



Original research article

Hydraulic fracturing and political conflict: News media coverage of topics and themes across nine states



Ramiro Berardo^{a,*}, Federico Holm^a, Tanya Heikkilä^b, Christopher M. Weible^b, Hongtao Yi^c, Jennifer Kagan^d, Catherine Chen^c, Jill Yordy^b

^a School of Environment and Natural Resources, The Ohio State University, Kottman Hall 316 B, Columbus, OH 43210, United States

^b School of Public Affairs, University of Colorado, Denver, United States

^c Glenn College of Public Affairs, The Ohio State University, United States

^d Public Administration Program, School of Social Sciences, University of Hawaii, Mānoa, United States

ARTICLE INFO

Keywords:

Hydraulic fracturing

Conflict

Text analysis

Topic modeling

Latent dirichlet allocation

United States

ABSTRACT

This article compares the topics that underlie public debate around hydraulic fracturing covered in newspapers across nine U.S. states over an eleven-year period. In analyzing more than 7000 newspaper articles using Latent Dirichlet Allocation (LDA) modeling, thirteen main topics emerge. While these topics fluctuate over time, their relative frequency and, hence, importance in the discourse remains largely constant. The environmental risks associated with the practice is the topic that receives more attention when all data are aggregated. We find that the frequency of the topics varies by state, and the nature of this variation is associated with the political leanings of the state, with media sources in Republican governed states more likely to report on the economic benefits associated with hydraulic fracturing. Finally, we show how all topics are associated with words that indicate the presence of conflict among stakeholders involved in discussions about the costs and benefits of hydraulic fracturing. In doing so, we describe how the association between topics and conflict varies according to which party governs the state, which provides evidence about the fundamental differences on how parties consider the practice of hydraulic fracturing in the states we study. We conclude the article by discussing the advantages and disadvantages of our methodological approach, which can be leveraged to discern trends in discussions about environmental and energy-related problems that exceed the specific case of hydraulic fracturing.

1. Introduction

In the last decade, the widespread and increasing use of hydraulic fracturing technology for oil and gas production in the United States has had sizable economic and socio-political effects. As thousands of new wells are drilled each year across the U.S., debates over the economic merits and environmental impacts of hydraulic fracturing continue, and proponents and opponents of hydraulic fracturing vie to influence governmental decisions that can deeply affect the future of oil and gas development.

The nature of these debates can be simplified into two polarized political positions. On one side, proponents have argued that hydraulic fracturing helps communities raise taxes, create jobs, increase income for landowners who lease their mineral rights to oil and gas companies, and increase business opportunities for the service sector, among other positive impacts [1,2]. On the other side, opponents have emphasized

the myriad negative effects of oil and gas production using hydraulic fracturing: the boom and bust economic cycles associated with it; public health effects resulting from occupational and community risks; social stress caused by the disruption of residents' sense of place and identity; air, noise and light pollution; overuse of water resources (in many cases in water-stressed areas) and reduction of water quality; increased seismic activity; and potentially noxious effects on wildlife in proximity to wells that are hydraulically fractured [3–9].

Given the novelty of hydraulic fracturing as a major technology for oil and gas production with uncertain and ambiguous benefits and costs, contentious debates continue about whether hydraulic fracturing should be widely used and, if so, how. For instance, in their nationally representative sample of over 1000 survey respondents, Boudet and colleagues [1] found that “among the minority who has formed an opinion, respondents were nearly split between support and opposition.” Moreover, Christenson and colleagues [10] point to research

* Corresponding author.

E-mail address: berardo.3@osu.edu (R. Berardo).

<https://doi.org/10.1016/j.erss.2020.101660>

Received 23 October 2019; Received in revised form 4 June 2020; Accepted 11 June 2020

2214-6296/ © 2020 Elsevier Ltd. All rights reserved.

suggesting that public discourse over hydraulic fracturing as reported in mass media outlets reflects sharp disagreements about the relative costs and benefits of the technology [11].

