

RESEARCH ARTICLE



OPEN ACCESS



Computational vs. qualitative: analyzing different approaches in identifying networked frames during the Covid-19 crisis

Hossein Kermani^a, Alireza Bayat Makou^b, Amirali Tafreshi^c, Amir Mohamad Ghodsi^d, Ali Atashzar^d and Ali Nojoumi^e

^aPolitical communication group, Department of communication, University of Vienna, Vienna, Austria; ^bFaculty of Electrical Engineering and Computer Science, Leibniz University, Hanover, Germany; ^cDepartment of communication, The University of Tehran, Tehran, Iran; ^dDepartment of communication, The University of Allameh Tabatabaee, Tehran, Iran; ^eMicrobiology Research Center, Pasteur Institute of Iran, Tehran, Iran

ABSTRACT

Despite the increasing adaption of automated text analysis in communication studies, its strengths and weaknesses in framing analysis are so far unknown. Fewer efforts have been made to automatic detection of networked frames. Drawing on the recent developments in this field, we harness a comparative exploration, using Latent Dirichlet Allocation (LDA) and a human-driven qualitative coding process on three different samples. Samples were comprised of a dataset of 4,165,177 million tweets collected from Iranian Twittersphere during the Coronavirus crisis, from 21 January, 2020 to 29 April, 2020. Findings showed that while LDA is reliable in identifying the most prominent networked frames, it misses to detect less dominant frames. Our investigation also confirmed that LDA works better on larger datasets and lexical semantics. Finally, we argued that LDA could give us some primary intuitions, but qualitative interpretations are indispensable for understanding the deeper layers of meaning.

ARTICLE HISTORY

Received 30 December 2021
Accepted 31 January 2023

KEYWORDS

Topic modeling; Latent Dirichlet Allocation (LDA); qualitative analysis; coronavirus; automated text analysis; twitter; framing analysis; Iran

Introduction

With the exponential growth of big data, automated methods have been used in the social sciences to analyze large text data, e.g. tweet corpora. In particular, natural language processing (NLP) offers researchers the ability to examine frames in the textual content. Researchers have explored the extent to which computational techniques can be used to discover media frames, but more research is needed if we are to understand the strengths and weaknesses of automated techniques for identifying frames in different contexts and languages. Indeed, the existing literature in this area has focused on mainstream languages such as English. As a result, we know less about other less-studied languages, such as Persian. To address these gaps, we have done a comparative analysis of the Latent Dirichlet Allocation (LDA) topic model with human-driven coding to identify networked frames on Iranian Twitter. This paper is organized as follows. First, we provide an overview of the promises and weaknesses of automated approaches to frame extraction and point out the existing gaps in this area. Then, we explain our approach to developing a comparative study to address these gaps. In the Methods section, we describe the processes of data collection, human-driven coding, and LDA implementation. Our empirical analyses focus on a large corpus of 4,165,177 Persian tweets collected during the early months of the coronavirus

CONTACT Hossein Kermani  hossein.kermani@univie.ac.at  University of Vienna, Kolingasse 14-16, Vienna 1030, Austria
 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/13645579.2023.2186566>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

crisis in Iran. Nicholls and Culpepper (2020) argue that our most important question is not the ability of computational methods to extract frames but rather what type of input data they are likely to succeed with. Therefore, we input three different samples of the same dataset to understand which ones LDA performs best on. The smallest of these samples were subject to intensive qualitative coding practice. This process was used to measure the performance of LDA in detecting frames and to validate the LDA result. After presenting the results, we discuss the contribution of this study to ongoing research on comparing computational techniques with traditional qualitative methods.

A computational approach to identifying networked frames

Frames are discursive packages people impose on their lived experiences to make sense of them (Chong & Druckman, 2007; Gamson & Modigliani, 1989). Despite the ongoing conceptual dispute among communication scholars (De Vreese, 2005; Lecheler & De Vreese, 2019), we follow Entman (1993, p. 52) in understanding framing as '[selecting] some aspects of a perceived reality and make them more salient in a communicating text.' Framing analysis is an important research strand in political science and communication studies (Lev-On, 2018). Framing analysis could enhance our knowledge of the extent to which and how media, politicians, etc., try to manipulate reality and from what point of view (Entman, 2003).

Social researchers have traditionally used qualitative methods such as thematic and content analysis in frame detection. However, with the advent of large text datasets, automated techniques, e.g. natural language processing (NLP), outperform traditional text analyses through their ability to analyze large text data with minimal time and effort (Barberá et al., 2021). Moreover, automatic methods are applicable to a wider variety of settings and can be used without extensive adaptation (Scharkow, 2013). On the other hand, the existing literature suggests that there is some exaggeration regarding the fact that machine learning approaches, e.g. topic modeling techniques, are inherently superior to the current manual process in terms of time and cost savings (Hilbert et al., 2019). Aside from several challenging machine-derived measurements (see De Grove et al., 2020; Grimmer & Stewart, 2013; Hilbert et al., 2019), machines have little understanding of human language, which brings the strengths of qualitative human-driven coding to the fore (Boyd & Crawford, 2012; Scharkow, 2013). Moreover, some studies show that a large dataset's carefully and deliberately coded sample can give similar results to what machines produce (De Grove et al., 2020).

Nevertheless, in recent years, automated methods (from supervised to unsupervised techniques) have gained importance in frame analysis. The use of supervised methods in framing analysis carries significant risk (Nicholls & Culpepper, 2020). Such methods rely on predefined frames, i.e. closed codebooks or pre-constructed scenarios, which tend to project the researchers' prior background, knowledge, and assumptions onto the dataset (Walter & Ophir, 2019). In this way, those frames that were not anticipated by the theory could be neglected, which is known as the 'drunk under the streetlight' syndrome (Nicholls & Culpepper, 2020).

Unsupervised methods are also used to identify grounded frames in text corpora. While some techniques, such as the k-means clustering algorithm and evolutionary factor analysis (EFA), are commonly used, topic modeling is much more accessible and reliable (Lind et al., 2022; Maier et al., 2018). There are several versions of topic models, especially in computer science, such as the Correlated Topic Model (CTM) (Blei & Lafferty, 2007). Nevertheless, two models remain the most important members of the topic modeling family to date: The structural Topic Model (STM) and Latent Dirichlet Allocation (LDA). The use of STM is somewhat idiosyncratic to political science and computational communication work. However, the two techniques are largely identical in their approach, e.g. both use the bag-of-words assumption, where the order of words in each document is ignored (Nicholls & Culpepper, 2020). Nevertheless, for the purposes of this study, we choose LDA.

