Government Information Quarterly 41 (2024) 101937

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Feature engineering from the perspective of agenda setting for predicting the success of online petitions

Philip Tin Yun Lee [a,*], Alvin Ying Lu [b], Feiyu E [c], Michael Chau [d]

[a] Department of Economics and Finance, Faculty of Commerce, Hong Kong Shue Yan University, North Point, Hong Kong
[b] Department of Information Systems and Analytics, School of Computing, National University of Singapore, Singapore
[c] Robert H. Smith School of Business, University of Maryland, United States
[d] Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Keywords:
Online petitions
Collective action
Agenda setting
Text mining
Feature engineering

ABSTRACT

This study draws on the issue expansion model and symbolism, both of which are influential concepts in the literature of public policy and agenda setting, to generate textual features for developing a predictive model of online petition success. Using a real-life dataset of an online petition platform, we show that the proposed model performs well in several important evaluation metrics when compared with benchmark models. This study offers several contributions. First, we present how to translate these concepts into textual features of petitions that can be understood by computers to improve prediction of petition success. The predictive models developed and the patterns of online petitioning identified enhance our understanding of collective actions on online petition platforms. In addition, we demonstrate that we can develop a better predictive model by adopting both supervised and unsupervised approaches of model development together with datasets that are exogenous from online petition platforms. Further examination of the predictive models in future may enable us to define vague concepts in a systematic way. On practical implications, our proposed text-mining model enables policy makers to handle a large volume of social data in a relatively objective manner. This is conducive to civic participation in e-democracy. The model may help policy makers identify potentially popular issues and prevent issue expansion at an early stage to mitigate the possible incursion of social cost. Moreover, by developing a predictive model based on our approach, citizens can compare different petition texts to determine their chances of success and post texts that have a higher predicted rate of success.

1. Introduction

Online petition platforms have been important tools in recent years, since they facilitate the processes of individuals gathering support and attention from others for any proposed changes of existing policies. In addition to government-managed e-petition platforms, such as We the People, some privately owned platforms, including Avaaz and Change.org, also exist to allow users to post their petitions online for free. Previous studies identified that many of the petitions on these privately owned platforms also target government officials for their responses to proposed changes to various policy issues (Halpin, Vromen, Vaughan, & Raissi, 2018; Horstink, 2017). These online petition platforms take advantage of the Internet and enable individuals to launch petition campaigns to reach a large population on the web using relatively affordable resources. Thus, people no longer have to rely on middle-man institutions, such as political parties and journalists, to solicit support from others and convey their messages to the governments and policy makers. Margetts, Hale, and John (2019) suggested that the use of these platforms "extends the ladder of political participation at the lower end, introducing new low-cost acts that were not possible in an earlier era" (p. 199), and an individual can raise an issue or participate in a campaign or debate "without belonging to anything, or even coming into contact with a political organization" (ibid). If the online views can be appropriately captured and responded to by the governments and policy makers in a timely manner, citizens will be motivated by the official feedback and participate in these petition platforms more often. This is one promising way in which online petition platforms can enhance civic participation.

* Corresponding author.
E-mail addresses: tylee@hksyu.edu (P.T.Y. Lee), luying@comp.nus.edu.sg (A.Y. Lu), u3005658@connect.hku.hk, feiyue@umd.edu (F. E), mchau@business.hku.hk (M. Chau).

https://doi.org/10.1016/j.giq.2024.101937
Received 18 February 2023; Received in revised form 26 April 2024; Accepted 30 April 2024
Available online 13 May 2024

Previous literature considered agenda setting as the first stage of a typical policy making process (Jann & Wegrich, 2017; Ngai & Lee, 2016). At this stage, when an issue secures adequate public attention, it will likely catch the attention of policy decisions makers, and the decision makers may feel the pressure to put the issues on their formal agendas for later decision on whether and how to act on the issues (Cobb, Ross, & Ross, 1976; Ragas, Tran, & Martin, 2014). Discussion about the first stage of agenda setting in the past literature pertains to traditional mass media, which was the most common device of social information diffusion among people during the age without the Internet (Lodge & Hood, 2002).

The traditional mass media, however, features hierarchical and elite control (Meraz, 2009; Neuman, Guggenheim, Jang, & Bae, 2014). The well-established media institutions were considered as gatekeepers with regard to information distribution to the public (Shoemaker & Vos, 2009). Thus, the content published by traditional media was generally more regulated and professional. The institutional interests of these traditional media sources may also be largely involved in the agenda setting stage (Grossman, 2022; Sparrow, 2006). In contrast, posting online petitions requires just a few clicks. The low threshold for engaging in online petition platforms aims to facilitate participation and deliberation among citizens (Vicente & Novo, 2014). While the literature on agenda setting features traditional mass media, we believe that the literature should also provide insights and ideas in the context of online petitioning.

Decades ago, only the more privileged people had the opportunity to gain access to the Internet. Access to the Internet has now become more of a necessity to many people. The difference between the online population and the offline population has become even more blurred. Recent studies have also demonstrated the close ties between traditional media agendas and online, social media agendas (Gilardi, Gessler, Kubli, & Müller, 2022; Vargo, Guo, & Amazeen, 2018). Arguably, the online issue expansion has been increasingly similar to the paper-based issue expansion to which longstanding concepts of agenda setting are anchored.

Unfortunately, only a few scholars of collective action have adopted the perspective of agenda setting to interpret and understand online petitioning. Concepts and findings of agenda setting have not been seriously used for predictive model development of online petition success. In recent years, researchers have examined various types of collective action through online platforms. Some popular research topics include crowdfunding (e.g., Hong, Hu, & Burtch, 2018; Siering, Koch, & Deokar, 2016) and crowdsourcing (e.g., Mo, Sarkar, & Menon, 2018). Online petitioning has also been considered as one kind of collective action that is enabled by web platforms (Brunsting & Postmes, 2002; Greijdanus et al., 2020). A number of recent studies of online petitions have examined how text-mining technologies can be used to understand a large volume of petitions initiated by the general public, e.g., Dumas et al. (2015), Dumas (2022), Hagen (2018), Harrison et al. (2022) and Suh, Park, and Jeon (2010). The results of these studies indicate a promising potential for the use of text-mining technologies in the context of online petitioning.

However, scholars appear to be cognitively attached to seeing online petitioning as online collective action. Thus, existing studies have developed insights into the knowledge base of online collective action. Only a few scholars have looked into the specific context of online petitions from the perspective of agenda setting – online petitioning is a virtual way to call for people's signatures and gather their support so as to urge established institutes to take action. Some examples include Dumas et al. (2015) and Dumas (2022). Both studies drew on Baumgartner and Jones' (2010) concept of punctuated equilibrium in agenda setting and used market basket analysis and social network analysis to investigate the co-signing phenomenon of petitions and how groups of petition supporters are formed. However, these researchers adopted a case study approach and thus focused on specific incidents and policy areas, i.e., gun control for Dumas et al. (2015) and the legalization of marijuana for Dumas (2022).

The non-virtual, paper-based calls for support are not new in the field of political communication, and scholars in these fields have examined many aspects of these calls (Negrine & Papathanassopoulos, 2011). Specifically, longstanding concepts, such as the issue expansion model (Cobb & Elder, 1972) and symbolism (Birkland, 2017; Schattschneider, 1975; Stone, 2002), have been widely discussed and examined among political scientists and communication researchers. This study draws on the issue expansion model (Cobb & Elder, 1972) and symbolism (Birkland, 2017; Schattschneider, 1975; Stone, 1989, 2002), both of which are influential concepts in the literature of agenda setting and policy-making. Based on these longstanding concepts, we adopt a feature engineering approach and propose a set of linguistic and semantic features of online petition content for prediction model development. In recent years, prediction models have been increasingly recognized as a way to advance our understanding of human behavior and support theory development (Dhar, 2013). Models developed on large amounts of data enable us to make fewer assumptions of the models and to avoid random errors (Salganik, 2019). Also, the models can facilitate scholars to identify patterns for supporting induction in theory building (Shrestha, He, Puranam, & von Krogh, 2021).