However, most of the research on these debates has focused on a single location (e.g., a single state or single shale play) and is often not comparative across states. Hence, we lack an understanding of how attributes of these debates differ across political jurisdictions. In this article, we aim to close this gap. We examine the topics of public debates about hydraulic fracturing across nine U.S. states over a period of more than a decade (from early 2007 to late 2017) and assess how much conflict is associated with each topic. The nine states we examine, California, Colorado, New York, North Dakota, Ohio, Oklahoma, Pennsylvania, Texas, and West Virginia, collectively hold most of the shale natural gas reserves in the U.S. (~76%) and are located in the most productive tight plays for crude oil production in the country [12]. In each of these states, proponents and opponents of hydraulic fracturing have engaged in lengthy and often contentious debates about its advantages and disadvantages. Given that debates often expand into extreme political contentiousness, disruption of the status quo, and degenerative forms of policymaking, the intensity of conflict per topic in these debates is deserving of scholarly understanding [13].

As has been shown by multiple studies of public discourse and narratives, written news media remains one of the most cost-efficient, consistent, and reliable data sources for gauging policy conflicts [14–16]. We engage in a comprehensive effort to code information available in written media outlets to describe the topics that dominate discussions about hydraulic fracturing, and the level of conflict associated with them. While there is outstanding previous research on the content of newspaper reporting on hydraulic fracturing in the U.S., the focus of prior studies has been mostly confined to a small group of newspapers with national reach [17–20], or to a larger group of outlets but only in states with access to the Marcellus shale [21,22]. By analyzing the content of reporting about hydraulic fracturing, the dominant topics, and the level of conflict associated with them over an extended period and in a large number of states, we produce the most comprehensive analysis to date of the public debate about hydraulic fracturing that takes place in written media outlets in the U.S.

Our goal is to identify topics that dominate reporting on hydraulic fracturing and to describe how the frequency of these topics varies across states and over time and associates with policy conflict. Four research questions guide our effort: (1) What major topics emerge in the news media? (2) At what frequencies have the topics appeared over time and across U.S. states? (3) How do the topic frequencies vary across states governed by different parties? And (4) To what extent do the topics associate with policy conflict? We answer these research questions by analyzing more than 7000 newspaper articles using Latent Dirichlet Allocation (LDA) modeling, which is a valuable approach to detect general tendencies in the content of public reporting and discussions around policy topics such as unconventional oil and gas development.

2. Policy debates in the news media in the U.S.

Scholars of politics and public policy have long researched how public attention to policy problems shapes policy making [23,24]. Both decision makers and policy stakeholders adjust their political strategies to accommodate new available information—some of which may be shaped by the shifts in public attention regarding particular issues of interest [25–28].

Changes in public discourse, which are often reflected in the news media, sometimes can lead to or trigger policy change [24]. For most attention-grabbing policy issues (e.g. climate change, abortion, gay marriage, capital punishment, immigration reform), news media reports reflect the tenor of public discussions about policy topics, which are frequently characterized by high levels of conflict, particularly when different parties in the discussion hold deeply entrenched and

opposing views [29,30].

Scholars have theoretically and empirically examined how conflicts become “contagious” across space and time [13]. Intense policy conflicts can shape the formation and structure of rival coalitions, their endurance over time and their capacity to learn about policy problems and possibly overcome their differences [31]. Recent approaches have developed the cognitive and behavioral characteristics of policy conflicts – i.e. how policy actors come to diverge on policy topics and what they do to advance their agendas once those disagreements become evident [32]. While policy conflicts are not necessarily harmful to the quality of decision-making processes, and can be an integral part of healthy public deliberations in complex governance systems [32,33], excessive levels of conflict can act as a destabilizing force in policy systems, leading to the breakdown of the deliberative processes that could have produced mutually advantageous policy decisions to actors with dissimilar goals [34,35].