LDA is an unsupervised probabilistic model that analyzes a set of documents and predicts the themes that make up the collection (Blei et al., 2003; Blei, 2012). LDA is increasingly used in communication studies because of its ability to quickly and efficiently explore the thematic

structures of a large set of text documents (Maier et al., 2018). A major advantage is that, in contrast to supervised methods, topics are inferred from a given collection without prior knowledge, thus avoiding the so-called streetlight syndrome (Maier et al., 2018). Another advantage of LDA is its mixed membership approach (Grimmer & Stewart, 2013, p. 18). Mixed membership means that each document can consist of more than one topic. This is a useful assumption when analyzing networked frames.

Some communication researchers are skeptical about the capabilities of LDA for frame detection, arguing that the output of the probabilistic algorithm is a topic, not a frame, as understood in communication research (Nicholls & Culpepper, 2020). A frame is a linguistic phenomenon that consequently carries a high degree of human meaning (Van Gorp, 2007). On the other hand, a topic is a computational construct representing a probability distribution over words in terms of NLP (Blei et al., 2003, p. 996). In particular, Walter and Ophir (2019) find the equation of topics with frames problematic. In their view, a topic can be a frame element but not necessarily an entire frame package. On the other hand, several studies argue that this computational construct is relatively adjacent to a frame (DiMaggio et al., 2013). Jurafsky and Martin (2020) described LDA as an extremely useful tool for discovering topical structures in documents. In other words, LDA explores the 'hidden' thematic structure of a given collection of texts (Maier et al., 2018, p. 93). In this sense, scholars have insisted on considering the themes revealed by these models as an operationalization of policy frames (DiMaggio et al., 2013; Gilardi et al., 2021). From another perspective, researchers suggest that using the results of LDA under certain conditions could be helpful in frame detection. Ylä-Anttila et al. (2022) argue that 'topics' of topic models can be used as a useful proxy for frames if (1) frames are operationalized as connections between concepts; (2) theme-specific data are used; and (3) topics are validated in terms of frame analysis.

Although existing research has mainly focused on different types of media frames (e.g. generic and topic-specific frames), little is known about the effectiveness of computational techniques in identifying a new type of framing that has emerged with the advent of social media. Networked framing is not cascading, vertical, restrictive, or organization-centered like mass media frames. Meraz and Papacharissi (2013) understand networked framing as a process whereby actors circulate information and add their own layers of information, knowledge, beliefs, and experiences on the fly. In contrast to the static and permanent nature of frames in mass media, networked frames, e.g. Twitter frames, are constantly revised, re-articulated, and re-circulated by both the masses and the elite (Meraz & Papacharissi, 2013). The particular nature of networked frames makes them even more difficult to identify using automated methods. To date, there is no further research evaluating the potential and weaknesses of computational techniques in identifying networked frames. Most researchers have focused on mainstream languages. Therefore, we do not know how automated techniques work for other, less studied languages such as Persian. This paper addresses these gaps by conducting a comparative analysis using LDA and human-driven qualitative coding in extracting networked frames in the Iranian Twittersphere.

Methods

Data and samples

We collected the data in the early months of the coronavirus crisis in Iran from Twitter REST-API. The emergence of COVID-19 on 31 December 2019, in China threw the entire world into complete shock. Almost immediately, many countries started to impose quarantine and stringent restrictions to stop the spread of Covid-19. Due to these prohibitions, many people did not go outside. Instead, they discussed this unfamiliar fact on social media. Social media data, in particular tweets, are especially well suited to studying how people frame crises of this nature in a networked space. No wonder social media are considered sites of rich 'cultural meaning-making' (Murthy, 2016). When the COVID-19 pandemic struck, social

media particularly took a prominent role, as people turned to these platforms to form collective consciousnesses, share stories and messages of hope, and discuss new events or natural disasters as they happened.

Iran was one of the countries that was seriously affected by this crisis. At the time of writing this paper, August 2022, the number of people who have died and gone infected hovers around 8 million, according to official figures. The first cases were reported on 19 February 2020. The research team had begun collecting data a month earlier because of suspicions that the virus had spread across the country earlier than the government would admit. This allowed us to study the framework even before the country was involved. Data collection began on 21 January 2020. We continued this process until 29 April 2020, for almost 4 months. We collected all tweets that contained corresponding hashtags (#Corona in Persian, with all its variations). Then, we filtered the data in Persian since the spelling of Corona is identical in Persian and Arabic. As a result, 4,165,177 tweets were collected.

In the next step, we created three different samples using social network analysis. These samples were part of our approach to validate and evaluate the results of this research. First, we created a retweet (RT) network¹ as the first sample ($n = 2,519,915$ tweets). Then, we used PageRank centrality (Easley & Kleinberg, 2010) to identify the 50 most influential users in the RT network. We formed the second sample and extracted all tweets sent by these users in the RT network ($n = 7,658$ tweets). Finally, we selected a random representative subset of sample 2. Based on the Cochran formula for calculating the sample size of a finite population, we selected 5,056 tweets (confidence level = 99%.) as sample 3. All samples were used in the computation phase, but the qualitative analysis was completed only with the last sample. While our selection of the 50 top users might be questionable in a discursive sense, that is beside the point in terms of the purposes of this study.

Human-driven coding process

To avoid the streetlight syndrome, we followed Van Gorp's (2010) two-stage mixed inductive-deductive approach to qualitative frame identification. In the first stage, we coded our content inductively, using open coding to elicit potential frames. This was followed by a deductive phase where we used the elements of these frame packages to code frame prevalence in a controlled and reproducible manner (Nicholls & Culpepper, 2020, p. 4). We added one more step to the first phase to scrutinize the process of qualitative interpretation. Subsequently, the samples were coded in three steps, namely in two inductive steps and one deductive step.

