Information Processing Systems workshop on human-propelled machine learning* (pp. 1–9). Montreal, Canada: Neural Information Processing Systems.

Chuang, J., Roberts, M. E., Stewart, B. M., Weiss, R., Tingley, D., Grimmer, J., & Heer, J. (2015). TopicCheck: Interactive alignment for assessing topic model stability. In *Proceedings of the conference of the North American chapter of the Association for Computational Linguistics - Human Language Technologies*, (pp. 175–184). Denver: Association for Computational Linguistics, https://doi.org/10.3115/v1/N15-1018

Collins, L., & Nerlich, B. (2015). Examining user comments for deliberative democracy: A corpus-driven analysis of the climate change debate online. *Environmental Communication*, 9(2), 189–207. https://doi.org/10.1080/17524032.2014.981560

Correa, T., Hinsley, A. W., & de, Z.ùñiga. H.G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2), 247–253. https://doi.org/10.1016/j.chb.2009.09.003

CSIRO (2018). Emergency Situation Awareness. Retrieved 2019-02-20, from https://esa.csiro.au/ausnz/about-public.html

Denef, S., Bayerl, P. S., & Kaptein, N. A. (2013). Social media and the police: Tweeting practices of British police forces during the August 2011 riots. In *Proceedings of the SIGCHI conference on human factors in computing systems*, (pp. 3471–3480). NY: ACM.

Fang, A., Macdonald, C., Ounis, I., & Habel, P. (2016a). Topics in Tweets: A user study of topic coherence metrics for Twitter data. In *38th European conference on IR research, ECIR 2016*, (pp. 429–504). Switzerland: Springer International Publishing.

Fang, A., Macdonald, C., Ounis, I., & Habel, P. (2016b). Using word embedding to evaluate the coherence of topics from Twitter data. In *Proceedings of the 39th international ACM SIGIR conference on research and development*

in information retrieval, (pp. 1057–1060). NY: ACM Press, https://doi.org/10.1145/2911451.2914729

Freelon, D., Lopez, L., Clark, M. D., & Jackson, S. J. (2018). How black Twitter and other social media communities interact with mainstream news (Tech. Rep.). Knight Foundation. Retrieved 2018-04-20, from https://knightfoundation.org/features/twittermedia

Gudivada, V. N., Baeza-Yates, R. A., & Raghavan, V. V. (2015). Big data: Promises and problems. *IEEE Computer*, *48*(3), 20–23.

Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. V. (2018). Augmenting qualitative text analysis with natural language processing: Methodological study. *Journal of Medical Internet Research*, *20*, 6. https://doi.org/10.2196/jmir.9702

Herring, S. C. (2009). Web content analysis: Expanding the paradigm. In Hunsinger, J., Klastrup, L., & Allen, M. (Eds.) *International handbook of internet research*, (pp. 233–249). Dordrecht: Springer.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*, (Vol. 51, pp. 211–218). Berkeley: ACM, https://doi.org/10.1109/BigData Congress.2015.21

Holmberg, K., & Hellsten, I. (2016). Integrating and differentiating meanings in tweeting about the fifth Intergovernmental Panel on Climate Change (IPCC) report. *First Monday*, *21*, 9. https://doi.org/10.5210/fm.v21i9.6603

Hoppe, R. (2009). Scientific advice and public policy: Expert advisers' and policymakers' discourses on boundary work. *Poiesis & Praxis*, *6*(3–4), 235–263. https://doi.org/10.1007/s10202-008-0053-3

Jang, S. M., & Hart, P. S. (2015). Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data. *Global Environmental Change*, *32*, 11–17. https://doi.org/10.1016/j.gloenvcha.2015.02.010

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251. https://doi.org/10.1016/j.bushor.2011.01.005

Kirilenko, A. P., & Stepchenkova, S. O. (2014). Public microblogging on climate change: One year of Twitter worldwide. *Global Environmental Change*, *26*, 171–182. https://doi.org/10.1016/j.gloenvcha.2014.02.008

Lee, D. D., & Seung, H. S. (2000). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, (pp. 556–562). Denver: Neural Information Processing Systems.