Hydraulic fracturing is a highly controversial topic, with public debates and discussions about the practice tending to be highly contentious. For instance, Brasier et al. [4] identified the potential for polarization and conflict among landowners. Jacquet [36] highlighted the likelihood for conflict at the local level, particularly when interpersonal trust and local leadership were not present to mediate debates. Heikkilä and Weible [37] examined divergent policy positions, unwillingness to compromise, and risk perceptions of stakeholders in Colorado and found how these characteristics of conflict could imperil shared views of the costs and benefits of hydraulic fracturing. Zilliox and Smith [38] showed that conflict could emerge even with collaborative-based efforts by local governments to involve citizens in processes designed to create formal agreements between local authorities and oil and gas companies to expand the use of hydraulic fracturing. Beyond the U.S., political disputes over hydraulic fracturing have occurred all over the world, in countries such as Argentina [39,40], Canada [20,41], South Africa [42,43], and France [44], to name a few.

Despite the plethora of studies examining how the increasing use of hydraulic fracturing can spark conflicts when citizens self-organize to oppose the practice, most of this research has not explicitly measured the breadth and level of conflict associated with news media discourse and reporting about hydraulic fracturing. Important work such as Blair and McCormack [45,46], Heikkilä et al. [47], and Gottlieb et al. [18] have analyzed newspaper content about hydraulic fracturing, but at a smaller geographic scale and without a specific focus on conflict.

In this article, we rely on a novel methodological approach that allows us to investigate how hydraulic fracturing discussions are portrayed in the media across U.S. states that have seen considerable conflict erupt in debates about the economic benefits and environmental costs of the practice. We leverage the advantages of automated content analysis tools and topic modeling to identify topics around which discussions are structured and examine how they change or not through time.

3. Methodological approach

To examine the breadth and levels of conflict in debates about hydraulic fracturing in the nine states where we focus our analysis, we use a topic modeling approach based on LDA. In general, topic modeling can be understood as a set of methods to uncover patterns in textual sources that reflect underlying themes or topics [48]. In recent years, topic modeling has become more attractive as a data-exploration and analysis approach as increased computational power has allowed for scaling up the technique to analyze large amounts of textual data. Topic modeling is based on latent semantic indexing (LSI), a previously developed method in Natural Language Processing [49,50] that assumes that the interdependence between words in a document can be explained by the underlying (or latent) topics contained in the document [45,51].

One of the most widely used algorithms in topic modeling is LDA,

which is a Bayesian mixed model for discrete data in which words and documents are assumed to be uncorrelated¹ [45,46]. LDA is a generative probabilistic model² that provides a powerful tool for discovering and exploring the underlying thematic structures in large datasets [46]. The basic idea behind LDA is that documents may contain multiple topics, each of which can be represented as a probability distribution of a vocabulary of terms [46,48]. The documents themselves are the observed data, whereas the topic structure – the topics, topic distributions per-document, and the topic assignments per-document per-word – is the underlying structure to be uncovered [52]. The key in LDA is that any document or piece of text is “composed by selecting words from possible bags-of-words, where each bag corresponds to a topic” [45].

In LDA, each document is modeled as a mixture of topics. The number of topics is determined by a topic-by-document distribution parameter. When the value of this parameter is set at a lower level, the model is forced to fit to a document using fewer topics than when the parameter value is higher. When the value of the parameter is increased, a larger number of topics results [46,53]. It is important to keep in mind that setting a lower parameter for a large corpus would not necessarily result in fewer revealed topics, and that the topics that do appear might not be as consistent as one might find if the value of the parameter were increased. The exercise of tweaking the parameter comes with a tradeoff, because a larger number of topics decreases the amount of data available to estimate each of the individual topics [45]. The larger the corpus of text to be analyzed, the more topics we can identify without facing issues related to insufficient data.

Uncovering the topic structure underlying newspaper articles on hydraulic fracturing allows us to explore how discussions and reporting about this technology have evolved over the years. Moreover, it also allows us to understand which are the critical topics associated with conflict in each state, because we pair the detection of topics with an analysis of conflict words in the articles we examine. We provide a more complete description of this process in the following sections.

