



# Characterizing partisan political narrative frameworks about COVID-19 on Twitter

Elise Jing<sup>1,2</sup> and Yong-Yeol Ahn<sup>1,3,4\*</sup> 

\*Correspondence: [yyahn@iu.edu](mailto:yyahn@iu.edu)

<sup>1</sup>Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, USA

<sup>3</sup>Indiana University Network Science Institute (IUNI), Bloomington, IN 47408, USA  
Full list of author information is available at the end of the article

## Abstract

The COVID-19 pandemic is a global crisis that has been testing every society and exposing the critical role of local politics in crisis response. In the United States, there has been a strong partisan divide between the Democratic and Republican party's narratives about the pandemic which resulted in polarization of individual behaviors and divergent policy adoption across regions. As shown in this case, as well as in most major social issues, strongly polarized narrative frameworks facilitate such narratives. To understand polarization and other social chasms, it is critical to dissect these diverging narratives. Here, taking the Democratic and Republican political social media posts about the pandemic as a case study, we demonstrate that a combination of computational methods can provide useful insights into the different contexts, framing, and characters and relationships that construct their narrative frameworks which individual posts source from. Leveraging a dataset of tweets from the politicians in the U.S., including the ex-president, members of Congress, and state governors, we found that the Democrats' narrative tends to be more concerned with the pandemic as well as financial and social support, while the Republicans discuss more about other political entities such as China. We then perform an automatic framing analysis to characterize the ways in which they frame their narratives, where we found that the Democrats emphasize the government's role in responding to the pandemic, and the Republicans emphasize the roles of individuals and support for small businesses. Finally, we present a semantic role analysis that uncovers the important characters and relationships in their narratives as well as how they facilitate a membership categorization process. Our findings concretely expose the gaps in the "elusive consensus" between the two parties. Our methodologies may be applied to computationally study narratives in various domains.

**Keywords:** COVID-19; Political discourse; Social media; Framing; Semantic role analysis

## 1 Introduction

Human beings make sense of the reality around them by constructing narratives using what they see, hear, and encounter [49]. However, narratives that evolve around different identities, cultures, religions, etc. are often at odds with each other [42]. One of the areas where contrasting narratives fiercely collide and fight is politics. Political communication often happens through narratives and stories, rather than logical reasoning [5, 24].

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

These narratives have a tremendous power in shaping people's stances and behaviors on important social issues [34]. In the age of social media, narratives can be circulated, mutated, and amplified with incredible intensity and speed [4, 15]. For example, during the COVID-19 crisis, social media sites including Twitter and Facebook are used by the anti-mask and anti-vaccine groups to organize multiple anti-mask protests [21]. The anti-mask and anti-vaccine narratives, accompanied by conspiracy theories, fake news, and unverified anecdotes, discouraged mask usage and vaccination heavily, which might have led to the loss of hundreds of thousands more lives [56]. Furthermore, such narratives often lead to collisions between partisan beliefs that strengthen political polarization [52]. As can be seen in the case of the pandemic narratives, understanding social conflicts and polarization is often impossible without understanding diverging narratives.

While many different definitions of narratives have been proposed, here we draw our definition of political narratives from the Narrative Policy Framework which defines a narrative as having "(i) a setting or context; (ii) a plot that introduces a temporal element, providing both the relationships between the setting and characters, and structuring causal mechanisms; (iii) characters who are fixers of the problem (heroes), causers of the problem (villains), or victims (those harmed by the problem); and (iv) the moral of the story" [30]. Using this definition, we can identify narratives used by politicians and political parties to convey their morals. For example, as summarized by Haidt et al. [23], the liberals and conservatives in the United States have the following contrasting narratives that their followers adopt: "*The majority of people used to be oppressed, treated unequally and with unjust; however, the courageous people fought against the powerful and freed a lot of the oppressed people. We as successors must continue their errand and fight for more equality in the society.*" (the "liberal progress" narrative from the liberals) Or,

*"People used to live in harmonious communities tied together by faith and tradition, however, this is broken by the modern lifestyle, science and the industrial revolutions. We must therefore hold to our values and resist these forces."* (the "community lost" narrative from the conservatives).

These narratives are not objective descriptions of history, but interpretations of the reality that fit with people's political beliefs. Additionally, even though the narratives are different and may be at conflict with each other, each of them achieve internal consistency and coherence [54], which makes them effective [17].

Traditional studies of political narratives are often based on political discourse analysis (PDA). PDA studies the role of spoken and written language in politics [10], focusing on the rhetoric features, styles, logic, metaphors, and contents of the political language [7]. While traditional PDA often draws its material from formal political language such as public speeches from national leaders [6], legislative debates [47], and newspaper articles [19], social media has gained increasing attention as many politicians turn to social media sites as their main online platforms for public communication [11], where they respond to issues raised by the media and public and promote their own agendas [2].

Among social media sites, Twitter has been one of the most important platforms for political communication during the last decade [48]. Politicians use Twitter to not only broadcast to, but also interact with and attract their audience directly [12, 45]. Such direct communication often benefits politicians; for instance, the usage of Twitter may increase the amount of donation that a politician receives and benefit their campaigns [26, 36]. For these reasons, as well as the succinct, swift, and amplifying nature of the Twitter discourse,

many politicians have been effectively using their tweets to spread their narratives [28]. While there have been studies on the hashtags [25], sentiments [33], and moral values [29] from the politicians' tweets, systematic studies of political narratives on Twitter are rare, although political science increasingly adopts text analysis methods [58].

While the scale of social media data provides great opportunities, it also poses many challenges. Traditional approaches to narrative studies through "close reading" [43] may allow deep understanding of narratives, but are labor-intensive and rely on subjective judgements. Such constraints may be addressed by computational methods, where we can automatically identify patterns in large datasets. For example, Shurafa et al. [53] studied hashtags and rhetoric devices used by U.S. Twitter users leaning towards the Democratic or Republican parties, and identified their framing preference regarding the COVID-19 crisis; Green et al. [20] identified key words from politicians' tweets, and showed that partisanship can be inferred by their word usage. However, these studies rely on word-level analysis and Twitter hashtags, while in-depth analysis of such narratives are rarely attempted.

Additionally, the brief nature of Twitter postings makes it unlikely for each of them to contain a complete narrative. Rather, each tweet may contain "fragments" of a larger narrative. While human readers can often infer the overarching narrative based on their reading of other tweets and background knowledge, it is difficult for computational models to do so. A similar challenge is identified by Tangherlini, et al. in their study of online conspiracy theories [55], where the complete narrative is often scattered in multiple short postings. Their response is to consider a narrative framework consisting of "cast of characters, the relationships between those characters, and the contexts in which those relationships arise", which individual postings sample from. Similarly, we consider two narrative frameworks for the U.S. Democratic and Republican parties, which are conceptualized by the aggregation of each party's tweets respectively, containing the contexts, characters, and relationships used by each party's narrative. Individual tweets draw their "ingredients" from this larger space, and allude to the complete narrative therein.

Following this intuition, we characterize the narrative frameworks for the two parties by analyzing collections of their tweets to identify three elements: context, framing, and characters and relationships. Our approach has two key differences from Tangherlini et al. [55] in that (i) we consider the context as the main topics and issues that each party engages with, instead of characterizing it with relationships. (ii) we examine framing separately as we consider it to be a central piece of political discourse, which shapes *how* political narratives are conveyed to the audience independent from *what* is communicated (we further elaborate on this below). In doing so, we aim to provide more nuanced analysis beyond the common term-based approaches.

First, we analyze the word frequencies in the tweets and identify the most characteristic words used by each party; this simple method allows us to see the most contrasting differences in each group's narratives at the level of "ingredients", which set up the contexts for their narrative frameworks.