The inductive steps followed Saldana's two-step approach (2015). In the first step, coders ($n = 3$) inductively developed a codesheet by closely reading the tweets. This codesheet was broad – and it included many open codes. After completing the open coding and team discussions, the team would identify sets of codes and create a preliminary codebook. By grouping the open codes into general themes, 71 themes were identified. This codebook was used as the baseline for the second round. In this round, the coders coded the entire sample based on precedent coding (Saldaña, 2015, p. 209). The goal of this step was to discover the core themes in the tweet sample. The result was a final codesheet with 16 networked frames. The 16 identified main frames and their 71 sub-frames (frames identified in the first step) are listed in Appendix 1. The main networked frames are also discussed in the Findings section.

We used the final codesheet in the deductive step to quantify the codes across all sample tweets. The result of this deductive step was used to identify the dominant themes in the sample. Of course, we did not rely only on these numbers. We also used the qualitative notes and memos from the previous rounds to gain a deeper understanding of the grounded frames. In addition, each tweet could be coded for multiple frames. Our decision was consistent with LDA mixed-membership. In addition, quantifying the qualitative codes creates a more consistent basis for comparing LDA topics with qualitative frames.

Topic modeling

We followed the comprehensive approach of Maier et al. (2018) to implement topic modeling in this study. They provide several steps for using LDA in communication research: appropriate pre-processing, selection of appropriate parameters, and evaluation and validation of LDA output.

First, we removed stop words, punctuation marks, numbers, emojis, smileys, and other special characters. As the Persian language lacks lowercase and uppercase characters, capitalization or lack thereof does not apply here. In addition, we did not perform stemming and lemmatization on the research dataset, as previous research has demonstrated that these features do not significantly improve the performance of topic models (Jurafsky & Martin, 2020; Schofield et al., 2017). Moreover, to our knowledge, no reliable dictionary performs these features in Persian.

Furthermore, the choice of the K-value, i.e. the number of topics to be generated by the LDA algorithm, is a critical decision that significantly affects the result. However, although several K-value-estimation techniques have been implemented in Python libraries (Baden et al., 2020; Maier et al., 2018; Rodriguez & Storer, 2020), there is not yet a standard method for calculating the K-value.

This study used the *LdaModel* algorithm embedded in Python's *Gensim* library. *Gensim* (Generate Similar) is a powerful Python library that processes plain texts and uses unsupervised machine learning algorithms (Jelodar et al., 2019). *Gensim* uses topic coherence to estimate K. Topic coherence measures the score of a single topic by measuring the degree of semantic similarity between high-scoring words in the topic. We calculated topic coherence for all samples in this study to determine the possible K-values. Our analyses yielded K = 16, 12, 9, and 7 for the RT network; K = 10, 7, and 3 for sample 2; and K = 18, 15, 12, and 8 for sample 3. In addition to these automatically selected values, we decided to add values of 5 and 20 for all samples. Thus, our efforts to select the best Ks are not limited to automated statistical algorithms. We then manually reviewed all tweets to see which K score worked best for each sample. Our evaluations revealed that the consistency of the results varies for different K-values. For some values, we got more scattered topics that are difficult to group. Nevertheless, after several rounds of checking the data for all possible Ks, we chose K = 12 for RT networks, K = 10 and K = 12 for samples 2 and 3, respectively. These values provide us with the most coherent sets of topics.

Topic model output in most cases includes a family of words in different categories that may appear inconsistent and meaningless. In other words, the topics do not speak for themselves. They must be interpreted by humans (Grimmer & Stewart, 2013). While some statistical approaches, e.g. topic coherence as we used in this research, have been proposed to capture the reliability of LDA (see Maier et al., 2018), we agree with Nicholls and Culpepper (2020), who argue that there are no quantitative measures that can calculate the quality of LDA topics (p. 8). Moreover, various obstacles, such as the lack of required dictionaries and software packages, make the available metrics even more inapplicable to Persian texts.

As a result, The vast majority of the literature suggests that manual labeling based on close reading of documents is a more reliable method for assessing LDA topics (e.g. Elgesem et al., 2015; Koltsova & Koltcov, 2013; Maier et al., 2018; Rodriguez & Storer, 2020). Following Roberts et al. (2014) and Gerring (2001), Nicholls and Culpepper (2020) propose three criteria for understanding the quality of LDA topics: 1) the suitability of topics to our theoretical understanding of a frame, 2) the internal coherence of each topic, and 3) the external differentiation from other topics.

We sought to ensure these criteria through a multi-level approach. First, we crawled through the top 10 words in each topic. This was followed by examining the top 200 tweets from each topic in each sample to gain a more profound understanding of their content and coherence. Maier et al. (2018) observed that this is a time-consuming but likely consistent and indispensable step in validating LDA output. Finally, we attempted to assign representative and reliable labels to LDA topics where possible.

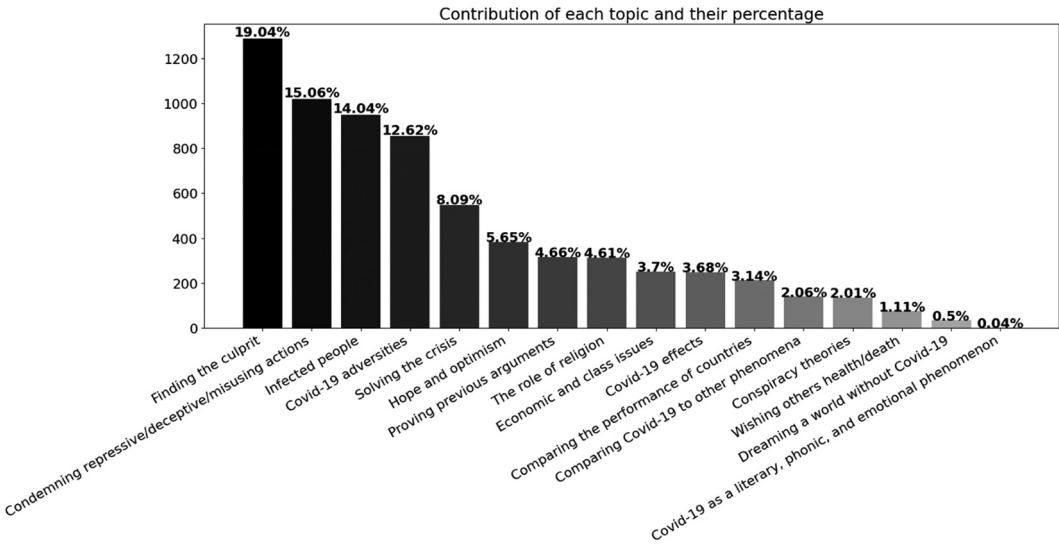


Figure 1. The frequency of networked frames in sample 3, resulting from human-driven coding.