3.1. Data sources and structure

The source data we use to detect topics related to hydraulic fracturing are newspaper articles from nine U.S. states that hold significant oil and gas reserves and where social conflict over hydraulic fracturing has been common: California, Colorado, North Dakota, New York, Oklahoma, Ohio, Pennsylvania, Texas and West Virginia. For each of these states, we collected data from 2007 to 2017 from two newspapers likely to offer extensive coverage on the topic of unconventional oil and gas exploration.³ The first newspaper in each state is the main paper in a major city or the state's capital city, where state policy debates and decision making are typically centered. The second newspaper is from a city or town in close proximity to oil and gas activities. By selecting newspapers in this manner, we aim to reduce the likelihood of obtaining biased analyses. Such biases are more likely when only local

¹ This is based on the “bag-of-words” assumption, which posits that the order in which the words appear in a document (or the order of documents themselves) can be neglected [46].

² The term “generative” is used to point out to “the imaginary random process by which the model assumes the documents arose” [52, pp. 78]. By understanding that the texts are generated by a random combination of topics (with a probability distribution over a set of words), the model allows to “reverse” the generative process and identify the topics by looking at the distribution of words [52].

³ 2007 is the year that U.S. news media started to cover the use of hydraulic fracturing in unconventional oil and gas development on a regular basis, which reflected the onset of this technology to help access previously untapped shale formations across the US. We extend our analysis to 2017, as a decade provides a reasonable time period to capture how a policy debate shifts or changes alongside changing political or socio-economic conditions.

news sources are used, since they tend to be more sensitive to local-level variables that might affect the way people frame discussions and debates [54].

To identify relevant articles in the 18 newspapers we chose, we used a two-step process. First, we used Newsbank to identify potentially relevant articles published between 2007 and 2017 using the following search terms: “fracking”, “hydraulic fracturing”, “shale oil”, “shale gas”, “oil drilling”, or “gas drilling”. Second, we reviewed the articles to ensure that they report or reflect discussions and debates about hydraulic fracturing in the state. In this process, we also excluded opinion pieces, such as op-eds and letters to the editor. This exercise resulted in 7616 newspaper articles across nine states, which we then subdivided into subsets of articles based on the year in which they were published. This task yielded 98 subcorpora (one subcorpus per year, per state, except for the state of New York, which had no fracking-related articles for 2007), with high variability in the number of articles present in each: the smallest subcorpus has 3 articles (Ohio, 2007), whereas the largest have 331 (Pennsylvania, 2011 and 2012). A table detailing the number of articles per year per state is available in the appendix.

To prepare the texts for topic modeling, our first step was to remove common English stop words, which are words that provide no information about the content of the text. We also created a custom list of extra words to be removed before fitting the models – again, highly common words that provide no meaningful information on the topic or context discussed.⁴ Second, we fixed the sequence of words to keep the order and structure in which words appear in the text, so we could build combinations of adjacent words. Finally, we reduced text to word bigrams for further analysis. This allows phrases that connote important meanings within the text, such as “climate change”, “severance tax” or “local ban” to be recoded as a single term (replacing the blank space between the two adjacent words with an underscore). This provides context when interpreting the overall theme of the group of bigrams, allowing key groups of words to be placed next to each other.

3.2. Topic model specification and implementation

To model and identify topics, we used Mallet, a Java-based package for statistical Natural Language Processing, document classification, clustering, topic modeling, and other machine learning applications for text analysis [55]. We used a Gibbs-sampling-based implementation of LDA to fit the models and identify topics. When fitting a model, Mallet allows the user to define the following input parameters:

- Number of topics: Since there is no set rule to define the number of topics, we followed an inductive approach based on coherence of the topics identified, combined with prior knowledge of the context in each state. Greater corpora allowed for the identification of a greater number of topics, so we fit numerous models to each corpus and settled on the number of topics that yielded maximum coherence within each topic.
- Interval optimization: This is the number of iterations between re-estimating Dirichlet hyperparameters.⁵ We found that 100 iterations were enough to get consistent results. The default is 0.

⁴ Stop words include those as my, myself, we, our, ours, than, too, the, he, she, very, etc. The additional list of words we excluded from the analysis includes words such as ‘Sen.’, ‘Rep.’, ‘Gov.’, ‘Dem.’, or common verbs such as ‘will’, ‘make’, ‘want’, ‘see’, ‘has’, ‘have’, etc. These words are not related to any topic in any meaningful way, and they add noise to the estimation of the models. Finally, we found that the words ‘fracking’, ‘oil’ and ‘gas’ were also evenly distributed across all topics and did not add any meaningful information to topic identification. This is not surprising, since they form the baseline criteria for finding relevant articles in the first place.