Next, we ask how they are *framed*. Framing analysis is a central piece in political discourse analysis [57]. Framing is about selectively presenting some aspects of an issue and make them more salient, in order to promote certain values, interpretations, or solutions [13]. For example, on the undocumented immigration issue, the Democrats often focus on the human rights aspect, while the Republicans often focus on the legality. Similar diver-

gence in framing across major political issues are widely recognized from the two parties. Hemphill et al. [25] showed that using Twitter data, a machine learning classifier can be trained to easily predict the partisanship of a politician from the frames that they use.

Traditional studies on political framing mostly rely on manual content analysis and discourse analysis to detect frames from texts [46], and are therefore confined to a small set of frames because the process is labor-intensive. Here, we employ the FrameAxis model [35], which was developed to facilitate this process by using word embeddings and antonymous word pairs. With this method, the overall *bias* (the alignment with a frame) and *intensity* (the strength of a frame) of a document with respect to many “microframes” can be computed. We apply the FrameAxis to identify important frames in the politicians’ tweets about COVID-19. For example, we found the microframe *dead vs. live* is used to discuss the deaths related to COVID-19, and the microframe *fast vs. slow* is used to discuss the spread of COVID-19.

Finally, we analyze the characters and relationships in each party’s narrative framework. We focus on the relationships captured by actions, the Agent (the one who initiates an action), and the Patient (the one being affected or the recipient of the action). For example, in the sentence *Mary sold the book*, *Mary* is the Agent, *book* is the patient, and the relationship is captured in the verb *sold*. The Agent–Patient–Action pattern appears to be universal in human cognition [8].

We use semantic role labeling (SRL) models to automatically identify Agents, Patients, and verbs in our dataset. Originated in traditional linguistics [16], SRL has attracted much interest from Computational Linguistics, leading to the development of large annotated corpora such as FrameNet [1] and PropBank [32]. Trained on such corpora, modern NLP platforms such as SENNA and AllenNLP can perform the SRL task with high accuracy [9, 18]. With the development of deep learning, SRL has been successfully applied to analyze events either as a stand-alone work or as part of an NLP pipeline [14, 27, 37]. As different semantic roles can refer to the same underlying character (e.g. “Kamala Harris” and “Vice President Harris” refer to the same person), other NLP techniques such as named entity recognition and coreference resolution are sometimes used to aggregate similar semantic roles and verbs [55].

We are especially interested in the characters that play key roles in the COVID-19 crisis and the relationships between them. For example, when the Democrats use the word “help”, *who are to be helped and who will help them?* Furthermore, how are these agents different in the Republican tweets? Our analysis shows the most prominent Agents and Patients in the Democratic/Republican narratives about the pandemic as well as the partisan differences. In particular, we identify a membership categorization process, namely the division between “us” and “them”, where “us” is often projected as the heroes and “them” as the villains in each party’s narratives. As the most general membership categories, they help people to organize their everyday knowledge and actions [51]. For example, the former President Donald Trump frequently used this categorization in his campaign: “They hate me. They hate you. They hate rallies and it’s all because they hate the idea of MAKING AMERICA GREAT AGAIN!” [38]. Our analysis reveals a similar process where memberships are established by the interaction between characters.

Overall, our work applies a set of computational methods to comprehensively describe the elements making up the two parties’ narrative frameworks, as well as how they diverge. Such divergence may be one of the “wedges” that exacerbate polarization in U.S. politics.

The combination of methods we employed here to explore political narratives are not limited to politics. The code we develop and publish would allow similar automatic analysis in various domains.

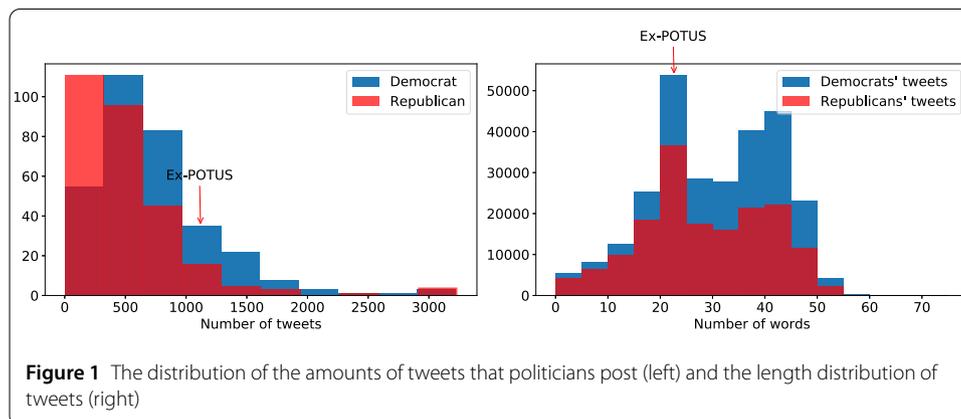
## 2 Data and methods

We collect data from major U.S. politicians on Twitter. Using the Twitter lists created by `cspan`,<sup>1</sup> we retrieve screen names of politicians including: U.S. Senators, House Representatives, state governors, and former President Trump. These Twitter accounts may be managed by the politicians or their staff, but in either case, they convey the messages from these politicians and are integral parts of their public images. We collect tweets from these accounts monthly starting in April 2020. In this study, we use tweets timestamped between February 1, 2020—one week after Wuhan’s lockdown started—to July 22, 2020. We use the full texts of tweets and only keep the English tweets.

The number of politicians’ tweets from each group is summarized in Table 1. We found that the Democratic politicians tend to post more compared to their Republican peers. Figure 1 shows the distribution of politicians’ posting frequencies and the length distribution of the tweets. We found a highly skewed distribution, where a few politicians tweet often and most only tweet occasionally. The majority of tweets have between 20–50 words for both groups.

**Table 1** The number of tweets posted by each group of politicians and the average number of tweets posted per person

	Number of tweets	Average number of tweets per politician
Senate (Republican)	34,329	635.7
Senate (Democratic)	38,539	<b>820.0</b>
House (Republican)	108,095	420.6
House (Democratic)	205,746	<b>635.2</b>
Governor (Republican)	23,397	899.9
Governor (Democratic)	30,398	<b>1085.6</b>
Former President Trump	1196	1196
Total (Republican)	167,017	494.1
Total (Democratic)	274,683	<b>704.3</b>



<sup>1</sup><https://twitter.com/cspan>

**Table 2** The number of COVID-19 related tweets and non-related tweets for each party

	Number of COVID-19 related tweets	Number of non-related tweets
Republican	37,854	127,275
Democratic	61,944	212,050

## 2.1 Filtering COVID-19 related tweets

Because we are most interested in the COVID-19 related political discourse, we identify COVID-19 related tweets by checking if “COVID” or “coronavirus” is present in a tweet (case insensitive). This may omit some tweets that are about the pandemic but do not mention the name, but it ensures that all tweets we consider are related to COVID-19. The number of COVID-19 related versus non-related tweets are show in Table 2.

## 2.2 Identifying over-represented terms

For an overall understanding of the topics and key issues that set up the contexts of each party’s narrative framework, we identify the over-represented words in their tweets. We use the log-odds ratios with informative Dirichlet priors [41] by computing the log-odds ratio of each word  $w$  in two corpora  $i$  and  $j$ , with a background corpus  $bg$  as prior. This is formally expressed as:

$$s_w = \log \frac{f_i + f_{bg}}{n_i + n_{bg} - f_i + f_{bg}} - \log \frac{f_j + f_{bg}}{n_j + n_{bg} - f_j + f_{bg}}, \quad (1)$$

where  $f_i$  is the frequency of the word in the target corpus; for example, words in the COVID-19 related Democratic tweets.  $f_{bg}$  is the frequency of the word in the background corpus. In this case, it is the combination of the Democratic and Republican tweets that are not related to COVID-19.  $n_i$  is the size of the target corpus, and  $n_{bg}$  is the size of the background corpus.  $f_j$  is the frequency of the word in the *other* corpus, in this case, the COVID-19 related Republican tweets; and  $n_j$  is the size of this corpus.