Findings

Qualitative results

Figure 1 shows the networked frames identified by human coders and the proportion of each in sample 3. A total of 16 networked frames (with 71 sub-frames) were identified during the qualitative human-driven coding process. See Appendix 1.1 for a complete list of main frames and sub-frames.

Figure 1 shows that ‘Finding the culprit (FC),’ ‘Condemning repressive/deceptive/misusing actions (RDM),’ and ‘Infected people (IP)’ were the most dominant frames in the study sample. The details of these frames are provided in appendix 1.2. In brief, the dominant frames consisted of critical narratives focused on the Iranian state’s engagement with the crisis. Iranian users also displayed personal sentiments, stories, and affective messages mixed with news and information. To a lesser extent, Iranian users devoted their attention to the situation around the world, focusing on China’s role in getting the world into deep trouble. In addition to these frames, the research team systematically examined all networked frames and their sub-frames. The result of this systematic examination was used throughout the study to draw more solid conclusions when comparing the qualitative results with the LDA results.

As we argued in the previous sections, qualitative coding of Persian tweets is not as straightforward and direct as we originally thought. Several challenges were encountered during the manual qualitative coding in this project.

First, vowels often play a crucial role in shaping the meaning of a Persian text. Unlike in English, vowels are not usually written in Persian scripts. As a result, the tone and meaning of a word or a sentence could be misconstrued by automatic word and sentence identification programs. For example, a user tweeted a message which could be roughly translated as ‘The regime is very good at controlling the virus!’. In plain language, this sentence can be interpreted as a tweet in favor of the regime. But here is the difference. The word ‘very’ is written in Persian as خیلی. In this form, this tweet is a serious message admiring the regime. But the user changed this word to خیلی. In this form, the user put a short vowel (ـِ) on the letter خ. The user also added the letter ع to emphasize this change. In this way, the user changed the tone of the text from serious to sarcastic. Unlike خیلی, the word خیلی is negatively charged. Therefore, this tweet should be coded as an anti-regime message berating the regime for its bungling incompetence in dealing with the COVID-19

pandemic. A common theme of this tweet is that the regime's incompetence has negatively affected the public health and well-being of people.

The above example shows the complexity of interpreting a Persian tweet. It illustrates how much work it is to understand Persian text. Such challenges are not limited to the problems posed by vowels. Sarcastic messages with embedded metaphors and ironies are common on Persian Twitter. For example, one user tweeted, 'Regarding the terrible spread of the coronavirus, God has announced through Jannati that no one should come to my party.' This message was sarcastic, and one should know the meaning to interpret it correctly. First, 'God's feast' refers to Ramadan, a month devoted to fasting and prayer, during which Muslims were particularly vulnerable to the spread of COVID-19 as they would gather to pray together. Because of the COVID-19 crisis, this was not possible at that time. Therefore, this user's most recent tweet posed a challenge to this religious metaphor. But there is more. The user did not just write that God announced it but that God announced it by Jannati. Ahmad Jannati is the chairman of the Guardian Council in Iran. He is 95 years old, and Iranians usually joke about him as immortal. In this tweet, the user wanted to imply that Jannati and God are the same age, so God instructed him to announce the news. This example shows how complicated and rich a short text can be in different discursive elements.

The above example also illustrates the problem of mixed frames. Tweets usually do not contain a clear and distinct frame. Users mix different topics together and vary their tone (e.g. serious, sarcastic, etc.). This makes it more difficult to identify the frames in each tweet. To illustrate, One user tweeted, 'The situation in Iran is critical! Do not expect anything from the regime! Please follow the health instructions yourself! Imam Khomeini hospital in Saghaz is completely quarantined! 12 people suspected to be infected with #Covid -19 are in this hospital.' In this tweet, we find at least three frames: FC (IRI), IP (statistics), and COVID-19 adversity (following health guidelines). Therefore, we believe that even human coders with a closed codebook are likely to code the same text differently. It is not easy to decide which frame is dominant in this text. We tried to reduce this problem by coding for multiple frames in this study. But the problem of dominant frames still exists.

Finally, we find users changing words and phrases out of context in creative and sometimes confusing ways. A tweet shows how the process works: 'And my thorns fell! If the situation is so critical, why did they not announce it?' When we hear this message, we should not think of our own genital hair but of Iranian users' use of the slang term 'My wool fell!' where 'wool' literally refers to genital hair. It is used to express a sense of shock, like 'holy shit.' The user has used 'thorn' as a metonym for 'wool' to mitigate wool's scatological and prurient connotations. Such a text would be impossible for a machine model to interpret in this way. Moreover, human coders should have a sense of user colloquialisms and innovations to understand the message fully.

The above challenges not only make it difficult for machine algorithms to understand the meaning behind a text but also make it difficult for human coders to encode the sample. The results show that human coders can identify more nuanced, subtle, and complex frames than a computer model would pick up on. Even so, human is better at processing the meaning of the text in tweets, where the words are rather brief and used more creatively. Therefore, they can identify all frames, regardless of their proportion. For example, the frame of Eastern countries' failure (in controlling the crisis) was identified in only two tweets. This frame was later classified with others in the main theme Comparison of Countries' Performance. Nevertheless, such frames tend to be eliminated by the computer model because of their small percentage. We will discuss in the next section to what extent the machine results are comparable to the human-driven codes.

LDA outputs

Below we present the results of the LDA topic model for each sample. Following Maier et al. (2018), we summarize the LDA output for each sample in a table. In addition, we have added additional columns to allow the analysis of the comparison of LDA topics with qualitative networked frames based on the work of Nicholls and Culpepper (2020). The first three columns contain some statistical information

Table 1. LDA result on sample 1 (RT network).