⁵ For a great discussion on the importance in the estimation and role of the hyperparameters controlling the prior conjugate Dirichlet distribution in LDA, see Wallach, et al. [53].

- c. Burn-in optimization: This is the number of iterations to run before first estimating the Dirichlet hyperparameters. We determined that 500 was an appropriate number of iterations.
- d. Random seed. We set a fixed random seed for every estimation to allow reproducibility.
- e. Number of words printed by topic: This is the number of words most likely to be printed for each topic following model estimation. We set the output to 20 words.
- f. Threshold ratio by topic: This parameter was used to prevent printing topics with proportions lower than a threshold value – i.e. the proportion of words in a document that belong to a topic. We use the default value, 5%.

Once the model estimation is complete, we can see the list of words (or bigrams in our case) grouped under each topic. Topics are numbered from 0 to n (n being the number of topics specified minus 1). We affix labels to each of the topics based on our interpretation of their content – i.e. based on the bigrams that predominate in each topic. This exercise is based on an iterative process of assessing the output of the models – varying the number of topics – to detect patterns of bigrams that are consistently grouped together, and corroborating that the grouping is adequate by exploring the content of the articles being modelled. This process also allows for properly labeling topics in which some of the bigrams are the same, yet they are used in different contexts.⁶

Mallet also estimates how prevalent each topic is within each document – i.e. each newspaper article – as part of the topic modeling routine. Specifically, it calculates the proportion of each topic in each article, which provides the basis to assess the relative importance of individual topics in the greater corpus.

4. Results

4.1. What topics emerge in the news media?

We begin with a basic overview of the number of articles published over time and by U.S. state, which is shown in Fig. 1.

Fig. 1 shows a relatively low number of articles on hydraulic fracturing published between 2007 and 2009, and a big increase between 2010 and 2012. The number of articles increases from less than 300 in 2009 to almost 1200 in 2012. After 2012, we see a slight decrease until 2015. In 2016 and 2017, this decline becomes more pronounced to levels roughly similar to those observed in 2010. The jump in attention from 2010 to 2012 is an example of the “media storms” events identified by Boydston et al [56]. The jump also is consistent with Mazur’s [11] observations of how a series of events, including the release of the film *Gasland* and the Deepwater Horizon disaster in early 2010, contributed to fueling debates about the potential economic benefits and environmental costs of hydraulic fracturing. However, we do not claim that these events solely caused this uptick of attention. Pennsylvania experienced the largest increase in media attention from 2009 to 2012. This is consistent with previous research describing how a series of events of possible contamination of drinking water from shale gas wells drew local attention and sparked protests across the state from 2009 onward [21].

Within these articles from Fig. 1, modeling uncovered a total of 13 topics present across the nine states and 11 years that we analyzed. Fig. 2 contains a depiction of the 12 most prevalent bigrams associated with each of the topics, illustrated through word clouds.⁷ The size and

color of the words reflect the more prominent bigrams in the topic word cloud (warmer colors denote higher frequency, as do larger sizes). Our labeling of the topics (“Clean Energy”, “Drilling”, etc.) reflects our understanding of the type of hydraulic fracturing-related issues the topics seem to cover (see Table A1 in the Appendix for a brief qualitative description of each topic, along with the most prominent words). Even though each bigram in each article is associated with only one topic, the same bigram (such as “environmental protection”) can be linked to more than one topic across different articles. To make sure that the associations are clear, we followed an inductive approach and assessed the content of the articles to understand the context in which the words were used.

Fig. 3 shows the aggregate number of topics that we identified by state. Based on Fig. 3, as the number of published articles increases from 2010 to 2012, the number of topics per year identified in our modeling consequently experiences an increase as well. Yet the number of aggregated topics per year across all states decreases at a slower rate than the decrease of reported stories. This might indicate that discussions and debates on hydraulic fracturing are becoming more varied (or nuanced) as newspapers engage in reporting that goes beyond covering the obvious topics of economic/fiscal benefits and environmental risks of hydraulic fracturing.