Furthermore, we compute the  $z$ -scores of the log odds ratio as:

$$z_w = \frac{s_w}{\sqrt{\frac{1}{f_i + f_{bg}} + \frac{1}{f_j + f_{bg}}}}, \quad (2)$$

where the denominator serves as an estimate of the variance of the log-odds ratio.

We choose the top 40 words with highest  $z$ -scores from each party’s COVID-related tweets as the most over-represented words. We exclude the politicians’ names and Twitter handles as they tend to be over-represented in each party’s tweets. To better explore these words and the topics they represent, we obtain their contextual embeddings using word embedding models. While many word embedding models are available, we choose the GloVe [50] embeddings as it is considered one of the most effective word embedding models [44] and is widely used. We use the pre-trained GloVe model with 6 billion tokens and a dimensionality of 300.

As many of the topic words are specific to the COVID-19 crisis, we train a new GloVe model on our tweet corpus for 500 epochs<sup>2</sup> to obtain embeddings for words not in the pre-trained GloVe model. Furthermore, for a consistent representation for terms related

<sup>2</sup>Training for less epochs result in less distinct clustering of the embeddings, but does not change the overall result.

to “COVID”, we compile a list of all tokens including “COVID” or “coronavirus” and replace them with “COVID” in the corpora. After removing emojis and words without embeddings, we show the top 35 words for each party.

To explore the topic words visually, we use the Uniform Manifold Approximation and Projection (UMAP), an effective [59] and efficient [3] dimensionality reduction method, to reduce the dimensionality of the GloVe embeddings. This method works by finding low-dimensional projections of the data that preserves their topological structures in high-dimensional space as much as possible [39]. We use the Python package `umap`. We plot the word embeddings with the dimensionality reduced to 2. With this visual aid, we identify and manually label six clusters for the Democratic tweets and three for the Republican tweets (see Sect. 3).

### 2.3 Microframe analysis

Most of the traditional framing analysis methods rely on “close reading” and manual examination of linguistic material, and are therefore challenging to apply to our dataset. Here, we employ the FrameAxis model [35], which allows an exploratory framing analysis through “microframes”. A microframe is operationalized as a pair of antonyms, such as “legal” and “illegal”, or “fast” and “slow”. In political science research, usage of antonyms has been successfully capturing political stances. For example, the Moral Foundations Theory uses five pairs of antonyms such as “Care/Harm” and “Fairness/Cheating” to serve as moral “axes” [22]. Here we use 1621 antonym pairs obtained from WordNet [40].

We then compute the *bias* and *intensity* of each microframe present in a document based on the vector representations of the microframes and other words in the text. We define the contribution of a word to a microframe as the cosine distance between the word vector  $w$  and the microframe’s vector  $f$  (see Kwak et al. [35] for details):

$$c_f^w = \frac{v_w \cdot v_f}{\|v_w\| \|v_f\|}. \quad (3)$$

The bias of a microframe is defined as the average contribution of all words in the document to the microframe. It captures the stance of a political argument; for example, a conservative document on the immigration issue may be biased towards *illegal* rather than *legal* in the *illegal versus legal* microframe. Formally, the bias is computed as

$$B_f^t = \frac{\sum_{w \in t} (n_w c_f^w)}{\sum_{w \in t} n_w}, \quad (4)$$

where  $t$  is a document,  $f$  is a microframe, and  $n_w$  is the number of occurrences of word  $w$  in  $t$ .

Meanwhile, the intensity of a microframe captures how strongly it is presented in a document, regardless of which “pole” the document is closer to. The intensity is computed using the second moment of the word contribution with a background corpus as baseline:

$$I_f^t = \frac{\sum_{w \in t} n_w (c_f^w - B_f^T)^2}{\sum_{w \in t} n_w}, \quad (5)$$

where  $B_f^T$  is the baseline microframe bias of the entire text corpus  $T$  on a microframe  $f$  for computing the second moment. As the squared term is included in the equation, the

words that are far from the baseline microframe bias—and close to either of the poles—contribute strongly to the microframe intensity.

Here we compute the bias and intensity for each COVID-19 related tweet, using a background of non-COVID-19 related tweets, for each microframe. We focus on the microframes with the largest difference in intensity between the two parties; for the Democratic party, we present the microframes where the intensity in Democratic tweets is higher than that in the Republican tweets, and vice versa. In addition to showing the microframes, we also show the top 3 tweets with the strongest intensity for each microframe.

## 2.4 Semantic role analysis

To identify important semantic roles, we use the Python package `Allennlp` [18] to perform semantic role labeling on our corpus. We focus on the verb, the Agents (*Arg0* in the `Allennlp` system), and the Patents (*Arg1*). To focus on the most common semantic roles, we only consider the Agents and Patents consisting of three or less tokens.

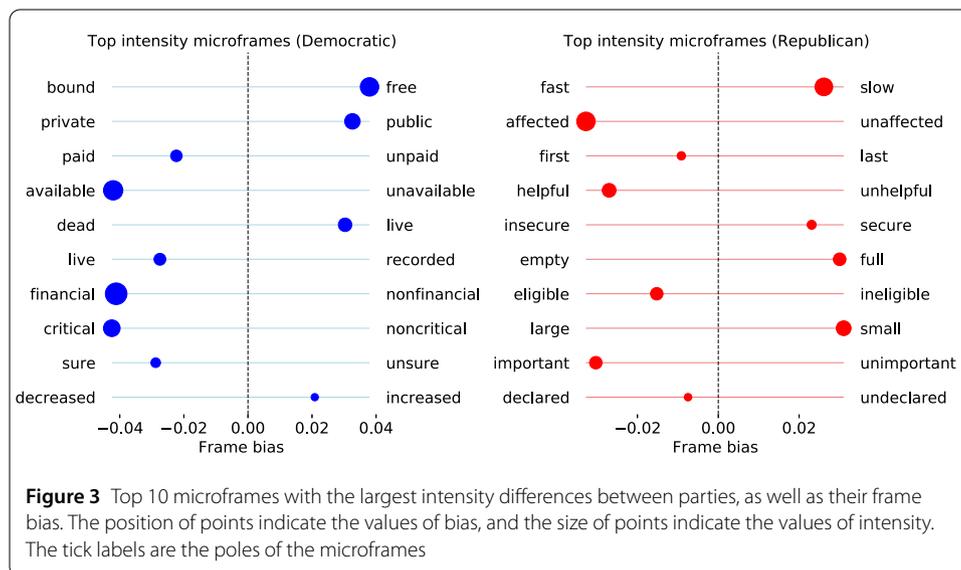
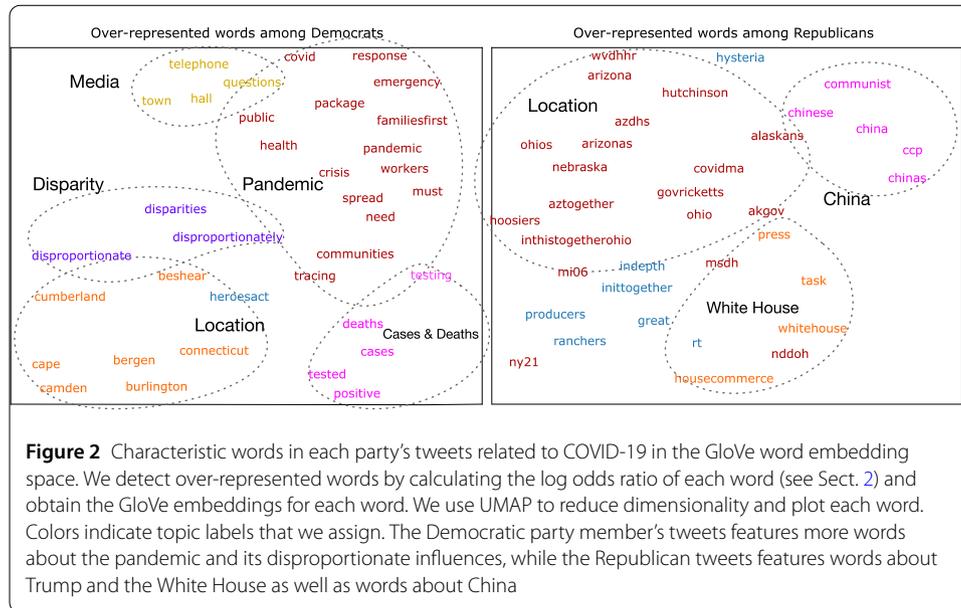
To obtain a list of semantic roles specifically related to the Democratic and the Republican party, we produce two lists of terms most similar to the words “Democrat”, “Democratic”, and “Republican” using the GloVe embedding model we described above. The terms most similar to “Democrat” and “Democratic” include “dems”, “housedemocrats”, “reddemocrats”, “democraticled”, “pelosi”, “speakerpelosi”, “nancy pelosi”, “chuck schumer”, “ralph northam”, “ayanna pressley”, “gwen moore”, and “senatedems”. The terms most similar to “Republican” include “gop”, “republicans”, “president”, “trump”, “donald trump”, “patrick mchenry”, “larry hogan”, “mitch mcconnell”, and “mcconnell” (case insensitive).