K	No of documents	Share	Label	Top words	Degree of coherence	Degree of compatibility	Compatible frame(s)
2	40378	0.114974	-	Life, alcohol, Nowruz, dad, problem, pray, easy, take, call, solution	Low	-	-
0	34008	0.096836	News and information	Hospital, Qom, infected, Tehran, infection, university, province, suspicious, closed, death	High	High	Infected people
7	31767	0.090455	Statistics	Health, announcement, statistics, country, death, ministry, infection, number, infected, minister	High	High	Infected people
1	30956	0.088146	Inefficiency of the state	Crisis, Islamic, republic, government, Rouhani, media, management, the state, news, lie	Medium	Medium	FC (IRI) Solving the crisis (Iran's power, government's good performance) RDM (media against Iran) Solving the crisis (government's good performance)
3	29359	0.083598	The state's confrontation with the COVID-19 crisis	Confrontation, fight, regime, identification, Khamenei, IRIGC, war, kit, France, England	Medium	Medium	FC (IRI) Covid-19 adversities (following sanitary guidelines) Covid-19 adversities (risks and properties) Covid-19 effects (global)
5	28806	0.082024	Prevention and health advices	Mask, treatment, prevention, vaccine, follow, transmitting, probability, advice, disinfection, infectious	High	High	FC (IRI, China)
8	28400	0.080868	Covid –19 properties	Though, authorities, cough, region, nose, fever, warm, inhale, stress, type	High	High	-
6	26678	0.075964	Coronavirus situation and effects in other countries	Outbreak, United States, circumstances, globe, disaster, Trump, spreading, prevention, conditions, decrease	Medium	Medium	-
4	25746	0.073311	Poor management	China, quarantine, world, country, control, Europe, dirty, abroad, cancel, Wuhan	Medium	Medium	-
9	25619	0.072949	-	Esfand*, Saturday, Korea, weather, my heart, day, dad, SMS, mother, southern	Very low	-	-
10	25313	0.072078	Religious and ideological messages	People, God, prison, defeat, political, flood, earthquake, shrine, west, heart	Medium	Medium	The role of religion (absurdness, importance)
11	24161	0.068797	-	People, election, Bahman, nation, God, Aban*, authorities, defeat, hard, viral	Very low	-	-

* Aban and Esfand are the names of some months in Persian calendar

about each topic. The proportion and number of the documents show the dominant topics in each sample in descending order. Each table has a column with a label. As we explained in the Methods section, these labels were assigned to each topic after a close examination of the top keywords and a qualitative interpretation of the top tweets. The next column shows the top 10 words in each topic. The ‘degree of coherence’ (DoC) column shows the degree of homogeneity of tweets in a topic. If the tweets culminating in a topic denote the same theme (represented by its label), the topic is more coherent. The last two columns show the degree of compatibility (DoCom) of a topic with some networked frames and equated frames. In other words, DoC indicates whether a topic is homogeneous enough to be described by a label. DoCom indicates whether this label has an equivalent in qualitative frames. In this way, each table provides a comprehensive understanding of the LDA result and its compatibility with the qualitative examination.

Table 1 shows some interesting results on the performance of LDA in the RT network. Despite the research team’s attempts to assign meaningful and reliable labels to the topics after a thorough qualitative analysis, some topics do not lend themselves to meaningful or reliable names. In fact, they are too incongruent to be labeled (topics 2, 9, and 11). In some cases, LDA classified some tweets with the same discursive practices, not frames. For instance, topic 2 is dedicated to users’ sarcasms and criticisms, mostly in the form of jokes and funny tweets. However, these messages were aimed at different targets: the state, religious figures and beliefs, and people, including the users themselves. As such, they contained various networked frames (sub-frames) such as FC (IRI, people, religious figures), RDM (China’s mendacity and secrecy, authorities abusing the situation, emphasis on not using security forces), and the role of religion (absurdity). Nevertheless, while LDA was able to identify that the underlying discursive practice of these tweets (sarcasm) was the same, it was not able to identify that the targets of the sarcastic tweets varied. While this result shows the weakness of LDA in detecting frames, it proves that LDA can be usefully employed to detect the underlying discursive practices in textual data. Such groupings may not be of interest from a frame analysis perspective, but they could help researchers study discursive practices in large datasets more efficiently. Research on stylistic variation in texts could particularly benefit from this strength of LDA. Nonetheless, there were more cases where the LDA topics did not even share the same discursive practices, and the included tweets were largely scattered.

Further analysis showed that LDA usually provides us with more straightforward and clear themes. Even coherent topics (at different levels) have some general labels. They can give us an initial understanding of the content but not an exact mapping. Table 1 shows that LDA performs better on lexical meanings than on compositional semantics.² In other words, the performance of LDA in classifying tweets containing simpler and one-dimensional information was more satisfactory. Examples include topics 5 and 8, in which users shared health instructions (including COVID-19 properties). Such messages rarely contained hidden political and cultural meanings. As a result, LDA successfully classified these topics within meaningful boundaries.

In contrast, topics with compositional meanings were less coherent and consistent with single and exclusive frames. Topic 1 includes many tweets that discuss the government’s inability to control the crisis. Many tweets with political and social connotations could be found within this topic. In general, two main groups fight over this subject: those defending the regime, and those questioning it. However, the human coders were able to find more nuanced ways to discuss these issues. For example, they were able to detect disagreement about the regime’s performance on these issues.

Moreover, the different DoC levels show that in all topics, there were tweets that deviated from the main theme. Such tweets could be found even in topics with a high coherence level. However, the higher the DoC score, the greater the chance of finding a representative label and equivalent networked framework for a topic. In the RT network, all topics with a DoC higher than ‘low’ have an analogous networked frame. Furthermore, while coherent topics could form some frames besides the identified networked frames, in this case that was not the case. All coherent topics in this sample corresponded to some networked frames and did not create any new frames.

On the contrary, there are several networked frames that were not found in the LDA topics. For example, none of the themes related to proving previous arguments or economic and class issues. This finding shows that using only LDA results to infer the underlying meaning in a text corpus can mislead researchers and lead to incorrect interpretations of such results. Our analyses show that qualitative interpretations are indispensable to understanding and identifying all embedded meanings in textual content, e.g. tweets. Nevertheless, it is worth noting that LDA topics might be associated with top frames. This means that LDA overlooked lesser featured networked frames.