In this study, using a real-life dataset of an online petition platform, we demonstrate how to translate previous observations and findings into textual features that can be understood by computers in the context of online petitions. Importing these features into classification models can improve the predictive results of online petition success. This enhances our pattern detection from online petition data and improves our understanding of public opinion formulation. Specifically, this study provides several contributions. In regard to theoretical implications, we differentiate online petitioning from other online collective action and identify its association with the previous literature on political communication. The specific context of online petitioning in agenda setting enables us to discover the long-standing concepts in political communication, i.e., the issue expansion model (Cobb & Elder, 1972) and the use of symbols (Birkland, 2017; Schattschneider, 1975; Stone, 1989, 2002). Drawing on the literature of these concepts, we develop predictive models of online petition success. The results advance our understanding of collective actions on online petition platforms.

In addition, some recent studies, e.g., Koenig and McLaughlin (2018) and Porten-Cheé, Kunst, Vromen, and Vaughan (2021), have tended to suggest that predictive results are valid only in selected policy categories. In this study, we demonstrate that we can develop a better predictive model by integrating both supervised and unsupervised approaches of model development. Further examination of the predictive models in future may enable us to define vague concepts in a systematic way.

Lastly, we demonstrate how we can import structured, regulated data of traditional media into models of social platforms of collective action. The inclusion of *New York Times* front-page story titles, which are exogenous from online petition platforms, follows the idea of big data and indicates the unexplored potential of including other exogenous data from sources of traditional media and online media for improved model development.

As for practical implications, our proposed text-mining model enables policy makers to handle a large volume of social data more objectively. Manual analysis of all online petitions is not feasible and largely involves human subjectivity. Previous studies have indicated that biases existed in officials monitoring and handling these online petitions (Feng, Wang, & Wang, 2023; Lu, Xu, & Wei, 2023). The officials can consider the prediction of the model as an evaluation metric for prioritizing the issues they need to address. They can then address the issues earlier and prevent public outcries. Moreover, the proposed model can be incorporated into data visualization and analytics tools for better policy-making in the future.

Furthermore, having a better predictive model of online petitions enables petition platform users to select the best from their drafts of petition texts. Users can evaluate the chance of petition success based solely on their texts before posting their petitions. With the predictive model, people can compare two or more petition texts on the same issue and choose to post the one with the higher likelihood of success. The ease of achieving petition success with the model will then enhance people's continuous use of the system and strengthen civic participation in the long term.

2. Literature review

2.1. Agenda-setting and online petitioning

Agenda setting, according to Birkland (2017), is "the process by which problems and alternative solutions gain or lose public and elite attention" (p. 63). In the context of policy- making, it is the first stage of the policy cycle, followed by policy formulation and decision making, implementation, and evaluation and termination (Jann & Wegrich, 2017). The public agenda is different from the agenda of political decision makers. The former agenda mostly determines what issues are to be put on the latter agenda. Those issues on the agenda of political decision makers then pass through the policy cycle and possibly lead to changes of relevant policies. Thus, the mechanism of agenda setting acts like a filter of policies that will be focused on by political decision makers in the future (Jann & Wegrich, 2017). Since public attention is limited, issues have to compete among themselves for places in the public agenda. Factors such as the public mood and events may affect the chances of gaining a "policy window" that enables issues to be placed on government agendas among other competing issues (Kingdon, 2010). Through online petition platforms, individuals can propose petitions that express their views on various societal issues. The petitions are aimed at gathering the wide attention of online users with the hope of ultimately leading to changes in the corresponding policies (Macintosh, 2004; Wright, 2015). Even if some petitions initially target the bad business practices of certain companies, the pressure will later be shifted to the government to implement new regulatory policies to rectify the wrong practices. As a result, policy changes are achieved. This sequence of policy changes has been considered to be similar across various democratic systems, although different countries have their own channels of policy development and decision-making institutes (True, Jones, & Baumgartner, 2019).

In the context of business operation, previous studies have found that public and media agendas influence management decision-making. It is suggested that companies may take proactive roles to conduct public relations activities to preserve their images on the news (Carroll & McCombs, 2003). Brown and Deegan (1998) selected firms from nine industries and reviewed their environmental disclosure over a 13-year span. They found that media agenda effectively draws public attention about specific firms' environmental performance, and the increased attention results in the firms following up to disclose more information on their environmental performance in their subsequent annual reports. Zyglidopoulos, Georgiadis, Carroll, and Siegel (2012) also found that media attention is associated with the strengthening of a firm's corporate social responsibility (CSR). In addition, Besiou, Hunter, & Van Wassenhove (2013) conducted three case studies of protests against multinational companies and suggested that online media successfully served as "web of watchdogs" that put pressure on managers to initiate negotiations with the stakeholders of the protests. Some recent studies shed light on social media agendas and showed a positive correlation between social media agendas and firms' behavior, such as decisions on CSR (Balasubramanian, Fang, & Yang, 2021; Saxton, Ren, & Guo, 2021).

2.2. Issue expansion model

Whether an issue in a petition can attract the public attention depends largely on how it is defined. According to Cobb and Elder (1972), issue expansion depends on certain defined characteristics of an issue. One such characteristic is the concreteness of the issue: "The more ambiguously an issue is defined, the greater the likelihood that it will reach an expanded public" (p. 112). Goals and objectives represent the two poles of concreteness-specificity continuum. Goals are vague terms and doctrines such as liberty and equality, whereas objectives are referred to as specific demands concerning the actions of political decision makers (Cobb & Elder, 1972). Ambiguous and abstract goals sometimes divert public attention away from whether proposed policies that address the issue can generate demonstrable benefits against costs. The spreading of an issue then becomes easier and requires no rational development (Edelman, 1974).

Another characteristic is temporal relevance: "(T)he more an issue is defined as having extended temporal relevance, the greater the chance that it will be exposed to a larger audience" (Cobb & Elder, 1972, p. 117). Actions for an issue can be described as the start of a long-lasting trend. Some people may not be affected by an issue, but they will be affected by the possible "spill-over" effect of the actions. Therefore, an issue that has strong temporal relevance can draw the attention of these types of people as they will see the issue as soon having a proximate impact on their lives (Zahariadis, 2016).

The third characteristic is linguistic complexity: "(T)he more non-technical an issue is defined to be, the greater the likelihood that it will be expanded to a larger public" (Cobb & Elder, 1972, p. 120). The use of technical language prohibits the general public from participating in the issue discussion, and therefore reduces the chance of issue expansion. Hartelius (2011) suggested that technical language reduces the persuasiveness of a message to the general public. Hansford and Coe (2019) demonstrated that linguistic complexity reduces public acceptance to court decisions. Boykoff (2011) also noted that expert terms led to a lack of public attention to climate change in developing countries.

The last characteristic is categorical precedence: "(T)he more an issue is defined as lacking a clear precedent, the greater the chance that it will be expanded to a larger population" (Cobb & Elder, 1972, p. 122). If an issue in a policy category has existed for long, the public may get the impression that the issue had been thoroughly examined but could not be resolved. On the other hand, if an issue appears to be unprecedented, the novelty of the issue can catch the attention of more people. This is also aligned with Kingdon's (2010) streams metaphor of agenda change. According to Kingdon (2010), a change of perception of an issue opens a window for policy change, implying a possible increase of public attention to the issue and public pressure on political decision makers for problem solving.

2.3. Symbolism

A symbol, in agenda-setting literature, is abstractly referred to as "anything that stands for something else. Its meaning depends on how people interpret it, use it, or respond to it" (Stone, 1989; Stone, 2002, p. 137). In political communication, symbols are different from textual signals in that symbols are not terms that have "narrowly circumscribed meanings" (Edelman, 1974, p. 299). Instead, symbols evoke beliefs and emotions generated through individual memories and perception (Edelman, 1974). Symbols can be used to elevate an issue to the public agenda by inducing the empathy of the media and the general public (Birkland, 2017; Schattschneider, 1975). For example, the words "me too" not only mean "we are the same," a narrowly circumscribed meaning, but also express the discontent about the prevalence of sexual violence against women (Lang, 2019). Cobb and Elder (1972) discussed several aspects of the use of symbols. They suggested that symbols with a long historical background are more likely to evoke the reactions and attention of the public. Stronger symbols have usually been used in a large number of issues. Cobb and Elder (1972) also argued that symbols can be used together for better issue expansion.