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, *57*(1), 34–52. https://doi.org/10.1080/08838151.2012.761702

Lietz, C. A., & Zayas, L. E. (2010). Evaluating qualitative research for social work practitioners. *Advances in Social Work*, *11*(2), 188–202.

Lorenzoni, I., & Pidgeon, N. F. (2006). Public views on climate change: European and USA perspectives. *Climatic Change*, *77*(1-2), 73–95. https://doi.org/10.1007/s10584-006-9072-z

Marwick, A. E. (2014). Ethnographic and qualitative research on Twitter. In Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (Eds.) *Twitter and society*, (Vol. 89, pp. 109–121). New York: Peter Lang.

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

McKechnie, L. E. F. (2008). Reactivity. In Given, L. (Ed.) *The SAGE encyclopedia of qualitative research methods*, (Vol. 2, pp. 729–730). Thousand Oaks California United States: SAGE Publications, Inc, https://doi.org/10.4135/9781412963909.n368.

McKenna, B., Myers, M. D., & Newman, M. (2017). Social media in qualitative research: Challenges and recommendations. *Information and Organization*, *27*(2), 87–99. https://doi.org/10.1016/j.infoandorg.2017.03.001

MySQL (2018). Full-Text Stopwords. Retrieved 2018-04-20, from https://dev.mysql.com/doc/refman/5.7/en/fulltext-stopwords.html

Newman, T. P. (2016). Tracking the release of IPCC AR5 on Twitter: Users, comments, and sources following the release of the Working Group I Summary for Policymakers. *Public Understanding of Science*, *26*(7), 1–11. https://doi.org/10.1177/0963662516628477

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the North American chapter of the Association for Computational Linguistics*, (pp. 100–108). Stroudsburg: Association for Computational Linguistics.

Nugroho, R., Yang, J., Zhao, W., Paris, C., & Nepal, S. (2017). What and with whom? Identifying topics in Twitter through both interactions and text. *IEEE Transactions on Services Computing*, 1–14. https://doi.org/10.1109/TSC.2017.2696531

Nugroho, R., Zhao, W., Yang, J., Paris, C., & Nepal, S. (2017). Using time-sensitive interactions to improve topic derivation in Twitter. *World Wide Web*, *20*(1), 61–87. https://doi.org/10.1007/s11280-016-0417-x

O'Neill, S., Williams, H. T. P., Kurz, T., Wiersma, B., & Boykoff, M. (2015). Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *Nature Climate Change*, *5*(4), 380–385. https://doi.org/10.1038/nclimate2535

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity*, *41*(2), 233–249. https://doi.org/10.1007/s11135-006-9000-3

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the international conference on Language Resources and Evaluation* (Vol. 5, pp. 1320–1326). European Language Resources Association.

Paris, C., Christensen, H., Batterham, P., & O'Dea, B. (2015). Exploring emotions in social media. In *2015 IEEE Conference on Collaboration and Internet Computing* (pp. 54–61). Hangzhou, China: IEEE. https://doi.org/10.1109/CIC.2015.43

Parker, C., Saundage, D., & Lee, C. Y. (2011). Can qualitative content analysis be adapted for use by social informaticians to study social media discourse? A position paper. In *Proceedings of the 22nd Australasian conference on information systems: Identifying the information systems discipline*, (pp. 1–7). Sydney: Association of Information Systems.

Pathak, N., Henry, M., & Volkova, S. (2017). Understanding social media's take on climate change through large-scale analysis of targeted opinions and emotions. In *The AAAI 2017 Spring symposium on artificial intelligence for social good*, (pp. 45–52). Stanford: Association for the Advancement of Artificial Intelligence.