4.2. At what frequencies have the topics appeared over time and across states?

Whereas Fig. 3 showed the aggregate number of topics, thereby providing a broad perspective on the content of media attention on hydraulic fracturing, Fig. 4 shows the relative salience of each topic over time. The topics of highest importance tend to be topics such as environmental risks, local politics, and economic benefits, while others such as state regulations, electricity generation, and clean energy tend to be of less relative importance. Some variation in topic salience occurs over time, but topics that captured a solid proportion of the news media agenda tended to stay on the agenda. For example, electricity generation never captured much agenda frequency over time whereas health risk maintained more than 10% of the relative frequency of the coverage.

To highlight the relative stability and high frequency of some topics over others, we distinguish between “hot” and “cold” topics based on the relative importance of each of them (see Figs. 5 and 6). “Hot” topics are those for which at some point there is at least 10% of the words in the corpus assigned to that particular topic, and “cold” topics are those that do not reach the 10% threshold. The threshold value is obviously artificial, but it is valuable to at least distinguish topics that dominate media coverage from those that do not. However, as coverage evolves, topics might become more or less dominant. The most important lesson from Figs. 5 and 6 (which can also be gleaned from Fig. 1) is that, in general, topics tend to remain relatively stable over time in terms of their importance. However, there are some changes worth noting. Two of the “hot” topics seem to have lost some prevalence. “environmental risks” has consistently captured the lion’s share of the agenda space, but in 2017 the relative importance of the topic decreased noticeably. The second topic that seems to have cooled off in the last 4 years of our analysis is “drilling”, which was prevalent in the early years but then dropped slightly (approximately 10%–5% of the relative frequency). Some of the topics that produce less coverage across our study seem to have gained greater relative importance. The two “cold” topics where we see a more consistent upward tendency are “infrastructure” and

⁶ The process was conducted separately by two individual coders, each of whom came up with an initial set of labels describing each of the topics. In a second step, the two coders met to solve any discrepancies, and converge on a label that represents the content of the topic.

⁷ The topic “Insurance” was not included in the figure. “Insurance” appeared

(footnote continued)

only in two articles in Ohio (2007) and it was identified as a relevant topic due to the size of the Ohio corpus of that year ($n=3$). This topic did not appear in any other article for the rest of the year under consideration, either in Ohio or in any other state.