The occurrence of focusing events is one of the main triggers of public attention to a specific issue (Baumgartner & Jones, 2010; Birkland, 1998; Cobb & Elder, 1972; Kingdon, 2010). A focusing event is defined as “an event that is sudden; relatively uncommon; can be reasonably defined as harmful or revealing the possibility of potentially greater future harms; has harms that are concentrated in a particular geographical area or community of interest; and that is known to policy makers and the public simultaneously” (Birkland, 1998, p. 54). Focusing events can be symbols, although not necessarily all focusing events are (Birkland, 1997). For example, Vox, the far-right party in Spain, made good use of selective historical events as political symbols to gain attention and popularity (Rodríguez-Temiño & Almansa-Sánchez, 2021). Media reporting of these events is commonly in headlines, and journalists tend to incorporate symbols into issues. The symbols facilitate issue expansion (Birkland, 1998).

2.4. Predictive models of online petitions

Previous researchers have commonly examined information diffusion in the contexts of online collective action. Recent studies, such as Hong and Hoban (2022), King and Wang (2023), and Costello and Lee (2022), have developed predictive models to identify patterns of information diffusion with the use of data analytical and machine learning methods. Hong and Hoban (2022) adopted the hierarchical attention network proposed by Yang et al. (2016) to leverage the hierarchical structure of written works on a donation-based crowdfunding platform. They showed that sentences of more compelling creative appeals tend to be more concrete. Concrete creative appeals are more effective in persuading readers to undertake actions. King and Wang (2023), who investigated misinformation diffusion on Twitter, used both regression analyses and machine learning methods to demonstrate that the readability of a tweet is positively correlated with the chances of retweet. They measured readability by the lexical density of tweet texts. The results generated by regression analyses and machine learning methods are largely consistent, although the results generated by one of the machine learning methods showed some inconsistencies with other methods. Moreover, Costello and Lee (2022) examined entrepreneurial narratives on crowdfunding platforms. Using machine learning methods, they identified causal features that help predict funding success on crowdfunding platforms. The researchers found that reading ease and the use of descriptive language in the descriptive sections of the narratives respectively increase funding amounts of crowdfunding projects.

Relatively few studies have explored how predictions of online petition popularity can be improved. Hagen et al. (2016) examined how various linguistic as well as semantic factors influence the popularity of online petitions. They made three concluding remarks. First, extreme language inhibits the success of petitions. Following Craig and Blankenship’s (2011) approach, Hagen et al. (2016) found that petitions that mention the words “much more,” “extremely,” “very,” and “wonderful” are less popular among the online audience. Second, names in petitions are not appealing to the online population. Using the StanfordCoreNLP NER tagger (Finkel, Grenager, & Manning, 2005), Hagen et al. (2016) extracted named entities from the petitions. These named entities included persons, locations, and organizations. Hagen et al. (2016) showed that only names of persons are correlated with the popularity of the petitions. Third, petitions which mentioned well-known topics or important events are more popular. Hagen et al. (2016) adopted a semi-automatic approach with the use of LDA to produce a list of topics. They then qualitatively analyzed and identified topics that were significantly correlated with the popularity of petitions.

Another relevant, recent study is that of Chen, Deng, Kwak, Elnoshokaty, and Wu (2019). Associating the words in petitions with the relevant categories of General Inquirer (Stone, Dunphy, & Smith, 1966), Chen et al. (2019), based on the dual-process theory of persuasion (Petty & Briñol, 2015), found that petitions with positive emotions and enlightening information are more appealing to online users. They also showed that the online population is not interested in moral and cognitive reasoning. Both Hagen et al. (2016) and Chen et al. (2019) used logistic regression models, in contrast to those more advanced data analytical and machine learning models used in existing studies of online collective actions, to make predictions of online petition success.

3. Method

3.1. Feature engineering for prediction model development

In this study, we adopt the feature engineering approach to identify domain-specific textual features for better prediction of online petition success. Feature engineering refers to the process of making better representations of predictors (i.e., features) that models can utilize to improve prediction results (Kuhn & Johnson, 2019). In the context of online petitions, words in each petition serve as predictors. We derive domain-specific textual features from the aforementioned literature of agenda setting. These features are divided into two types, namely linguistic features and semantic features. Using a real-life dataset of an online petition platform, we compare models with benchmark features against models with both our proposed features and benchmark features. The proposed textual features are useful predictors if the latter models outperform the former models in various evaluation metrics. A model with better predictability can enrich our knowledge of the pattern of online collective actions in the context of online petitioning.

3.1.1. Benchmark features

Few studies have examined the textual characteristics that lead to higher popularity in the context of online petitioning. Hagen et al. (2016) found that linguistic *extremity* negatively affects popularity. This *extremity* linguistic feature was measured by whether a petition contained any of the following words: “much more,” “extremely,” “very,” or “wonderful.” Moreover, the study showed that the linguistic features of *repetition* of words and *internet activity* (i.e., whether a petition mentions words including “http,” “www,” “html,” or “Youtube”) have negative effects on popularity. The features *urgency* (i.e., whether a petition mentions words such as “immediately,” “immediate,” “urgent,” or their synonyms) and *sentiment* positively affect popularity. However, except for *extremity*, the effects of the remaining linguistic features disappeared when topic variables were incorporated into the predictive model. Hagen et al. (2016) also discovered that petitions with more names of *persons* are likely to be unpopular. The name entities were identified using Stanford CoreNLP NER tagger (Finkel et al., 2005).

Chen et al. (2019) developed a multi-appeal model composed of cognitive appeals, moral appeals, and emotional appeals to predict the popularity and success of online petitions. To measure the strength of these petitions’ appeals, they used categories in General Inquirer that were relevant to these appeals. General Inquirer (GI) is a linguistic and content analytical tool commonly used in similar studies (Chen et al., 2019). Specifically, they found four significant appeal factors, namely negative emotion (an emotional appeal factor), linguistic modality (a moral appeal factor), enlightenment (a cognitive appeal), and understatement (a cognitive appeal).

3.1.2. Linguistic features

A low level of linguistic concreteness leads to a higher chance of attracting public attention (Cobb & Elder, 1972; Edelman, 1974). The concreteness of a petition is represented by the frequency of abstract words appearing in a petition. Using a larger number of abstract words indicates a lower level of concreteness of an issue. The list of abstract words is derived from GI and comprises those words that are tagged abstract in GI.

A petition with stronger temporal relevance is more likely to attract people’s attention (Cobb & Elder, 1972). Temporal relevance is represented by the frequency of words related to “future” in a petition. The repository of words related to “future” was prepared using WordNet, which is a large lexical database of English words in which all the words are organized in synonym sets. Each set expresses a distinct concept, in which items share similar conceptual-semantic meaning and are lexically related (Miller, 1998). In addition, we used the online Oxford Dictionary[1] to complement the synonym sets.

Petitions with fewer technical words are more appealing to the general public, since they do not require exposure to the domains to obtain a clear understanding of their wording (Boykoff, 2011; Cobb & Elder, 1972; Hansford & Coe, 2019; Hartelius, 2011). Linguistic complexity is represented by the frequency of technical words in a petition. By technical words, we mean those words that are highly relevant to specific domain knowledge. Understanding technical words requires some previous exposure to the domains. The repository of technical words was prepared using the online Oxford Dictionary. Any technical words were tagged with a “domain” label in the online Oxford Dictionary. The “domain” labels indicated that readers are expected to have domain knowledge in order to fully understand the meanings of the words. We calculated the number of words with the domain tag in each petition. More words of this kind indicate a higher level of complexity of a petition.

3.1.3. Semantic features

Semantic features were identified through two approaches. First, we adopted an expert-assisted approach to address categorical precedence as suggested in Cobb and Elder (1972). A code book was developed to determine an issue’s policy category. Two political science faculty members were invited to formulate the code book with reference to code books of The Policy Agendas Project (2017). For each policy category, the code book includes a list of keywords. Appendix A shows the policy categories and examples of keywords in each category. We measured the political category of each petition by calculating the frequency of keywords that co-occurred in the petition description and the keyword list of each political category. As a result, each petition was mapped into a political dimension with 24 distinct topics. The zero value on one dimension indicates the irrelevance of the current petition regarding this political category. This enabled us to observe differences among all petitions in this political categorical knowledge base.

Second, an unsupervised approach was used to identify symbols in political communication literature (Edelman, 1974; Stone, 1989, 2002). Symbols were obtained by training an LDA model. LDA is a generative probabilistic model that expresses documents via a distribution of topics, and each topic is further represented via a probabilistic distribution of words (Blei, Ng, & Jordan, 2003). Not only did we extract topics using user-generated content from online petition websites, as suggested by Hagen et al. (2016), but we also used textual data written by professional journalists. We argue that symbols can be found more easily from textual data that is largely composed of symbols through topic modelling.