Pearce, W. (2014). Scientific data and its limits: Rethinking the use of evidence in local climate change policy. *Evidence and Policy: A Journal of Research, Debate and Practice*, *10*(2), 187–203. https://doi.org/10.1332/174426514X13990326347801

Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 Report. *PLOS ONE*, *9*(4), e94785. https://doi.org/10.1371/journal.pone.0094785

Pistrang, N., & Barker, C. (2012). Varieties of qualitative research: A pragmatic approach to selecting methods. In Cooper, H., Camic, P. M., Long, D. L., Panter, A. T., Rindskopf, D., & Sher, K. J. (Eds.) *APA handbook of research methods in psychology, vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*, (pp. 5–18). Washington: American Psychological Association.

Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: Methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, *16*(3), 197–214. https://doi.org/10.1080/13645579.2013.774172

Ramage, D., Dumais, S. T., & Liebling, D. J. (2010). Characterizing microblogs with topic models. In *International AAAI conference on weblogs and social media*, (Vol. 10, pp. 131–137). Washington: AAAI.

Schneider, R. O. (2011). Climate change: An emergency management perspective. *Disaster Prevention and Management: An International Journal*, *20*(1), 53–62. https://doi.org/10.1108/0965356111 1111081

Sharma, E., Saha, K., Ernala, S. K., Ghoshal, S., & De Choudhury, M. (2017). Analyzing ideological discourse on social media: A case study of the abortion debate. In *Annual Conference on Computational Social Science*. Santa Fe: Computational Social Science.

Sisco, M. R., Bosetti, V., & Weber, E. U. (2017). Do extreme weather events generate attention to climate change? *Climatic Change*, *143*(1-2), 227–241. https://doi.org/10.1007/s10584-017-1984-2

Swain, J. (2017). Mapped: The climate change conversation on Twitter in 2016. Retrieved 2019-02-20, from https://www.carbonbrief.org/mapped-the-climate-change-conversation-on-twitter-in-2016

Veltri, G. A., & Atanasova, D. (2017). Climate change on Twitter: Content, media ecology and information sharing behaviour. *Public Understanding of Science*, *26*(6), 721–737. https://doi.org/10.11 77/0963662515613702

Weinberg, B. D., & Pehlivan, E. (2011). Social spending: Managing the social media mix. *Business Horizons*, *54*(3), 275–282. https://doi.org/10.1016/j.bushor.2011.01.008

Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). TwitterRank: Finding topic-sensitive influential Twitterers. In *Proceedings of the Third ACM international conference on web search and data mining*, (pp. 261–270). New York: ACM, https://doi.org/10.1145/17184 87.1718520

Williams, H. T., McMurray, J. R., Kurz, T., & Hugo Lambert, F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, *32*, 126–138. https://doi.org/10.1016/j.gloenvcha.2015.03.006

Yin, S., & Kaynak, O. (2015). Big data for modern industry: Challenges and trends [point of view]. *Proceedings of the IEEE*, *103*(2), 143–146. https://doi.org/10.1109/JPROC.2015.2388958

Affiliations

Matthew Andreotta[1,2] · Robertus Nugroho[2,3] · Mark J. Hurlstone[1] · Fabio Boschetti[4] · Simon Farrell[1] · Iain Walker[5] · Cecile Paris[2]

Robertus Nugroho
askrobertus@gmail.com

Mark J. Hurlstone
mark.hurlstone@uwa.edu.au

Fabio Boschetti
Fabio.Boschetti@csiro.au

Simon Farrell
simon.farrell@uwa.edu.au

Cecile Paris
Cecile.Paris@csiro.au

[1] School of Psychological Science, University of Western Australia, 35 Stirling Highway, Perth, WA, 6009, Australia

[2] Data61, CSIRO, Corner Vimiera and Pembroke Streets, Marsfield, NSW, 2122, Australia

[3] Faculty of Computer Science, Soegijapranata Catholic University, Semarang, Indonesia

[4] Ocean & Atmosphere, CSIRO, Indian Ocean Marine Research Centre, The University of Western Australia, Crawley, WA 6009, Australia

[5] School of Psychology and Counselling, University of Canberra, Canberra, Australia