We identify important verbs by considering the top 100 most frequent verbs in each party’s tweets. We obtain the GloVe embeddings for each verb in the same manner as we describe above. We then use UMAP to reduce the dimensionality of the embeddings, and use the k-means clustering algorithm to group the verbs from each party into 15 clusters. This produces clusters of verbs that are semantically close to each other in daily usage, but also indicates some verb usage that are specific to parliamentary politics.

## 3 Results

First, we look at the most characteristic words found in each party’s tweets. We start with comparing each word’s dense rank [31] in the COVID-related Democratic and Republican tweets and the background corpus to find words over-represented in the COVID-related tweets. While these tweets unsurprisingly features many shared words between parties as shown in Table S1, we notice that the two parties have different focuses. We therefore use the log-odds ratio to identify the most representative words for each party in Fig. 2.

We find that the Democratic tweets have over-represented words related to media, such as “telephone”, “town hall”, and “facebook”, while a similar cluster for the Republican tweets appear to be related to the White House and its press conferences, such as “whitehouse” and “press”. Additionally, each party has words related to states, cities, and public figures from these places in the U.S. Meanwhile, the largest category in the Democratic tweets appears to be about the pandemic, such as “health”, “response”, “covid”, “emergency”, etc. Another cluster including “disparities” and “disproportionately” also suggest that they discuss issues about social and racial inequalities more. In the Republican case, few words such as “initttogether” appears to be directly related to the pandemic. Only the phrases and hashtags for certain region such as “covidma” and “inthisttogetherohio” are detected, indicating much less active narrative regarding the pandemic from the Republicans.



Lastly, both parties have some unique categories; the Democratic tweets has a cluster related to testing, specifically, including words such as “tested” and “positive”. The Republican tweets has a particular cluster about China and the Chinese Communist Party, reflecting the ex-president’s narrative against China.

The overrepresented words give us a sense of the topics and issues that set up the context for each party’s narrative frameworks. Our analysis of the framing used in each party’s tweets reveals the ways in which they shape their narratives. While the two parties share many common microframes about the pandemic, such as *new versus worn* and *endemic versus epidemic* (see Figure S2), here we focus on the microframes that one party uses significantly more than the other. Figure 3 shows the bias and intensity for each of the top ten microframes we identify (see Sect. 2). For example, the Democratic tweets features

the *public versus private* frame more intensely than the republican tweets, and at the same time they are more biased towards “public” rather than “private”.

Since it is hard to interpret the pole words without context, we also show the tweets with the highest intensity for each microframe in Table 3. Combining the pole words and tweet texts, we find that the Democratic frames strongly feature the economic relief during the pandemic, discussing topics such as financial relief, increased funds for support, free testing, etc., which are picked up by the microframe pole words including *free*, *financial*, *increased*, and *paid*. Additionally, the *public versus private* microframe identifies the emphasis on the public aspect of the pandemic and its response. They also frequently tweet about live events and town hall meetings, invoking the *live* frame. Taken together, we interpret that they emphasize the roles that the government should play regarding the pandemic, contrasting to the Republican framing that we discuss below.

Republican microframes include aid for *small* business, the *eligibility* for financial aid, and *securing* the economy and nation. “*Slowing* the spread” appears to be the top slogan used in Republican tweets, emphasizing the roles that individuals play, which contrasts the Democratic narrative. Additionally, the top tweets about *declaring* national emergency, *important* information, and *full* statements also suggests that the Republicans tend to use Twitter as a channel for formal announcements.

Finally, we examine the characters in each party’s narrative frameworks—people who need healthcare, travelers, voters, etc—and their relationships. For insights into how these characters are represented in the politicians’ tweets, we explore the semantic roles in these tweets, in particular, the Agents and Patients. We explore the most frequent Agents and Patients in both parties’ tweets in Figure S1. We find many common semantic roles as personal pronouns, but also notice some unique semantic roles, such as “the resources” and “lives” in Democratic tweets, and “COVID” and “relief” in Republican ones. Furthermore, the Republican tweets often mention the Agent “Democrats”, and the Democratic tweets often use “Trump” and “the president”.

For a more detailed analysis of the semantic roles, we consider the combinations of an Agent, a verb, and a Patient in each party’s tweets. We use the frequency for each combination to identify the most characteristic combinations. We found 321,913 unique combinations in the Democratic tweets and 82,821 unique combinations in Republican tweets. Table 4 shows the top combinations whose frequency in Democratic tweets is higher than in Republican tweets, and vice versa.

We find that most of the top combinations from Democratic tweets convey a message of “they” need support and “we” do everything we can to provide the resources, save lives, etc, further confirming the emphasis on the public response to the pandemic that we found in our framing analysis. Meanwhile, the combinations from Republicans are more diverse, featuring combating COVID, holding press conference, and aiding small businesses. Additionally, one combination discusses the threat of socialism.