Our qualitative interpretations also showed that most of the topics were highly interlinked, e.g. topics 2, 1, 3, and 4. Consistent with previous findings, this result indicates that LDA cannot generate separate and comprehensive topics. Our analyses indicated that a predominant theme was supported in one topic but was also present in lesser amounts in several other topics. This was mainly due to the mixed membership approach. It also happens with human-driven coding, but our analyses yielded a greater number of overlapping topics. For example, topics 0 and 7 or 5 and 8 are highly identical, and it makes sense to treat them as a single topic. LDA missed this. Moreover, the most common qualitative frames are found in all LDA topics, e.g. FC and IP. This shows that LDA is a good technique for finding the most dominant themes but at the expense of the less prominent topics.

Samples 2 and 3: Tables 2 and 3 in Appendix 2 provide information on how LDA works for the next samples. Since these tables are very large and contain more or less the same information as Table 1, we have included them in a separate appendix.

The results show that the number of topics with lower DoC is higher in samples 2 and 3. This shows that the performance of LDA does not improve when the number of documents is reduced. Surprisingly, LDA performs worse in such cases. This result also shows that we can easily apply LDA to the RT network or to the entire datasets in other studies and obtain the same or even better results.

The analysis of samples 2 and 3 also confirms most of the results of the analysis of the RT network (sample 1). For example, LDA also works better for lexical meanings in these samples or in the case of overlapping topics. In particular, the results show that LDA is a suitable technique for finding the most dominant themes in the text. Moreover, the discursive and qualitative analyses of all samples confirm that the top keywords cannot provide even a half-acceptable understanding of the content of the tweets. There is a significant gap between the top words and the actual tweets on each topic. One reason for this could be the fact that there are some tweets in each topic with a high contribution percentage (CP),³ which probably influence the process of selecting the top keywords. For example, topic 5 in Sample 3, a topic with a very low DoC, has one tweet with a CP of 24. Then the CP drops significantly to 15. The first tweet contains words like 'apparatus, IRGC, test, and produce.' Most tweets on this topic lack these words, but they are among the top keywords. For example, the 6th tweet on this topic, with a CP of 14, does not contain any of the top keywords.

The gap between the top keywords and the grounded themes highlights the importance of qualitative interpretation of LDA topics. In particular, we cannot infer the meaning of a topic by looking at the top word in the Persian text. Therefore, human examination of the tweets per LDA topic is an indispensable phase in evaluating the LDA results. In the last section, we will discuss these and other results to show how this research advances our understanding of computational methods vis-a-vis qualitative interpretations.

Discussion

While computational techniques have been used extensively in the social sciences to challenge, supplement, or extend long-standing theories, their application in frame analysis is still uncertain (van Atteveldt & Peng, 2018). Nonetheless, most research in this area has focused on the automated detection of different types of frames, e.g. generic and context-specific frames in news articles (Burscher et al., 2014; Walter & Ophir, 2019). As a result, researchers have relatively failed to investigate the capabilities of topic models in automatically detecting networked frames, which are another form of framing in social media. Other relevant questions also remain unanswered, such as

to what extent does a single LDA topic correspond to a networked frame? What impact does the imported dataset have on the reliability of LDA results? Since previous research has mainly focused on English and other common languages, the ability of topic models to interpret other languages remains unknown. We focus on Persian Twitter to fill this gap as another contribution of this study. We have attempted to advance the existing literature by combining a comparative approach based on LDA topic models and human-driven qualitative coding.

First, this research provides more insight into the extent to which a topic can be considered a frame. While LDA categorizes tweets into loose categories, it cannot produce pure frames with clear boundaries. Frames are something more than LDA raw topics. Indeed, our results support previous studies that challenged the idea of equating LDA topics with frames (Walter & Ophir, 2019). Consistent with previous studies (Nicholls & Culpepper, 2020), we have shown that there is always a gap between LDA topics and grounded frames. There is no reliable statistical method for measuring this distance to date. We emphasize that human interpretation is crucial in deciding how far LDA outputs are from networked frames.

In our three cases, only four of the topics have DoC scores of the highest level. Therefore, they are able to more easily find their correspondence in networked frames. Regardless of the low number of these topics, there are even tweets that deviate from the predominant themes in such highly coherent cases. It is more difficult to equate topics with lower DoC with networked frames. In such cases, manual review of tweets and investigation of topics are essential. The high number of topics with very low, low, and medium DoC confirms that LDA is not a powerful method to identify networked frames at a satisfactory level.

Our analyses also indicated that some topics should be merged to form a frame. In contrast, other topics span multiple frames. For topics with very low and low DoC, the number of frames in a topic is so high that no label can even describe it. It also shows that LDA topics usually overlap. The same frames can be found in many topics, which significantly reduces the performance of LDA in frame detection.

Nevertheless, the fact of overlapping frames provides an advantage in automated frame detection. The results show that LDA can identify the most important topics in large text data. This study's qualitative analysis showed that FC, RDM, and IP are the most networked frames on Persian Twitter. LDA topics could be largely equated with them. Thus, while LDA does not detect less dominant frames, it can be used to quickly identify the most prominent frames. The results also reflect previous studies arguing that the topic model works better with lexical meanings than with compositional semantics (Rodriguez & Storer, 2020). Overall, we argue that LDA is a satisfactory method for analyzing the lexical web of meanings and identifying top frames. As a result, we question the idea of equating topics with frames, as has been claimed in some research (DiMaggio et al., 2013; Gilardi et al., 2021). However, we have demonstrated that this method effectively extracts the most important themes in large text data. If researchers want to go further and better understand their data, human coding is required.

Human interpretation is important for other purposes as well. We have shown that there is a discrepancy between the top keywords in each topic and the grounded themes. A qualitative review is the most reliable way to measure this gap. In line with Maier et al. (2018), we believe that eliminating human-driven analysis seriously challenges the validity of LDA results. This means that the trade-off between streetlight syndrome and fully automated systems is inevitable. A researcher should decide to what extent to rely on fully automated systems depending on research goals, time, and other resources.