Specifically, we used descriptive contents from successful petitions as they are assumed to be composed of more political symbols that successfully gained public attention by inducing the empathy of the media and the general public (Birkland, 2017; Schattschneider, 1975). We also used the corpus from the *New York Times*’ front-page story titles between 1996 and 2006 (The Policy Agendas Project, 2017). The corpus written by professional journalists is believed to reflect the issues’ symbols more consistently and succinctly than user-generated online petitions. In particular, if there were any focusing events, the titles of the front-page stories would be highly likely to mention the events. These focusing events are commonly symbols (Baumgartner & Jones, 2010; Birkland, 1998; Cobb & Elder, 1972; Kingdon, 2010), and therefore words that named the focusing events constitute literal symbols.

To express each petition regarding its potential semantic information, we extracted the top 20 topics with the highest topic-term probability values after training the LDA model using the strategy mentioned above. The number of topics was determined by comparing the 15-topic words for each topic in topic models, including models with 5, 10, 15, 20, 25, and 30 topics, respectively. We did not examine models with more than 30 topics, given that these models would have too high a perplexity level, which is not desirable (Blei et al., 2003). The words of each topic in the model that contained 20 topics were considered as the most coherent among 3 coders, who are postgraduate students with thorough training of coding. Appendix B shows examples of words in each of the 20 topics. Our approach resembles the approach of Hagen et al. (2016) of determining the appropriate number of topics generated by LDA. However, we did not follow Hagen et al. (2016) entirely to delete topics from the whole set of topics generated by the unsupervised LDA. We were inclined to preserve the unsupervised nature of LDA to mitigate the involvement of human subjectivity. Using these 20 topics, we successfully mapped each document (petition) into a 20-dimensional space of potential semantic topics. We statistically expressed each petition in a probability distribution of topics. This approach allowed us to quantitatively capture the extents of different topics addressed in the petitions, and we imported such features into prediction models.

3.2. Experiment design

We evaluated the practical value of our proposed features using a dataset collected from Avaaz (Aragón et al., 2018), one of the most influential online petition websites. The website allows individuals to advocate their political views and to gather support from other users. The dataset consists of around fifty thousand English petitions across a six-year span. We tagged a petition as “successful” as long as the number of signatures reached the targets set by the petition authors. Otherwise, a petition was labeled as “failed.” Thus, the evaluation of proposed features generated binary classification results. Five thousand petitions were randomly selected from the successful and unsuccessful groups respectively to examine the performance of our proposed semantic features and the linguistic features. The benchmark features include Hagen et al.’s (2016) identified linguistic variables and named-entity variables as well as Chen et al.’s (2019) features of cognitive appeal, moral appeal, and emotional appeal. The final dataset contained 10,000 petitions with their success indicator, their benchmark features, and our proposed features (linguistic and semantic).

We first used 10-fold cross validation to show the effectiveness of the full feature set, which consists of linguistic features, semantic features, and benchmark features. We then compared the performance of the classification models trained by the full feature set with the performance of the models trained only by the benchmark features. Six classification models were used in the experiment: Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbor (KNN), Naïve Bayesian (NB), Decision Tree (DT), and Random Forest (RF). This set of classification models was also used in previous studies such as that of Lash and Zhao (2016). Fig. 1 shows a data-oriented illustration of our overall experimental set-up.

3.3. Results

Four commonly used performance metrics of classification, namely precision, recall, F1-score and AUC, were used to evaluate the performance of the predictive models. Table 1 shows the average scores (i.e., precision, recall, F1-score, and AUC) of the classification models achieved by importing the different sets of textual features into the models. Table 2 presents the detailed scores of each classification model with different sets of textual features.

The persistently outstanding performances of the full set of our proposed features against counterparts signify the usefulness of the proposed features in predicting petition success. Table 1 indicates that our proposed full set of textual features outperforms the benchmark features with respect to precision, recall, F1-score, as well as AUC on average. The four average scores of the full set are higher than or equal

[1] We accessed the Oxford Dictionary via its API (https://developer.oxforddictionaries.com/).

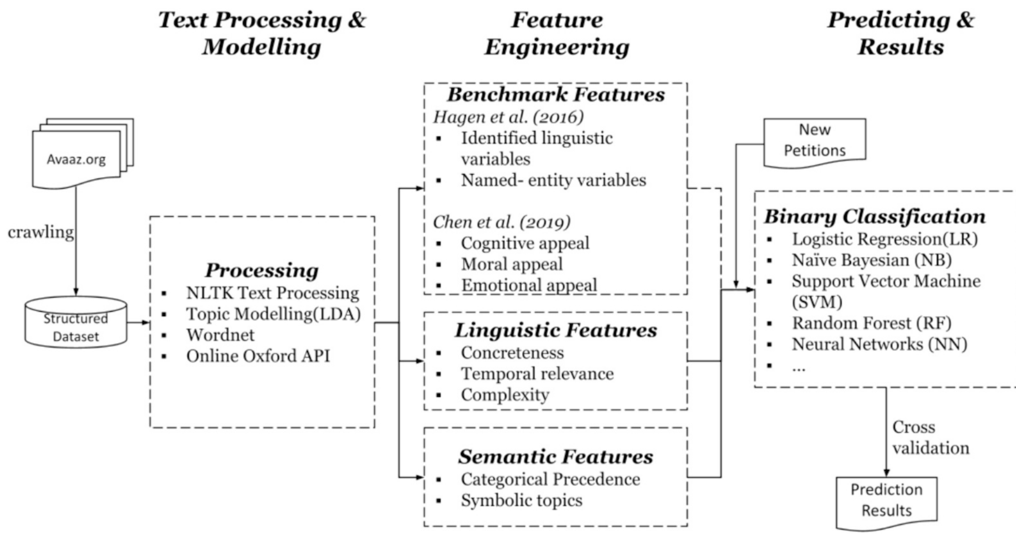

Fig. 1. Data-oriented illustration of the overall experimental set-up.

Table 1
The average performance scores of classification models achieved by importing the different sets of textual features into the models.

Performance Metrics	Features	The Average Scores of Classification Models
	BF	0.62
	BF & SF	0.65
	BF & LF	0.63
Precision	BF & SF & LF	**0.65**
	BF	0.73
	BF & SF	0.72
	BF & LF	0.69
Recall	BF & SF & LF	**0.74**
	BF	0.67
	BF & SF	0.68
	BF & LF	0.65
F1-score	BF & SF & LF	**0.70**
	BF	0.66
	BF & SF	0.71
	BF & LF	0.65
AUC	BF & SF & LF	**0.71**

SVM: Support Vector Machine; NN: Neural Network; KNN: K-nearby Neighbors; NB: Naïve Bayes; DT: Decision Tree; RF: Random Forest; BF: Benchmark Features; SF: Semantic Features; LF: Linguistic Features.

Table 2
The individual performance scores of each classification model with different sets of textual features.

Performance Metrics	Features	Classification Models					
		SVM	NN	KNN	NB	DT	RF
	BF	0.62	0.62	0.61	0.65	0.62	0.62
	BF & SF	0.63	0.68	**0.64**	0.63	0.62	0.68
Precision	BF & LF	0.63	0.63	0.60	0.65	0.62	0.63
	BF & SF & LF	**0.64**	0.69	0.63	**0.65**	**0.63**	**0.68**
	BF	0.83	0.82	0.66	0.50	**0.78**	**0.79**
	BF & SF	0.83	0.75	0.71	0.79	0.61	0.63
Recall	BF & LF	0.83	0.81	0.69	0.48	0.62	0.68
	BF & SF & LF	**0.83**	0.84	**0.72**	**0.80**	0.62	0.64
	BF	0.71	0.71	0.63	0.56	**0.69**	**0.69**
	BF & SF	0.72	0.71	0.67	0.70	0.61	0.65
F1-score	BF & LF	0.71	0.71	0.64	0.55	0.62	0.65
	BF & SF & LF	**0.72**	0.80	**0.67**	**0.70**	0.62	0.66
	BF	0.71	0.69	0.48	0.70	**0.69**	0.69
	BF & SF	0.75	0.74	**0.70**	0.72	0.62	0.73
AUC	BF & LF	0.71	0.71	0.46	0.70	0.66	0.68
	BF & SF & LF	0.75	**0.74**	0.69	**0.72**	0.62	**0.73**

SVM: Support Vector Machine; NN: Neural Network; KNN: K-nearby Neighbors; NB: Naïve Bayes; DT: Decision Tree; RF: Random Forest; BF: Baseline Features; SF: Semantic Features; LF: Linguistic Features.

to the respective average scores of the benchmark features, the set of the benchmark features and our proposed linguistic features, and the set of the benchmark features and our proposed semantic features.