From Figure S1, we also notice that the Agents often contains personal pronouns such as “I”, “we”, “they”, and both parties frequently discuss the opposite party, such as the Agent “Trump” from Democratic tweets, and “Democrats” from Republican tweets, evoking a membership categorization process. We therefore focus on the personal pronouns as Agents that we group into two categories—*us*, including the personal pronouns “I”, “we”, “us”, “our”, and “ours”, and *them*, including the words “they”, “their”, and “them”. Addition-

**Table 3** Three top tweets from each microframe with the largest difference in intensity between two parties. URLs, emojis, and some special characters are omitted

Democratic microframe	Republican microframe
<u>bound-free</u> "Free COVID testing is available near you."	<u>fast-slow</u> "Do your part to slow the spread of the Wuhan COVID:"
"Today's free COVID testing sites"	"We all need to do our part to slow the spread of the COVID. here's what you can do to help:"
"Testing, testing, testing. the bill makes sure that COVID testing is free for all Americans."	"rt @housegop: are you doing your part to slow the spread of the COVID?"
<u>decreased-increased</u> "Check to see if you qualify for paid sick leave because of the COVID here"	<u>declared-undeclared</u> "My statement after president @realdonaldtrump declared a #nationalemergency to respond to COVID."
"Stand with @pattymurray and @sengillibrand and support the paid leave act to provide additional support to workers & businesses for paid family and sick leave during the COVID outbreak."	"The first public health emergency was declared on March 6 and allows the state to increase coordination across all levels of government in the state's response to COVID:"
"Today, the house will vote on our next COVID response legislation to provide Americans wpaid family and medical leave, increased federal medicaid funds to support our state public health partners, free testing, & emergency sick leave for those impacted by the virus"	"President @realdonaldtrump has declared today as national day of prayer. Please join me in praying for our country as we continue to respond to the COVID pandemic."
<u>sure-unsure</u> "#MD02 constituents, unsure where to turn for local COVID resources? check out the below graphic for the hotline for your county."	<u>important-unimportant</u> "Important information for you and your family about the COVID"
"The least the president can do is make sure they have the equipment they need. COVID 3/3"	"Important information from @cdcgov regarding COVID"
"The response to COVID needs to help all Americans. i'm working with my colleagues to make sure that it does."	"Important COVID update from the @deptofdefense in the thread below."
<u>critical-noncritical</u> "rt @frankpallone: @WHO is critical in the fight against the COVID pandemic. Trump must work with the world's premier public health..."	<u>large-small</u> "If you own or work for a small business affected by the COVID pandemic, visit my website for information on support for small businesses"
"rt @uazmedphx: to address the critical needs of the Navajo nation during the COVID outbreak, #uazmedphx, @reppregstanton, as well as?"	"Visit learn about the EPCC's grant program for small businesses impacted by the COVID find more helpful EPCC small business resources"
"It is critical that we ensure those who have access to any COVID vaccine are not the privileged few, but the many who actually need it most."	"Welcomed news for Georgia small business owners. @sbagov emergency loans are now available to impacted businesses in all 159 counties. COVID"
<u>financial-nonfinancial</u> "Thank you, @abigaildisney, for looking out for the most vulnerable affected by the financial repercussions of COVID."	<u>eligible-ineligible</u> "Alabamians laid off or unpaid due to COVID are eligible for unemployment compensation"
"rt @repmalinowski: "the COVID will prey not just on the health of Americans but their financial wellbeing. In its next bill responding..."	"rt @oronline: if you work in Pennsylvania and the novel COVID has affected your job, you may be eligible for benefits."
"May 1 is quickly approaching, and I know that many marylanders are experiencing severe financial hardship because of the COVID. In this thread you'll find information about financial assistance available in MD."	"Small businesses: you may be eligible for up to \$2 million in @sbagov low-interest loans if your business has been affected by the COVID. These loans can help fill your working capital needs. Non-profits may also be eligible. Apply online here:"

**Table 3** (Continued)

Democratic microframe	Republican microframe
<p><u>live–recorded</u></p> <p>“Tune in now: I’m hosting a Facebook live town hall with @repbillfoster and @repcasten. We will be answering your questions on COVID. Watch live here:”</p> <p>“tune in now for my Facebook live COVID town hall with @stevelockhartmd of @sutterhealth:”</p> <p>“I am #live now on Facebook addressing your questions and concerns about the COVID. Tune in here:”</p>	<p><u>empty–full</u></p> <p>“Read here: my full statement in support of the COVID relief legislation the House just passed.”</p> <p>“My full statement on presumptive COVID cases in South Dakota”</p> <p>“See my full statement on president @realdonaldtrump’s new actions to fight COVID here”</p>
<p><u>dead–live</u></p> <p>“Tune in now: I’m hosting a Facebook live town hall with @repbillfoster and @repcasten. We will be answering your questions on COVID. Watch live here:”</p> <p>“I am #live now on Facebook addressing your questions and concerns about the COVID. tune in here:”</p> <p>“As of 2pm today 1,700 people in my state new jersey are tragically dead from COVID and 16,642 Americans are dead across the country.”</p>	<p><u>insecure–secure</u></p> <p>“rt @waysandmeansgop: in the phase three package to secure our economy as we fight against COVID, @ustreasury secretary @stevenmnuchi?”</p> <p>“I also thank our brave frontline @tsa officers for the risks they face on our behalf, continuing to keep our nation safe &amp; secure in the COVID pandemic.”</p> <p>“rt @waysandmeansgop: Dems voted against the phase three package to secure our economy as we fight against COVID. This package include?”</p>
<p><u>available–unavailable</u></p> <p>“More information is available from @cdcgov here: COVIDupdates COVIDUS”</p> <p>“Free COVID testing is available near you.”</p> <p>“rt @sfpelosi: 77,000 Americans killed by COVID unavailable for comment.”</p>	<p><u>helpful–unhelpful</u></p> <p>“This is a helpful resource for hoosiers to stay updated on COVID”</p> <p>“Here’s some helpful information on COVID for pregnant women and parents from the @cdcgov. You can find these and other resources on my website at”</p> <p>“Continue to follow @cdcgov for the latest updates on the COVID and helpful information. #MI06”</p>
<p><u>paid–unpaid</u></p> <p>“Check to see if you qualify for paid sick leave because of the COVID here”</p> <p>“rt @facttank: new: as COVID spreads, which U.S. workers have paid sick leave? and which don’t?”</p> <p>“I stand with @pattymurray and @sengillibrand and support the paid leave act to provide additional support to workers &amp; businesses for paid family and sick leave during the COVID outbreak.”</p>	<p><u>first–last</u></p> <p>“Love this. @starbucks is fueling our first responders on the frontlines of the COVID crisis! #initttogether”</p> <p>“rt @chadsabadie: @repabraham: the first responders, you bring calm to chaos COVID”</p> <p>“rt @woodtv: @rephuizenga pitches COVID aid bill for doctors, nurses and other first responders:”</p>
<p><u>private–public</u></p> <p>“rt @indivisibleteam: medicines, like the COVID vaccine, that are developed with public money should benefit public health, not create?”</p> <p>“@unitedwaydenver @cohealth coloradans can call the cohhelp line for the latest public health information on the COVID at 1-877-462-2911.”</p> <p>“rt @bryan_pietsch: healthcare workers battling the COVID would have their public and private student loans forgiven under a new bill?”</p>	<p><u>affected–unaffected</u></p> <p>“If you own or work for a small business affected by the COVID pandemic, visit my website for information on support for small businesses”</p> <p>“If you own a small business and your operations are being affected by COVID you may be able to get assistance from @sbagov. More info here:”</p> <p>“Appeared on @foxbusiness to discuss congressional action being taken to help Americans affected by COVID”</p>

**Table 4** Top Agent, verb, and Patient combinations in Democratic and Republican tweets extracted by semantic role labeling with largest differences in frequency. The left column shows the combinations where the frequencies in Democratic tweets are larger than the frequencies in Republican tweets, and vice versa. Most combinations in Democratic tweets focus on resources and support, while combinations in Republican tweets discuss combating COVID, news updates, support for small businesses, and the threat of socialism

Top Democratic combinations	Top Republican combinations
they, need, the resources	we, combat, covid
we, can, everything	I, holding, a news conference
we, do, more	covid, impacted, small businesses
they, need, the support	we, fight, covid
we, do, everything we can	governor hutchinson, provides, update
we, save, lives	we, moving, tax day
I, joined, my colleagues	I, provide, a covid update
we, do, what	I, holding, a press conference
those who, need, it	socialism, destroys, nations
we, recommit, ourselves	covid, affected, those