This research also confirms the existing literature that argues that computational methods work better with clearly delineated and general concepts (De Grove et al., 2020; Grimmer & Stewart, 2013). LDA provided such topics in all cases. This is also consistent with our findings mentioned above. Moreover, our study showed that the larger the sample size, the more reliable LDA produces results in the RT network. While we expected LDA to perform more efficiently on smaller datasets, its performance was worse on smaller and

more coherent samples. Nicholls and Culpepper (2020) also argued that topic model outputs are not highly acceptable even with more targeted datasets. While they used datasets from different sources, we imported samples of the same dataset to evaluate the automated results. However, the results showed that LDA performed better on a larger, logically less-targeted dataset.

While our focus was on comparing LDA outputs with qualitative interpretations, which is a gap in the existing literature, this study also extended our knowledge of applying computational methods to less dominant languages. LDA shows slight weakness for Persian compared to other prominent languages such as English (Burscher et al., 2014; Nicholls & Culpepper, 2020; Rodriguez & Storer, 2020; Walter & Ophir, 2019). Compared to other languages, Persian has fewer computational tools and resources. For example, while Rodriguez and Storer (2020) used LIWC⁴ software for semantic evaluation of their results, we cannot do the same for Persian. Moreover, to our knowledge, there is no dictionary for tokenization or lemmatization in Persian. That could impair the efficiency of computational techniques in Persian. Therefore, we believe that more tools should be developed for lesser-known languages to enable researchers in such areas to contribute more effectively to ongoing research. In this study, we omitted lemmatization and tokenization, as they are too resource intensive and previous studies have not found them to affect the performance of the model. However, creating and using a tool to do these things can deepen our understanding.

In line with Nicholls and Culpepper (2020), who argue for more research to understand the efficiency of LDA in analyzing different input data, we sought to provide more empirical evidence by studying networked frames on Twitter in a little-studied domain and on three different samples. As mentioned earlier, this research shows that the performance of LDA in detecting networked frames is similarly efficient as in extracting other types of frames in news articles (Burggraaff & Trilling, 2020; Burscher et al., 2014; Walter & Ophir, 2019). This finding proves that the weakness of LDA in understanding human meaning is something fundamental. It is not likely to depend on the type of frame studied, notwithstanding some minor differences in results. The results in three different samples underscore this. The results of performing LDA on all samples were not completely acceptable. Moreover, we have shown that while LDA performs slightly better on English and other major languages, its performance on Persian texts is weak. Further research is needed to demonstrate how LDA can be improved to produce more coherent and reliable topics across different languages and datasets.

We also acknowledge that various decisions, such as the choice of K , may affect the results. However, this limitation is inevitable as current topic modeling approaches have not satisfactorily addressed this issue (Maier et al., 2018). In addition, we used Gensim to perform LDA. Using other tools, such as the *lda* package in R, could lead to different results. However, we believe that using such methods would likely not significantly affect the results of this study.

Despite these caveats, our study provides LDA as an acceptable technique for detecting the most important frames, especially those containing lexical meaning, even in larger datasets in Persian. Of course, the results showed that the performance of the LDA model does not satisfy all the requirements necessary to be considered satisfactory. The results also showed the importance of human interpretation in several steps of automated frame extraction. LDA is reliable in deriving the main grounded themes when researchers want a quick and less accurate analysis. Systematic and intensive qualitative coding (of text) and interpretation (of LDA outputs) are essential to identify more detailed and nuanced grounded frames. While the study mainly concerns monolingual topic modeling, it could be employed in developing approaches in multilingual topic modeling. Lind et al., (2022) discussed the various approaches to multilingual topic modeling. They proposed the 'Polylingual Topic Model' (PLTM) as a valid approach in this field. The result of this investigation could be used in line with this approach to enhance multilingual topic modeling including big Persian text. In this way, this study sets more directions for further investigation, particularly in understudied languages.

Notes

1. The retweet network (RT) is a directed graph G. The nodes are Twitter users and the edges are retweet links between users. An edge is directed from user A, who posts a tweet, to user B, who retweets it.
2. In short, lexical semantics refers to the meaning attributed to individual words in a text, as well as the disambiguation of such words through contextualization (Johnson, 2007). Compositional semantics refers to the way words are combined to create larger meanings (Pelletier, 1994).
3. The CP shows the influence of a tweet on the composition of a topic. A higher value of CP means that the tweet plays a more important role in the formation of a topic by LDA.
4. Linguistic Inquiry and Word Count (LIWC) is a word counting software program that refers to a dictionary of grammatical, psychological, and content word categories.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This project (Hossein Kermani) has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101029945.) This paper reflects only the authors' views; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains.

Notes on contributors

Hossein Kermani, Ph.D. in Social Communication Science (University of Tehran), is a MSCA postdoctoral researcher at the University of Vienna. He is studying social media, digital repression, computational propaganda, and political activism in Iran. His research mainly revolves around the discursive power of social media in making meaning, shaping practices, changing the microphysics of power, and playing with the political, cultural, and social structures in Iran. Following this line of research, he has done several studies to shed light on Iranians' everyday life on social media.

Alireza Bayat Makou is a Master's student of Computer Science at the Leibniz University of Hannover. He has worked in the field of Machine Learning and Data Mining for 3 years and his Bachelor's thesis is on semi-supervised bot detection on Twitter. His research interests are Natural Language Processing, Human-inspired AI, Cognitive Linguistics, and Social Network Analysis.

Amirali Tafreshi is a Ph.D. candidate at the University of Tehran in Iran. His research interests include Big Data, discourse analysis, text mining, and data mining regarding epistemological issues in social network sites as well as fake action in social media.

Amir Mohamad Ghodsi is a Master's Student at the University of Allameh Tabataba'i. His Master thesis is about discourse networks and the role of brokers in Twitter conversations. His research interests include social network analysis, discourse analysis, social media discourses, digital governance.

Ali Atehzar is an Iranian researcher in digital journalism. His research concentrates on the intersection of social media use and journalism practices. Furthermore, he has studied the fields of social listening, sentiment analysis, and social network analysis.

Ali Nojourni graduated from Tehran University in Iran and then continued his education up to Ph.D. at the University College London affiliated to the University of London. He currently works at the Microbiology Research Center, Pasteur Institute of Iran (IPI). Also, he has been working on issues related to innovation, social media, technology parks and incubators.