As shown in Table 2, the combination of the proposed full set of features and an NN model generates the best performance with respect to precision (0.69), recall (0.84), and F1-score (0.80) in comparison to other classification models. Its three scores outperform the respective scores achieved by any other combination of classification models and sets of textual features. The combination of the proposed full set of features and an SVM model generates one of the best performances with respect to AUC (0.75). In summary, our proposed set of features stands out from other sets of features with regard to the four perspectives of performance metrics, namely precision, recall, F1-score, and AUC.

The inclusion of either linguistic or semantic features does not always improve model performance in comparison to mere benchmark features. For example, in terms of recall, the NN model with the benchmark features (0.82) performs better than the model that uses the benchmark features along with linguistic features (0.81) or the model that uses the benchmark features along with semantic features (0.75). Only in a model that has both linguistic and semantic features along with benchmark features can the performance of the model prevail.

Furthermore, although the NB model with the benchmark features generates the highest precision rate (0.65) against other models with benchmark features (i.e., 0.62 for SVM, NN, DT, and RF and 0.61 for KNN), the NB model with the benchmark features has the worst performance against other models with the benchmark features in terms of recall (0.50) and F1-score (0.56). These results indicate the importance of using various evaluation metrics. The inclusion of F1-score, AUC, and recall in evaluation provides us with a more thorough view of the performance of the predictive models.

4. Discussion

Previous studies, such as those of Liang and Kee (2018) and Okunloye, Kee, Cummins, and Zhang (2023), indicated that language complexity is useful in predicting the popularity of an online petition.

Liang and Kee (2018) advocated a positive relationship between the complexity and the popularity. On the other hand, Okunloye et al. (2023) argued that linguistic complexity may lead to better information diffusion, suggesting that the difference may depend on the environment in which a piece of information is introduced. Regardless of the mixed findings, we suggest that the complexity of petitions together with other textual features proposed in this study can facilitate the discovery of online petitioning patterns.

Our results area also consistent with previous findings that future consideration is associated with civic participation. Some recent studies have indicated that future orientation is linked to greater concern with societal issues related to the Covid-19 pandemic (Lalot, Abrams, Ahvenharju, & Minkkinen, 2021) and moral consideration of artificial intelligence (Pauketat & Anthis, 2022). Specifically, Knudsen and Christensen (2021) indicated a strong association between people with a stronger consideration of distant future consequences and their non-institutionalized political participation. Although official response to a popular petition is not guaranteed on privately owned petition platforms, we are inclined to believe that privately owned petition platforms are considered as a non-institutionalized political channel for expressing views and engaging in civic participation. Our results are aligned with previous findings that temporal relevance in petition texts together with other proposed predictive textual features help improve the prediction of petition success and the identification of online petitioning patterns.

Some researchers have highlighted that narratives and framing of policy issues, e.g., gun control, can be rather dynamic across years (Lin & Chung, 2020). We adopted the time dimension to view policy categories and included policy categorical precedence as one semantic feature in our model. Although some researchers, e.g., Vromen, Halpin, and Vaughan (2022), have suggested that no relationship exists between topics of online petitions and government policy agendas, our results indicate that the inclusion of semantic features, including categorical precedence and symbolic topics, enhances prediction of petition success. Arguably, the popularity of different policy issues vary across years. Inclusion of semantic information into model development should be beneficiary, provided that we create appropriate features and incorporate them into the models.

Our results complement those of recent studies on online collective action and information diffusion. For example, King & Wang, 2023 investigation of misinformation diffusion on Twitter was aligned with our results that textual complexity together with other proposed textual features help improve predictive models of online petition success. In addition, Costello and Lee's (2022) findings were consistent with our results that the two linguistic features, namely complexity and concreteness, help develop a better predictive model of online petition success. Interestingly, Costello and Lee (2022) found that the use of vague language in entrepreneurial narratives reduces the success rate of online crowdfunding. They argued that the use of vague language reduces the quality of information presented in the entrepreneurial narratives and leads to investors having the impression of low credibility. This contrasts with Cobb and Elder's (1972) notion that linguistic concreteness reduces issue expansion. We suspect that the contrasting findings may be attributed to the financial commitment commonly involved in online crowdfunding platforms. This echoes the reason why online petitioning should be differentiated from other online collective actions. Nevertheless, the findings of both Costello and Lee (2022) and our study are inclined towards the notion that linguistic concreteness is influential in predicting the outcomes of online collective action.

Lastly, the positive effects on predictive performance in the models of online petition success indicate the usefulness of the long-standing wisdom of agenda setting in the current age of big data and social media. We interpret the results to suggest that social media and websites have become increasingly formal sources of information in the eyes of the online users. Thus, it is plausible that the responses of the general public to online petitions have become increasingly similar to their responses to articles published by the traditional media in the past. Previous scholars, e.g., Neuman et al. (2014), Harrison et al. (2022), and Gilardi et al. (2022), considered traditional media agendas and social media agendas as separate entities and investigated the antecedental relationships among them. We take another perspective to understand the dynamics among these agendas and examine the similarity between them. Integrating our results to the traditional wisdoms of agenda setting on petition popularity, we posit that people have possibly gotten used to these social platforms and have considered them formal sources of information like those of traditional media outlets. Thus, their responses and the pattern of issue expansion have become similar in the contexts of traditional agenda setting as well as agenda setting on social platforms of collective action.

5. Implications

5.1. Theoretical implications

Regarding theoretical implications, this study proposes a set of textual features, which it then uses to develop predictive models of online petition success. The results demonstrate that some relevant existing literature can be discovered for model development only if we view platforms of online collective action from different perspectives. Whereas existing studies largely viewed online petitioning from the perspective of online collective action, we adopt the perspective of agenda setting and thereby identify longstanding concepts, including the issue expansion model and symbolism, as relevant to online petitioning. These concepts help us identify relevant textual features for model development. We demonstrate how we can appropriately translate the insights of the previous literature into textual features that can be comprehended by computers for better prediction. A better predictive model developed and the pattern of online petitioning identified can advance our understanding of collective actions in the context of online petitioning.

It is noteworthy that some recent studies of online petitioning have mixed results. These studies tend to suggest that online petitions should be further classified into different issue categories. For example, Porten-Cheé et al. (2021) showed that popularity cues, i.e., numbers of signatures, of petitions lead to more participants of online petition platforms signing the petitions if the petitions are related to climate change. Yet the same conclusion could not be applied to petitions of social welfare. Koenig and McLaughlin (2018) distinguished between petition-supporting behavior and information-seeking behavior. They hypothesized that when petitions generate more anger among readers, the readers are more likely to sign/forward the petitions and less likely to seek further information about the petitions. The researchers found support for the former outcome but not the latter one. In addition, the researchers hypothesized that when petitions generate more anxiety among readers, the readers are less likely to sign/forward the petitions and more likely to seek more information. The results of the study were able to validate the latter outcome but not the former one. However, the non-systematic approach of pre-defining the petitions into issue categories adopted in their study could involve considerable human subjectivity. Moreover, predictive models developed for a specific issue category limit their potential usefulness for other policy issues.

In this study, we incorporate semantic features generated by both supervised (based on the code book prepared by two political science researchers) and unsupervised approaches (LDA) into predictive models. The supervised approach enabled us to incorporate experts' domain knowledge into the predicative models. Furthermore, to reduce potential inconsistency and subjectivity in manually classifying petitions into issue categories from time to time, we adopted the unsupervised approach, which aims to cluster the petitions into different topics of symbols. In comparison to those recent studies, which faced difficulties in validating their hypotheses in general datasets that cover issues across various policy areas, we demonstrate the use of advanced text-mining techniques, such as LDA, to complement our supervised approach for enhanced adaptability of our predictive models to the general datasets. Furthermore, the predictive models open the way to identify issue categories more systematically through detailed examination of the models. The pattern of online petitioning in each category may not necessarily be the same.