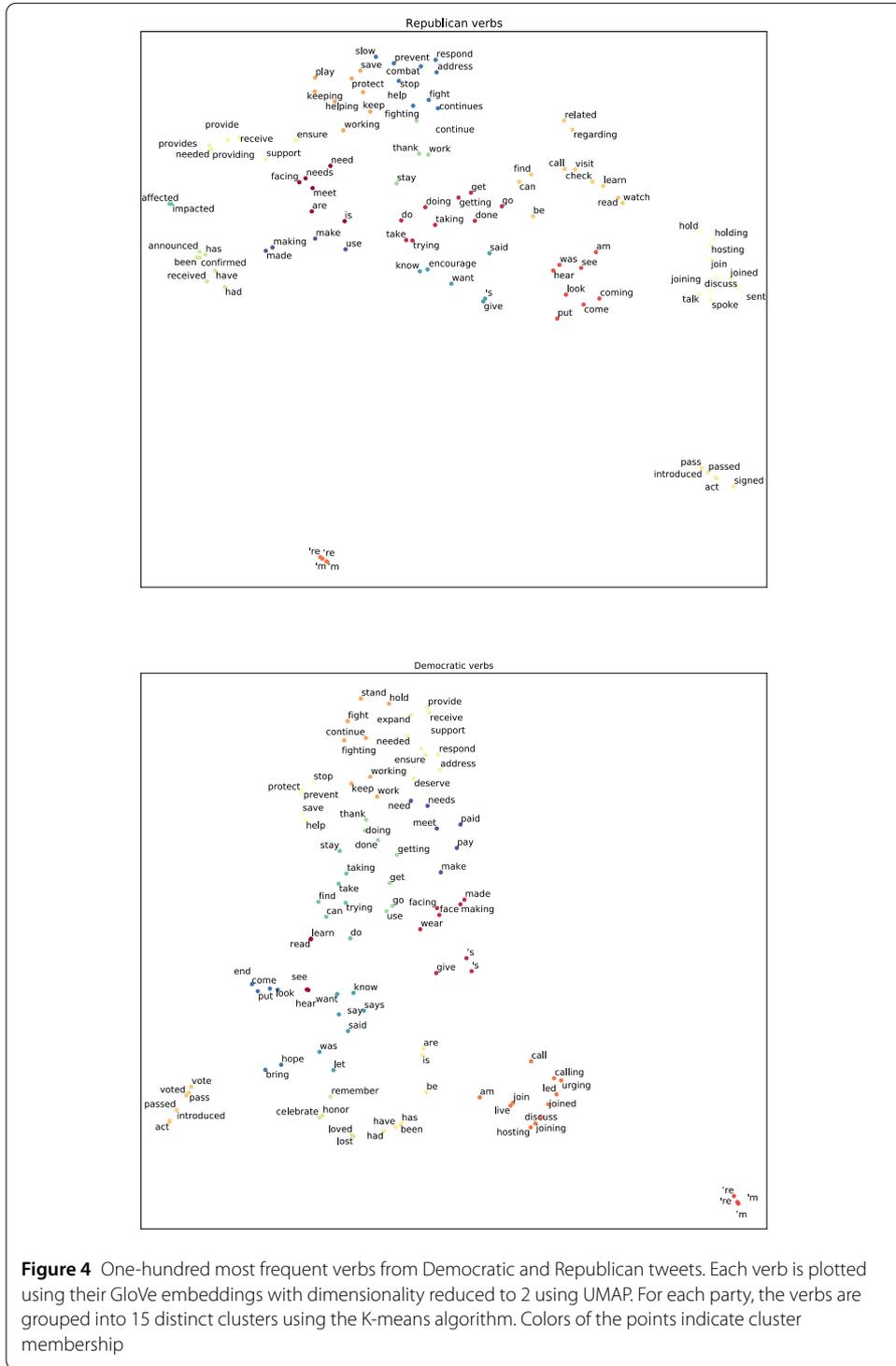
ally, we compile two lists of words associated with “Democrats” for Republicans, and vice versa (see Sect. 2).

We choose specific verbs for a more focused investigation. To leverage the semantic similarities between verbs, we consider the verb clusters that we create from the GloVe embeddings of verbs (see Sect. 2 for details). These clusters are shown in Fig. 4. Based on the proximity between verbs and examination of their Patients, we choose three sets of verbs that are most relevant to the pandemic, as well as having a number of diverse semantic roles as their Patients. We then consider the Patients with highest frequency for each set of verbs.

We begin by examining the Patients for the verbs “help”, “save”, and “protect” in Fig. 5. For both of the “us” and “them” categories, we find a strong shared theme about curbing the pandemic, such as saving lives, helping Americans and public health. Despite some party-specific Patients such as “#DACA” and “oil companies”, these semantic roles indicate an overlap in both parties’ tweets when it comes to protecting American people (although the way they frame help can be different as we discuss above).

We then move to the set of verbs “stop”, “slow”, and “prevent”. While both parties share a common theme in “stop the spread”, we observe many inter-partisan exchanges for both categories. For example, the Democrats discuss stop “mass employment” and “gun violence”, and the Republicans discuss stop “terrorism” as part of their own agendas. In the “them” category, the Democrats accuse the Republicans of stopping Fauci and “doing stock buybacks”, and the Republicans calls for the other party to stop “attacking president Trump” and “the deceptive mailers”. Compared to the previous set, this set of verbs has much less common Patients between two parties.

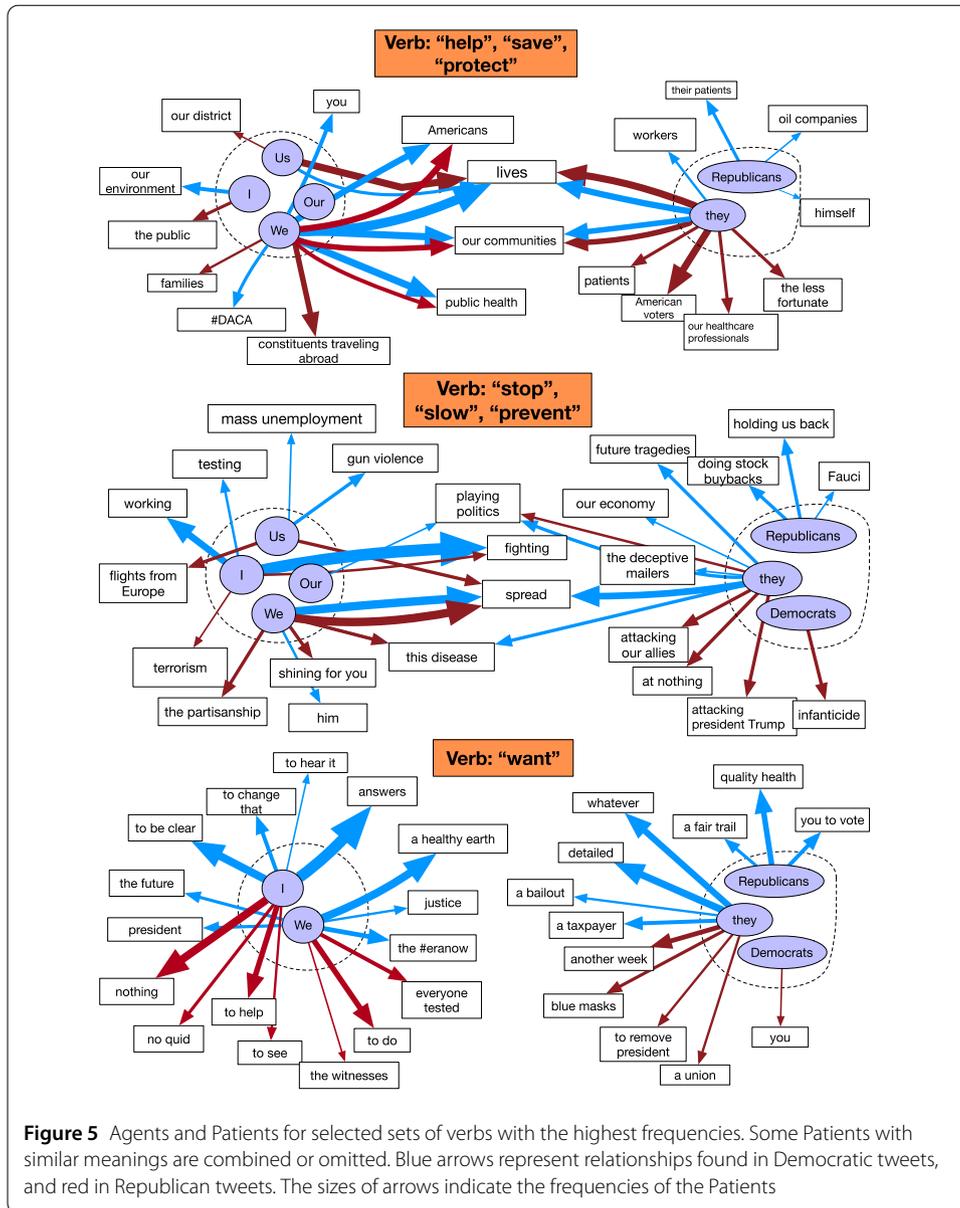
Finally, we check the verb “want” and find that the Patients are rather distinctive for both categories. In the “us” category, the Democrats emphasizes “answers”, “justice”, “a healthy earth”, and calling for the Equal Rights Amendment. Meanwhile, the Republicans do not have such strong callings, potentially due to the ruling/opposition party dynamics. In the “them” category, we see strong partisan messages about the opposite party, such as the Republican tweets discussing the Democrats’ “blue masks” and “to remove president”. This verb does not have any shared Patients, hinting at the different agendas from each party.