References

- Baden, C., Kligler-Vilenchik, N., & Yarchi, M. (2020). Hybrid content analysis: TOWard a strategy for the theory-driven, computer-assisted classification of large text corpora. *Communication Methods and Measures*, 14 (3), 165–183. <https://doi.org/10.1080/19312458.2020.1803247>

- Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, 16(1), 1–18. <https://doi.org/10.1080/19312458.2021.2015574>
- Barberá, P., Boydston, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated text classification of news articles: A practical guide. *Political Analysis*, 29(1), 19–42. <https://doi.org/10.1017/pan.2020.8>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1). <https://doi.org/10.1214/07-AOAS114>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://www.jmlr.org/papers/v3/blei03a.html>
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Burggraaff, C., & Trilling, D. (2020). Through a different gate: An automated content analysis of how online news and print news differ. *Journalism*, 21(1), 112–129. <https://doi.org/10.1177/1464884917716699>
- Burscher, B., Odijk, D., Vliegthart, R., de Rijke, M., & de Vreese, C. H. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8(3), 190–206. <https://doi.org/10.1080/19312458.2014.937527>
- Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10(1), 103–126. <https://doi.org/10.1146/annurev.polisci.10.072805.103054>
- De Grove, F., Boghe, K., & De Marez, L. (2020). (What) can journalism studies learn from supervised machine learning? *Journalism Studies*, 21(7), 912–927. <https://doi.org/10.1080/1461670X.2020.1743737>
- De Vreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal*, 13(1), 51–62. <https://doi.org/10.1075/idjdd.13.1.06vre>
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41(6), 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>
- Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511761942>
- Elgesem, D., Steskal, L., & Diakopoulos, N. (2015). Structure and content of the discourse on climate change in the blogosphere: The big picture. *Environmental Communication*, 9(2), 169–188. <https://doi.org/10.1080/17524032.2014.983536>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Entman, R. M. (2003). Cascading activation: Contesting the white house's frame after 9/11. *Political Communication*, 20(4), 415–432. <https://doi.org/10.1080/10584600390244176>
- Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *The American Journal of Sociology*, 95(1), 1–37. <https://doi.org/10.1086/229213>
- Gerring, J. (2001). *Social science methodology*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815492>
- Gilardi, F., Shipan, C. R., & Wüest, B. (2021). Policy diffusion: The issue definition stage. *American Journal of Political Science*, 65(1), 21–35. <https://doi.org/10.1111/ajps.12521>
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>
- Hilbert, M., Blumenstock, J., Diesner, J., Frey, S., González-Bailón, S., Pan, J., Shen, C. C., Smaldino, P. E., & Zhu, J. J. H. (2019). Computational communication Science: A methodological catalyzer for a maturing discipline. *International Journal of Communication*, 13(2019), 3912–3934.
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>
- Johnson, K. (2007). An overview of lexical semantics. *Philosophy Compass*, 3(1), 119–134. <https://doi.org/10.1111/j.1747-9991.2007.00101.x>
- Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing* (3rd ed.). <https://web.stanford.edu/~jurafsky/slp3/>
- Koltsova, O., & Koltcov, S. (2013). Mapping the public agenda with topic modeling: The case of the Russian livejournal. *Policy & Internet*, 5(2), 207–227. <https://doi.org/10.1002/1944-2866.POI331>
- Lecheler, S., & De Vreese, C. H. (2019). *News framing effects*. Routledge.
- Lev-On, A. (2018). The anti-social network? Framing social media in wartime. *Social Media and Society*, 4(3), 205630511880031. <https://doi.org/10.1177/2056305118800311>
- Lind, F., Eberl, J. -M., Eisele, O., Heidenreich, T., Galyga, S., & Boomgaarden, H. G. (2022). Building the Bridge: Topic modeling for comparative research. *Communication Methods and Measures*, 16(2), 96–114. <https://doi.org/10.1080/19312458.2021.1965973>

- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Meraz, S., & Papacharissi, Z. (2013). Networked Gatekeeping and Networked Framing on #egypt. *The International Journal of Press/politics*, 18(2), 138–166. <https://doi.org/10.1177/1940161212474472>
- Murthy, D. (2016). The ontology of tweets: Mixed-Method approaches to the study of twitter. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 559–572). SAGE Publications Ltd. <https://doi.org/10.4135/9781473983847.n33>
- Nicholls, T., & Culpepper, P. D. (2020). Computational identification of media frames: Strengths, weaknesses, and opportunities. *Political Communication*, 1–23. <https://doi.org/10.1080/10584609.2020.1812777>
- Pelletier, F. J. (1994). The principle of semantic compositionality. *Topoi*, 13(1), 11–24. <https://doi.org/10.1007/BF00763644>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Rodriguez, M. Y., & Storer, H. (2020). A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. *Journal of Technology in Human Services*, 38(1), 54–86. <https://doi.org/10.1080/15228835.2019.1616350>
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. SAGE.
- Scharkow, M. (2013). Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality & Quantity*, 47(2), 761–773. <https://doi.org/10.1007/s11135-011-9545-7>
- Schofield, A., Magnusson, M., & Mimno, D. (2017). Pulling out the stops: Rethinking stopword removal for topic models. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, Valencia, Spain, 432–436.
- van Atteveldt, W., & Peng, T. -Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, 12(2–3), 81–92. <https://doi.org/10.1080/19312458.2018.1458084>
- Van Gorp, B. (2007). The constructionist approach to framing: Bringing culture back in. *Journal of Communication*, 57(1), 60–78. https://doi.org/10.1111/j.1460-2466.2006.00329_5.x
- Van Gorp, B. (2010). Strategies to take subjectivity out of framing analysis. In P. D'Angelo & J. A. Kuypers (Eds.), *Doing news framing analysis* (pp. 100–125). Routledge.
- Walter, D., & Ophir, Y. (2019). News Frame analysis: An inductive mixed-method computational approach. *Communication Methods and Measures*, 13(4), 248–266. <https://doi.org/10.1080/19312458.2019.1639145>
- Ylä-Anttila, T., Eranti, V., & Kukkonen, A. (2022). Topic modeling for frame analysis: A study of media debates on climate change in India and USA. *Global Media and Communication*, 18(1), 91–112. <https://doi.org/10.1177/17427665211023984>