Additionally, the proposed predictive model development adopts a big-data approach and includes textual data that is exogenous from online petition platforms. Symbols are theoretically difficult to be identified by computers due to their vagueness. We recognize that many focusing events have symbolic meaning, and these events are commonly reported in front-page stories of the media. Thus, with the use of the *New York Times* corpus, which is exogenous from the petition platform, we strengthen the set of semantic features in our model. It is common to employ user-generated data on social media to make predictions of collective action outcomes, e.g., Mortensen, Neumayer, and Poell (2018), Syed and Silva (2023), and Tarafdar and Kajal Ray (2021). Less common is the use of professional journalists' texts to enhance the prediction of online petition success. Our study serves as a pointer on how standard, structured texts from the traditional mass media can be included in model development in the context of online petitioning. There may exist a systematic way to define or identify symbols by further exploration of the predictive models.

5.2. Practical implications

The behavioral data on these online petition platforms is something that was not available in the past. Governments and policy makers should take advantage of the ample supply of citizens' online behavioral data to develop good predictive models to identify those tiny "ripple" petitions that can bring major shocks and surprise to political and social stability and address the issues of these petitions in advance to mitigate the potential loss generated by the shocks and surprise. Yet it is infeasible to handle the large volume of petition data without the assistance of computers.

Our proposed text-mining model enables policy makers to handle a large volume of social data. It also helps them to monitor petitions in a relatively objective manner. Certain biases have existed with officials monitoring and handling these online petitions. Previous studies indicate that government officials tend to address petitions with less detailed content (Feng et al., 2023). Moreover, petitions with simple words are more likely to get a response from the officials, but those with complex words can get a higher level of officials' response (Lu et al., 2023). Appropriate analysis of these online petition platforms and other relevant social media is vital to understanding online activism (Hale, Margetts, & Yasseri, 2013). The results of predictive models of petition success may serve as a more objective metric to be considered in prioritizing various issues within governmental agendas. Based on the prediction of the model, they can address potentially popular issues at an early stage and prevent issue expansion, and thereby mitigate the social costs generated by public outcries concerning such issues.

Similar to some other early detection systems in assessing public concern, e.g., Collier et al. (2008) and Osakwe, Ikhapoh, Arora, and Bubu (2021), during pandemic seasons, our proposed model, i.e., the combination of the full set of textual features and the NN/SVM model, can serve as a core component of early detection systems of gauging public attention. Several scholars (e.g., Hagen, Keller, Yerden, & Luna-Reyes, 2019; Simonofski, Fink, & Burnay, 2021) have proposed the use of online petition data with data analytics and visualization tools for improved policymaking. Our predictive models can be incorporated into those tools to further support policymaking.

Lastly, this study facilitates activists to evaluate their online petitions to secure adequate support for their proposed policy changes in the competitive environment of online petitioning. Before posting their petition onto the online platforms, they can use our proposed model to evaluate their chances of success. Our model takes only the texts of the petitions into consideration and therefore does not necessitate in-depth knowledge of the social network related to the platform. Given that social network data involves privacy concerns, access to these data has become rather restricted in recent years. A model based solely on petition texts may be handier and more user-friendly. In comparison to models that rely on social networking data, our model stands out as being more sustainable.

With the prediction model, users can compare the predicted chance of success of two or more petition texts and post the best petition onto online petition platforms. The ease of attaining petition success with the use of the model will motivate citizens to continuously use online petition platforms and will strengthen civic participation in the long term. This helps fulfill the promised potential of e-participation (Kim & Lee, 2012; Macintosh, 2004), the goal of which is to enable a larger population to engage in democratic debates.

6. Conclusion and future directions

This study has some limitations. First, the findings may not be entirely applicable to petition platforms owned by governments whose responses to those highly popular petitions are sometimes guaranteed after certain criteria are met, although responses from the governments do not necessarily result in real policy changes. Avaaz is a privately owned petition platform, and governments are not obliged to respond to petitions on Avaaz. Nevertheless, the Avaaz platform was designed to enable its community to put pressure on policy makers for policy changes.[2] Previous studies also indicated that petitions of Avaaz targeted the government officials for policy changes (Halpin et al., 2018; Horstink, 2017). Second, our model development does not take the social media associated with petition platforms into consideration. Including data and features generated from social media would have strengthened the model development. Access to social media data, however, has become more restricted due to the emerging concerns about user privacy. Models based on these data will be seriously affected by the change of data access policies of these social media. Also, our proposed model can only help users evaluate the chance of their petition succeeding. Although people can compare the expected chance of success of two or more petition texts, determine the best petition, and post it onto online petition platforms with the use of our proposed model, our model does not inform people of how to improve their petition writing. Such guidance could support petition success and increase engagement in online democracy.

Furthermore, although we identify the pattern that the proposed set of textual features can help predict online petition success from the big data, we admit that interpreting relationships between each feature and the petition outcomes is challenging. As recognized by recent scholars, such as Dhar (2013), Salganik (2019) and Shrestha et al. (2021), the complexity of the prediction models, albeit conducive to prediction performance, may result in inadequate interpretability of results. For example, our results indicated that our proposed set of textual features with the NN model perform the best among other models in terms of precision, recall and F1 score. Unlike a linear regression model, the NN model is good at fitting arbitrarily complex polynomials which are difficult to interpret and explain. Thus, the influences of individual textual features on petition success require further investigation in future. Lastly, whereas we successfully use topics as proxies for political symbols for model development, the current research method cannot guarantee that the political symbols are fully imported into the models. Scholars can conduct further studies on the development of models based solely on political symbols in the future.

Although the long-standing agenda-setting literature features traditional, well-established media, this study shows that past literature can still provide us with insights and ideas about how to improve the predictive models of petition success in the online environment where

[2] https://secure.avaaz.org/page/en/about/.

traditional media does not dominate. Future scholars may consider incorporating the concept of participatory journalism[3] into our proposed model and include social media data of petition platform users in the model development. These users, thanks to their strong interest in current affairs and politics, may report and comment on recent policy development in various policy areas in a real-time manner. Their behavior of reporting and commenting on social media may provide researchers with further insights. Previous researchers in agenda setting have widely discussed the important role of the mass media in setting the public agenda, gaining public attention, and subsequently influencing government actions on certain issues (e.g., Downs, 1996; Lodge & Hood, 2002). The mass media was able to direct readers' thinking, highlight the salience of certain issues, and imply genuine reasons for issues (McCombs & Shaw, 1972). However, the traditional, well-established mass media has been reported to have a rather weak influence on the online environment (e.g., Meraz, 2009; Wu, Atkin, Lau, Lin, & Mou, 2013). On the other hand, the social media users can now play the roles of reporters in the online environment. The roles of the mass media's online outlets and those self-motivated reporters on the social media have not been considered in this study. Future researchers may examine the interaction among online petition content, responses of users on different social media as well as the mass media content, and translate the interaction into features that can be incorporated into classification models for better prediction. Lastly, the individual factors of platform users can be taken into consideration to develop a better usage model of online petition platforms. Factors, such as users' achievement goals and their social capital have been reported to affect their usage behavior on platforms of collective action (e.g., Choi & Song, 2020; Lee, Lui, Chau, & Tsin, 2023). It is believed that incorporation of platform users' individual differences may further improve the predictive models.

CRediT authorship contribution statement

Philip Tin Yun Lee: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Formal analysis. **Alvin Ying Lu:** Formal analysis, Methodology, Writing – original draft. **Feiyu E:** Formal analysis. **Michael Chau:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

None.

Acknowledgements

This project is supported in part by the University Research Grant of Hong Kong Shue Yan University (Ref. No.: URG/22/01).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.giq.2024.101937.

References

Aragón, P., Sáez-Trumper, D., Redi, M., Hale, S., Gómez, V., & Kaltenbrunner, A. (2018, June). Online petitioning through data exploration and what we found there: A dataset of petitions from avaaz.org. In *Twelfth international AAAI conference on web and social media. Palo Alto, CA*.

Balasubramanian, S. K., Fang, Y., & Yang, Z. (2021). Twitter presence and experience improve corporate social responsibility outcomes. *Journal of Business Ethics, 173*, 737–757.

Baumgartner, F. R., & Jones, B. D. (2010). *Agendas and instability in American politics*. University of Chicago Press.

Besiou, M., Hunter, M. L., & Van Wassenhove, L. N. (2013). A web of watchdogs: Stakeholder media networks and agenda-setting in response to corporate initiatives. *Journal of Business Ethics, 118*, 709–729.