**Figure 4** One-hundred most frequent verbs from Democratic and Republican tweets. Each verb is plotted using their GloVe embeddings with dimensionality reduced to 2 using UMAP. For each party, the verbs are grouped into 15 distinct clusters using the K-means algorithm. Colors of the points indicate cluster membership

#### 4 Discussion

In this work, we characterize the political narrative frameworks about the COVID-19 crisis constructed by two major U.S. parties, demonstrating that a suite of relatively simple natural language processing methods can be applied to a large dataset to produce useful insights into the diverging narrative frameworks. We examine each narrative framework from three aspects: context, framing, and characters and relationships. We show



that the Democratic narrative framework contains more discussion about the pandemic overall, whereas the Republican one includes more mentioning of other political entities. In terms of framing, the Democratic narrative focuses on the financial relief and public health service during the COVID-19 crisis, whereas the Republican narrative emphasizes small business and the role of individuals. When we consider the semantic agents, these different focuses are further exposed, and we also found that while both parties find a common ground in battling the pandemic, they also have distinct agendas and political goals, and use their narratives to criticize the other party.

Our work demonstrates that computational methods can automatically extract strong signatures of political narratives that fit the key theories of political science, providing a useful "recipe" for computational narrative analysis. In addition, we also provide empirical analysis about diverging narrative frameworks in U.S. politics during the pandemic. Our results confirming our intuition, commonsense, and social theories about American

politics is a strong evidence for the effectiveness of the tools that we employ. By using an integrated set of computational methods, we bridge the gap between sophisticated NLP methodologies and real-world social problems.

Our study has several key limitations. One limitation of our FrameAxis model is not being able to distinguish word senses; for example, it is not able to separate “live” as the antonym of “dead”, and “live” as the antonym of “recorded”. This may lead to confusion when both word senses are widely used in the corpora. tweets with very different topics may also be identified under the same microframe, such as in the case of *available versus unavailable*, where the availability of COVID testing and availability for comment are mixed together. Such limitations may be partially addressed by using contextualized word embeddings such as ELMO or BERT, and will be an interesting future work.

Our semantic agent analysis use modern SRL tools to automatically identify semantic roles, but the interpretation of such roles remain a challenging task. For example, in Fig. 5, manual examination is required to select the Agents and verbs, as well as inferring their context. We are also limited to showing several small sets of verbs and their semantic roles. Additionally, when we examine the membership categorization, some semantic roles such as “they” may refer to a third group, instead of one of the parties, and these could not be identified by our model. More automatic ways of analyzing and exploring the SRL data can therefore be fruitful future research.

As we are not working on well-established tasks with systematic benchmarks, and because the tools are exploratory in nature (i.e., they serve as discovery tools and should be combined with human expertise in most cases), it is difficult to quantitatively evaluate them, although we have more rigorous evaluation tasks for our FrameAxis model [35]. We believe that designing systematic benchmarks for narrative analysis is a challenging, yet important future work. Nevertheless, even with these limitations, our set of methods provide an effective way to systematically characterize narrative frameworks that can be applied not only to the political communication domain, but to other domains as well.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1140/epjds/s13688-021-00308-4>.

**Additional file 1.** Supplementary information (PDF 79 kB)

## Acknowledgements

We thank Sandra Kübler, Xiaozhong Liu, Minje Kim, Haewoon Kwak, Jisun An, Byungkyu Lee, Matthew Josefy, and the anonymous reviewers for their insightful comments.

## Funding

Not applicable.

## Abbreviations

**PDA**, Political Discourse Analysis; **SRL**, Semantic Role Labeling; **NLP**, Natural Language Processing; **UMAP**, Uniform Manifold Approximation and Projection; **GloVe**, Global Vectors for Word Representation.

## Availability of data and materials

Our dataset and code are available at <https://github.com/yzjing/covid19-politics>.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

EJ and YYA designed the study. EJ collected the data and performed the analysis. EJ and YYA wrote the paper. All authors read and approved the final manuscript.

### Author details

<sup>1</sup>Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, USA. <sup>2</sup>Present address: Sirius XM, 1221 Avenue of the Americas 37th Floor, New York, NY 10020, USA. <sup>3</sup>Indiana University Network Science Institute (IUNI), Bloomington, IN 47408, USA. <sup>4</sup>Connection Science, Massachusetts Institute of Technology, 02139 Cambridge, USA.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 March 2021 Accepted: 12 October 2021 Published online: 30 October 2021