Birkland, T. A. (1997). *After disaster: Agenda setting, public policy, and focusing events*. Georgetown University Press.

Birkland, T. A. (1998). Focusing events, mobilization, and agenda setting. *Journal of Public Policy, 18*(1), 53–74.

Birkland, T. A. (2017). Agenda setting in public policy. In F. Frank, J. M. Gerald, & S. S. Mara (Eds.), *Handbook of public policy analysis: Theory, politics, and methods* (pp. 63–78). CRC Press.

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

Bowman, S., & Willis, C. (2003). *We media. How audiences are shaping the future of news and information*. American Press Institute.

Boykoff, M. T. (2011). *Who speaks for the climate?: Making sense of media reporting on climate change*. Cambridge University Press.

Brown, N., & Deegan, C. (1998). The public disclosure of environmental performance information—A dual test of media agenda setting theory and legitimacy theory. *Accounting and Business Research, 29*(1), 21–41.

Brunsting, S., & Postmes, T. (2002). Social movement participation in the digital age: Predicting offline and online collective action. *Small Group Research, 33*(5), 525–554.

Carroll, C. E., & McCombs, M. (2003). Agenda-setting effects of business news on the public's images and opinions about major corporations. *Corporate Reputation Review, 6*, 36–46.

Chen, Y., Deng, S., Kwak, D., Elnoshokaty, A., & Wu, J. (2019). A multi-appeal model of persuasion for online petition success: A linguistic Cue-based approach. *Journal of the Association for Information Systems, 20*(2), 105–131.

Choi, J. C., & Song, C. (2020). Factors explaining why some citizens engage in E-participation, while others do not. *Government Information Quarterly, 37*(4), Article 101524.

Cobb, R., Ross, J. K., & Ross, M. H. (1976). Agenda building as a comparative political process. *American Political Science Review, 70*(1), 126–138.

Cobb, R. W., & Elder, C. D. (1972). *Participation in American politics: The dynamics of agenda-building*. Johns Hopkins University Press.

Collier, N., Doan, S., Kawazoe, A., Goodwin, R. M., Conway, M., Tateno, Y., … Taniguchi, K. (2008). BioCaster: Detecting public health rumors with a web-based text mining system. *Bioinformatics, 24*(24), 2940–2941.

Costello, F. J., & Lee, K. C. (2022). Exploring investors' expectancies and its impact on project funding success likelihood in crowdfunding by using text analytics and Bayesian networks. *Decision Support Systems, 154*, Article 113695.

Craig, T. Y., & Blankenship, K. L. (2011). Language and persuasion: Linguistic extremity influences message processing and behavioral intentions. *Journal of Language and Social Psychology, 30*(3), 290–310.

Dhar, V. (2013). Data science and prediction. *Communications of the ACM, 56*(12), 64–73.

Downs, A. (1996). Up and down with ecology: The "issue-attention cycle". In P. Peretz (Ed.), *The politics of American economic policy making* (pp. 48–59). M. E. Sharpe, Inc.

Dumas, C. (2022). E-petitioning as online collective action in we the people: The case of the legalization of marijuana in the US. In *DG. O 2022: The 23rd annual international conference on digital government research* (pp. 151–165).

Dumas, C. L., LaManna, D., Harrison, T. M., Ravi, S., Kotfila, C., Gervais, N., … Chen, F. (2015). Examining political mobilization of online communities through e-petitioning behavior in we the people. *Big Data & Society, 2*(2). https://doi.org/10.1177/2053951715598170

Edelman, M. (1974). The political language of the helping professions. *Politics and Society, 4*(3), 295–310.

Feng, X., Wang, C., & Wang, J. (2023). Understanding how the expression of online citizen petitions influences the government responses in China: An empirical study with automatic text analytics. *Information Processing & Management, 60*(3), Article 103330.

Finkel, J. R., Grenager, T., & Manning, C. D. (2005, June). Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)* (pp. 363–370).

Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2022). Social media and political agenda setting. *Political Communication, 39*(1), 39–60.

Greijdanus, H., de Matos Fernandes, C. A., Turner-Zwinkels, F., Honari, A., Roos, C. A., Rosenbusch, H., & Postmes, T. (2020). The psychology of online activism and social movements: Relations between online and offline collective action. *Current Opinion in Psychology, 35*, 49–54.

Grossman, E. (2022). Media and policy making in the digital age. *Annual Review of Political Science, 25*, 443–461.

Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing & Management, 54*(6), 1292–1307.

Hagen, L., Harrison, T. M., Uzuner, Ö., May, W., Fake, T., & Katragadda, S. (2016). E-petition popularity: Do linguistic and semantic factors matter? *Government Information Quarterly, 33*(4), 783–795.

Hagen, L., Keller, T. E., Yerden, X., & Luna-Reyes, L. F. (2019). Open data visualizations and analytics as tools for policy-making. *Government Information Quarterly, 36*(4), Article 101387.

[3] In Bowman and Willis' (2003) book *We Media, How audiences are shaping the future of news and information,* they defined participatory journalism as "the act of a citizen, or group of citizens, playing an active role in the process of collecting, reporting, analyzing and disseminating news and information" (p. 9).

Hale, S. A., Margetts, H., & Yasseri, T. (2013, May). Petition growth and success rates on the UK no. 10 downing street website. In *Proceedings of the 5th annual ACM web science conference* (pp. 132–138).
Halpin, D., Vromen, A., Vaughan, M., & Raissi, M. (2018). Online petitioning and politics: The development of change. Org in Australia. *Australian Journal of Political Science, 53*(4), 428–445.
Hansford, T. G., & Coe, C. (2019). Linguistic complexity, information processing, and public acceptance of supreme court decisions. *Political Psychology, 40*(2), 395–412.
Harrison, T. M., Dumas, C., DePaula, N., Fake, T., May, W., Atrey, A., … Ravi, S. S. (2022). Exploring E-petitioning and media: The case of # BringBackOurGirls. *Government Information Quarterly, 39*(1), Article 101569.
Hartelius, J. (2011). Rhetorics of expertise. *Social Epistemology, 25*(3), 211–215.
Hong, J., & Hoban, P. R. (2022). Writing more compelling creative appeals: A deep learning-based approach. *Marketing Science, 41*(5), 941–965.
Hong, Y., Hu, Y., & Burtch, G. (2018). Embeddedness, pro-sociality, and social influence: Evidence from online crowdfunding. *MIS Quarterly, 42*(6), 1211–1224.
Horstink, L. (2017). Online participation and the new global democracy: Avaaz, a case study. *Global Society, 31*(1), 101–124.
Jann, W., & Wegrich, K. (2017). Agenda setting in public policy. In F. Frank, J. M. Gerald, & S. S. Mara (Eds.), *Handbook of public policy analysis: Theory, politics, and methods* (pp. 43–62). Routledge.
Kim, S., & Lee, J. (2012). E-participation, transparency, and trust in local government. *Public Administration Review, 72*(6), 819–828.
King, K. K., & Wang, B. (2023). Diffusion of real versus misinformation during a crisis event: A big data-driven approach. *International Journal of Information Management, 71*, 102390.
Kingdon, J. W. (2010). *Agendas, alternatives, and public policies* (2nd ed.). Pearson.
Knudsen, M. S., & Christensen, H. S. (2021). Future orientation and political participation: The moderating role of political trust. *Frontiers in Political Science, 3*, Article 791467.
Koenig, A., & McLaughlin, B. (2018). Change is an emotional state of mind: Behavioral responses to online petitions. *New Media & Society, 20*(4), 1658–1675.
Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*.
Lalot, F., Abrams, D., Ahvenharju, S., & Minkkinen, M. (2021). Being future-conscious during a global crisis: The protective effect of heightened futures consciousness in the COVID-19 pandemic. *Personality and Individual Differences, 178*, Article 110862.
Lang, H. (2019). # MeToo: A case study in re-embodying information. *Computers and Composition, 53*, 9–20.
Lash, M. T., & Zhao, K. (2016). Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems, 33*(3), 874–903.
Lee, P. T. Y., Lui, R. W. C., Chau, M., & Tsin, B. H. Y. (2023). Exploring the effects of different achievement goals on contributor participation in crowdsourcing. *Information Technology & People, 36*(3), 1179–1199.
Liang, Y., & Kee, K. F. (2018). Developing and validating the ABC framework of information diffusion on social media. *New Media & Society, 20*(1), 272–292.
Lin, Y. R., & Chung, W. T. (2020). The dynamics of twitter users' gun narratives across major mass shooting events. *Humanities and Social Sciences Communications, 7*(1), 1–16.
Lodge, M., & Hood, C. (2002). Pavlovian policy responses to media feeding frenzies? Dangerous dogs regulation in comparative perspective. *Journal of Contingencies & Crisis Management, 10*(1), 1–13.
Lu, L., Xu, J., & Wei, J. (2023). Understanding the effects of the textual complexity on government communication: Insights from China's online public service platform. *Telematics and Informatics, 102028*.
Macintosh, A. (2004). Characterizing e-participation in policy-making. In *37th Annual Hawaii international conference on system sciences, 2004. Proceedings*. IEEE.. pp. 10-pp.
Margetts, H., Hale, S., & John, P. (2019). How social media shapes political participation and the democratic landscape. In *Society and the internet: How networks of information and communication are changing our lives* (p. 197).
McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly, 36*(2), 176–187.
Meraz, S. (2009). Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication, 14*(3), 682–707.
Miller, G. (1998). *WordNet: An electronic lexical database*. MIT press.
Mo, J., Sarkar, S., & Menon, S. (2018). Know when to run: Recommendations in crowdsourcing contests. *MIS Quarterly, 42*(3), 919–944.
Mortensen, M., Neumayer, C., & Poell, T. (Eds.). (2018). *Social media materialities and protest: Critical reflections*. Routledge.
Negrine, R., & Papathanassopoulos, S. (2011). The transformation of political communication. In *Media perspectives for the 21st century* (pp. 41–54). Routledge.
Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication, 64*(2), 193–214.
Ngai, E. W. T., & Lee, P. T. Y. (2016, January). A review of the literature on applications of text mining in policy making. In *In proceedings of 20th Pacific Asia conference on information systems, PACIS 2016*. Pacific Asia Conference on Information Systems.
Okunloye, O., Kee, K. F., Cummins, R. G., & Zhang, W. (2023). The linguistic and message features driving information diffusion on twitter: The case of# RevolutionNow in Nigeria. *International Journal of Communication, 17*, 22.
Osakwe, Z. T., Ikhapoh, I., Arora, B. K., & Bubu, O. M. (2021). Identifying public concerns and reactions during the COVID-19 pandemic on twitter: A text-mining analysis. *Public Health Nursing, 38*(2), 145–151.
Pauketat, J. V., & Anthis, J. R. (2022). Predicting the moral consideration of artificial intelligences. *Computers in Human Behavior, 136*, Article 107372.
Petty, R. E., & Briñol, P. (2015). Emotion and persuasion: Cognitive and meta-cognitive processes impact attitudes. *Cognition and Emotion, 29*(1), 1–26.
Porten-Cheé, P., Kunst, M., Vromen, A., & Vaughan, M. (2021). The effects of narratives and popularity cues on signing online petitions in two advanced democracies. *Information, Communication & Society*, 1–21.
Ragas, M. W., Tran, H. L., & Martin, J. A. (2014). Media-induced or search-driven? A study of online agenda-setting effects during the BP oil disaster. *Journalism Studies, 15*(1), 48–63.
Rodríguez-Temiño, I., & Almansa-Sánchez, J. (2021). The use of past events as political symbols in Spain. The example of Vox and the need for a new archaeology of ethnicity. *International Journal of Heritage Studies, 27*(10), 1064–1078.
Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.
Saxton, G. D., Ren, C., & Guo, C. (2021). Responding to diffused stakeholders on social media: Connective power and firm reactions to CSR-related twitter messages. *Journal of Business Ethics, 172*, 229–252.
Schattschneider, E. E. (1975). *The semi-sovereign people: A realist's view of democracy in America*. Cengage Learning.
Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping theory*. Routledge.
Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science, 32*(3), 856–880.
Siering, M., Koch, J. A., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems, 33*(2), 421–455.
Simonofski, A., Fink, J., & Burnay, C. (2021). Supporting policy-making with social media and e-participation platforms data: A policy analytics framework. *Government Information Quarterly, 38*(3), Article 101590.
Sparrow, B. H. (2006). A research agenda for an institutional media. *Political Communication, 23*(2), 145–157.
Stone, D. A. (1989). Causal stories and the formation of policy agendas. *Political Science Quarterly, 104*(2), 281–300.
Stone, D. A. (2002). *Policy paradox: The art of political decision making*. Norton.
Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. Oxford, UK: MIT Press.
Suh, J. H., Park, C. H., & Jeon, S. H. (2010). Applying text and data mining techniques to forecasting the trend of petitions filed to e-people. *Expert Systems with Applications, 37*(10), 7255–7268.
Syed, R., & Silva, L. (2023). Social movement sustainability on social media: An analysis of the women's march movement on twitter. *Journal of the Association for Information Systems, 24*(1), 249–293.
Tarafdar, M., & Kajal Ray, D. (2021). Role of social media in social protest cycles: A sociomaterial examination. *Information Systems Research, 32*(3), 1066–1090.
The Policy Agendas Project. (2017). Retrieved from https://www.comparativeagendas.info/ Accessed January 2, 2023.
True, J. L., Jones, B. D., & Baumgartner, F. R. (2019). Punctuated-equilibrium theory: Explaining stability and change in public policymaking. In *Theories of the policy process* (pp. 155–187). Routledge.
Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society, 20*(5), 2028–2049.
Vicente, M. R., & Novo, A. (2014). An empirical analysis of e-participation. The role of social networks and e-government over citizens' online engagement. *Government Information Quarterly, 31*(3), 379–387.
Vromen, A., Halpin, D., & Vaughan, M. (2022). What kinds of issues do citizens successfully raise via online petitions?. In *Crowdsourced politics: The rise of online petitions & micro-donations* (pp. 73–94). Singapore: Springer Nature Singapore.
Wright, S. (2015). E-petitions. In S. Coleman, & D. Freelon (Eds.), *Handbook of digital politics* (pp. 136–150). Edward Elgar Publishing.
Wu, Y., Atkin, D., Lau, T. Y., Lin, C., & Mou, Y. (2013). Agenda setting and micro-blog use: An analysis of the relationship between Sina Weibo and newspaper agendas in China. *The Journal of Social Media in Society, 2*(2), 8–25.
Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 1480–1489).
Zahariadis, N. (2016). Setting the agenda on agenda setting: Definitions, concepts, and controversies. In *Handbook of public policy agenda setting* (pp. 1–22). Edward Elgar Publishing.
Zyglidopoulos, S. C., Georgiadis, A. P., Carroll, C. E., & Siegel, D. S. (2012). Does media attention drive corporate social responsibility? *Journal of Business Research, 65*(11), 1622–1627.

Philip Tin Yun Lee (tylee@hksyu.edu) is an assistant professor in Faculty of Commerce at Hong Kong Shue Yan University. He received his Ph.D. in management information systems from the University of Hong Kong, where he worked on research topics related to text mining and mixed-motivational systems. He also holds a master's degree in Public Policy & Administration from LSE and a master's degree in Computer Science from the University of Hong Kong. He has published papers in various conferences and journals, such as Internet Research and IT and People.

Alvin Ying Lu (luying@comp.nus.edu.sg) is currently a PhD candidate in the Department of Information Systems and Analytics at the National University of Singapore. He has published research in conferences and journals in the field of Information Systems. He utilizes multiple methodologies to study the impact of emerging technologies on human

behavior. His research interests include online political governance, decentralized social networks, and the IT labor market.

Feiyu E (u3005658@connect.hku.hk) is currently a post-doctoral fellow at the University of Maryland. She received her Ph.D. in Management of Information Systems from the University of Hong Kong. Her research interests focus on the topics of social media information analysis and disaster management, with research methods and tools drawn from reference disciplines including Information Systems (e.g., Design Science, Econometrics and Survey) and Computer Science (e.g., Text Analytics and Machine Learning).

Michael Chau (mchau@business.hku.hk) is a professor in the Faculty of Business and Economics at the University of Hong Kong. He received his Ph.D. in management information systems from the University of Arizona. His interests focus on the intersection of information systems, computer science, business analytics, and information science, with a focus on the applications of data, text, and web mining in various domains. His research has resulted in over 120 publications in journals and conferences. He was the Program Co-chair of ICIS 2013 and the Founding Co-Chair of the Pacific Asia Workshop on Intelligence and Security Informatics series.