### References

1. Baker CF, Fillmore CJ, Lowe JB (1998) The Berkeley framenet project. In: 36th annual meeting of the association for computational linguistics and 17th international conference on computational linguistics, volume 1, pp 86–90
2. Barberá P, Casas A, Nagler J, Egan PJ, Bonneau R, Jost JT, Tucker JA (2019) Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. *Am Polit Sci Rev* 113(4):883–901
3. Becht E, Dutertre C-A, Kwok IW, Ng LG, Ginhoux F, Newell EW (2018) Evaluation of UMAP as an alternative to t-SNE for single-cell data. *BioRxiv*, 298430
4. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W (2015) Science vs conspiracy: collective narratives in the age of misinformation. *PLoS ONE* 10(2):e0118093
5. Bruner JS (2009) *Actual minds, possible worlds*. Harvard University Press, Cambridge
6. Charteris-Black J (2004) Why "an angel rides in the whirlwind and directs the storm?": a corpus-based comparative study of metaphor in British and American political discourse. In: *Advances in corpus linguistics*. Brill, Leiden, pp 133–150
7. Charteris-Black J (2018) *Analysing political speeches*. Macmillan International Higher Education
8. Cohn N, Paczynski M (2013) Prediction, events, and the advantage of agents: the processing of semantic roles in visual narrative. *Cogn Psychol* 67(3):73–97
9. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (almost) from scratch. *J Mach Learn Res* 12:2493–2537
10. Dunmire PL (2012) Political discourse analysis: exploring the language of politics and the politics of language. *Lang Linguist Compass* 6(11):735–751
11. Enli G (2017) Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. *Eur J Commun* 32(1):50–61
12. Enli GS, Skogerbø E (2013) Personalized campaigns in party-centered politics. *Inf Commun Soc* 16(5):757–774
13. Entman RM (1993) Framing: toward clarification of a fractured paradigm. *J Commun* 43(4):51–58
14. Exner P, Nugues P (2011) Using semantic role labeling to extract events from Wikipedia. In: *Proceedings of the workshop on detection, representation, and exploitation of events in the semantic web (Derive 2011)*. Workshop in conjunction with the 10th international semantic web conference, pp 23–24
15. Fan C, Jiang Y, Yang Y, Zhang C, Mostafavi A (2020) Crowd or hubs: information diffusion patterns in online social networks in disasters. *Int J Disaster Risk Reduct* 46:101498
16. Fillmore CJ (1967) The case for case. In: *Proceedings of the Texas symposium, on language universals*, April 13–15. Holt, Rinehart & Winston, New York
17. Fisher WR (1987) *Human communication as narration: toward a philosophy of reason, value, and action*. University of South Carolina Press
18. Gardner M, Grus J, Neumann M, Tafjord O, Dasigi P, Liu NF, Peters M, Schmitz M, Zettlemoyer LS (2017) Allennlp: a deep semantic natural language processing platform. [arXiv:1803.07640](https://arxiv.org/abs/1803.07640)
19. Garretson G, Ådel A (2008) 8. Who's speaking?: evidentiality in US newspapers during the 2004 presidential campaign. In: *Corpora and discourse*. Benjamins, Amsterdam, pp 157–187
20. Green J, Edgerton J, Naftel D, Shoub K, Cranmer SJ (2020) Elusive consensus: polarization in elite communication on the COVID-19 pandemic. *Sci Adv* 6(28):eabc2717
21. Grimes DR (2020) Health disinformation & social media: the crucial role of information hygiene in mitigating conspiracy theory and infodemics. *EMBO Rep* 21(11):e51819
22. Haidt J, Graham J (2007) When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Soc Justice Res* 20(1):98–116
23. Haidt J, Graham J, Joseph C (2009) Above and below left–right: ideological narratives and moral foundations. *Psychol Inq* 20(2–3):110–119
24. Haidt J, Joseph C et al (2007) The moral mind: how five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In: *The innate mind*, vol. 3, pp 367–391
25. Hemphill L, Culotta A, Heston M (2013) Framing in social media: how the US Congress uses Twitter hashtags to frame political issues. Available at SSRN 2317335
26. Hong S (2013) Who benefits from Twitter? Social media and political competition in the US House of Representatives. *Gov Inf Q* 30(4):464–472
27. Hung S-H, Lin C-H, Hong J-S (2010) Web mining for event-based commonsense knowledge using lexico-syntactic pattern matching and semantic role labeling. *Expert Syst Appl* 37(1):341–347
28. Johnson K, Goldwasser D (2016) All I know about politics is what I read in Twitter: weakly supervised models for extracting politicians' stances from Twitter. In: *Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers*, pp 2966–2977

29. Johnson K, Goldwasser D (2018) Classification of moral foundations in microblog political discourse. In: Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: long papers), pp 720–730
30. Jones MD, McBeth MK (2010) A narrative policy framework: clear enough to be wrong? *Policy Stud J* 38(2):329–353
31. Kessler JS (2017) Scattertext: a browser-based tool for visualizing how corpora differ
32. Kingsbury P, Palmer M (2002) From TreeBank to PropBank. In: LREC, pp 1989–1993. Citeseer
33. Kouloumpis E, Wilson T, Moore J (2011) Twitter sentiment analysis: the good the bad and the omg! In: Fifth international AAAI conference on weblogs and social media
34. Kubin E, Puryear C, Schein C, Gray K (2021) Personal experiences bridge moral and political divides better than facts. *Proc Natl Acad Sci* 118(6):e2008389118
35. Kwak H, An J, Jing E, Ahn Y-Y (2021) FrameAxis: characterizing microframe bias and intensity with word embedding. *PeerJ Comput Sci* 7:e644
36. Lee E-J, Shin SY (2012) Are they talking to me? Cognitive and affective effects of interactivity in politicians' Twitter communication. *Cyberpsychol Behav Soc Netw* 15(10):515–520
37. Llorens H, Saquete E, Navarro-Colorado B (2013) Applying semantic knowledge to the automatic processing of temporal expressions and events in natural language. *Inf Process Manag* 49(1):179–197
38. Madhani A (2020) Trump turns virus conversation into 'US vs. THEM' debate. <https://apnews.com/article/fe8d83b196f703520495ab7a92ba4d4cc>
39. McInnes L, Healy J, Melville J (2018) UMAP: uniform manifold approximation and projection for dimension reduction. arXiv preprint. [arXiv:1802.03426](https://arxiv.org/abs/1802.03426)
40. Miller GA (1995) WordNet: a lexical database for English. *Commun ACM* 38(11):39–41
41. Monroe BL, Colaresi MP, Quinn KM (2008) Fightin' words: lexical feature selection and evaluation for identifying the content of political conflict. *Polit Anal* 16(4):372–403
42. Moon D (2012) Who am I and who are we? Conflicting narratives of collective selfhood in stigmatized groups. *Am J Sociol* 117(5):1336–1379
43. Moretti F (2000) Conjectures on world literature. *New Left Rev* 1:54–68
44. Naili M, Chaibi AH, Ghezala HHB (2017) Comparative study of word embedding methods in topic segmentation. *Proc Comput Sci* 112:340–349
45. Ott BL (2017) The age of Twitter: Donald J. Trump and the politics of debasement. *Crit Stud Media Commun* 34(1):59–68
46. Pan Z, Kosicki GM (1993) Framing analysis: an approach to news discourse. *Polit Commun* 10(1):55–75
47. Park HS, Liu X, Vedlitz A (2014) Analyzing climate change debates in the US Congress: party control and mobilizing networks. *Risk Hazards Crisis Public Policy* 5(3):239–258
48. Parmelee JH, Bichard SL (2011) Politics and the Twitter revolution: how tweets influence the relationship between political leaders and the public. Lexington Books
49. Patterson M, Monroe KR (1998) Narrative in political science. *Annu Rev Pol Sci* 1(1):315–331
50. Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543
51. Sacks H (1992) Lectures on conversation (ed Jefferson G). Blackwell, Oxford
52. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrocioni W (2018) Polarization of the vaccination debate on Facebook. *Vaccine* 36(25):3606–3612
53. Shurafa C, Darwish K, Zaghrouani W (2020) Political framing: US COVID19 blame game. In: International conference on social informatics. Springer, Berlin, pp 333–351.
54. Smith LD (1989) A narrative analysis of the party platforms: the democrats and republicans of 1984. *Commun Q* 37(2):91–99
55. Tangherlini TR, Shahsavari S, Shahbazi B, Ebrahimzadeh E, Roychowdhury V (2020) An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: bridgegate, pizzagate and storytelling on the web. *PLoS ONE* 15(6):e0233879
56. IHME COVID-19 forecasting team (2020) Modeling COVID-19 scenarios for the United States. *Nat Med* 27:94–105
57. Wang J (2016) New political and communication agenda for political discourse analysis: critical reflections on critical discourse analysis and political discourse analysis. *Int J Commun* 10:2766–2784
58. Wilkerson J, Casas A (2017) Large-scale computerized text analysis in political science: opportunities and challenges. *Annu Rev Pol Sci* 20:529–544
59. Yang Y, Sun H, Zhang Y, Zhang T, Gong J, Wei Y, Duan Y-G, Shu M, Yang Y, Wu D et al (2021) Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. *bioRxiv*

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---