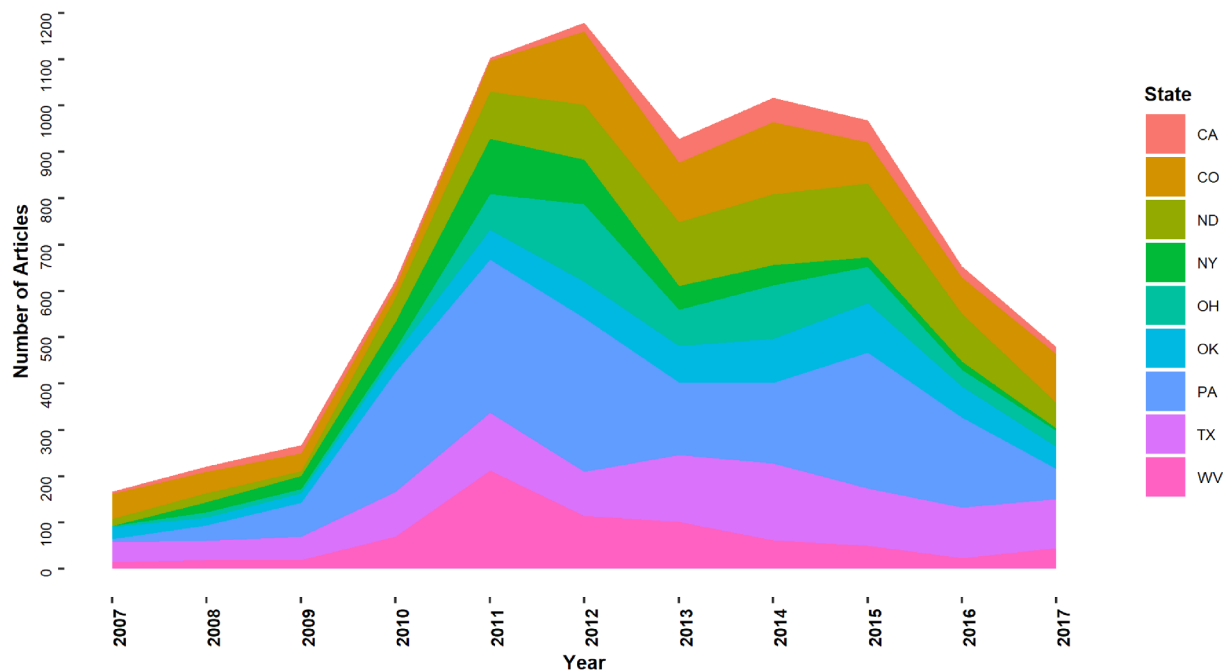


Fig. 1. Number of articles on hydraulic fracturing, per state and year (2007–2017).



Fig. 2. Word clouds for each topic associated with hydraulic fracturing.

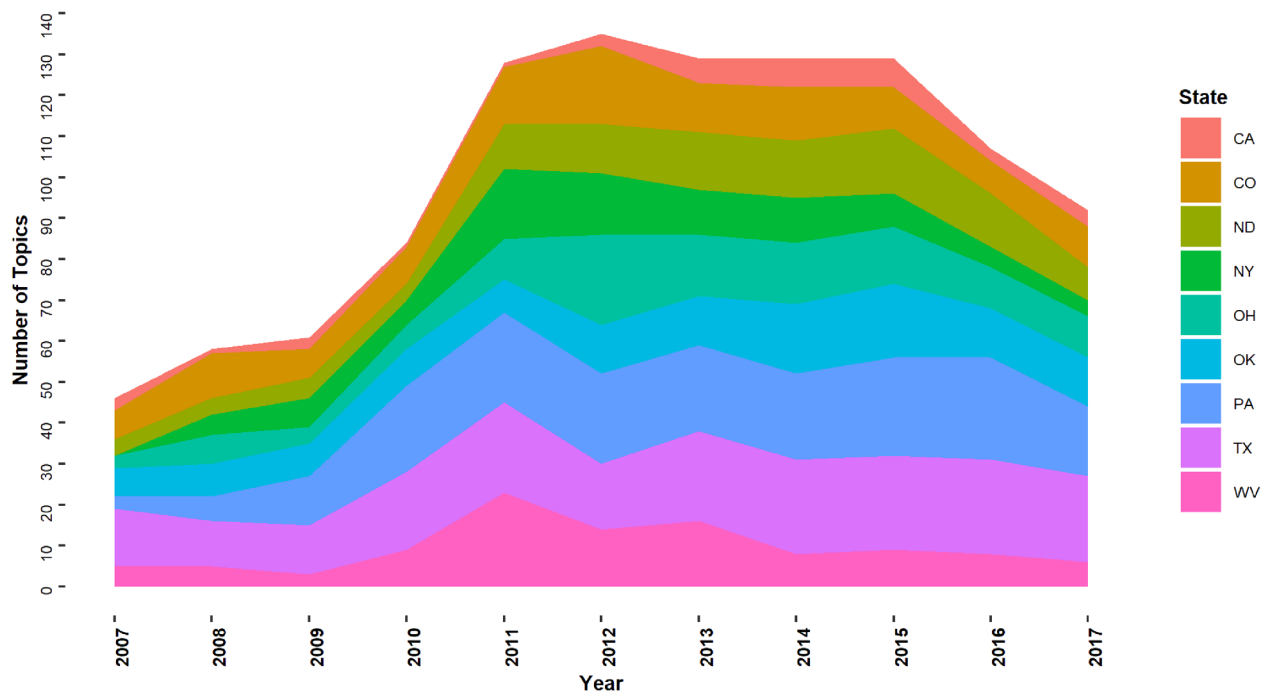


Fig. 3. Number of topics associated with hydraulic fracturing, per state and year (2007–2017).

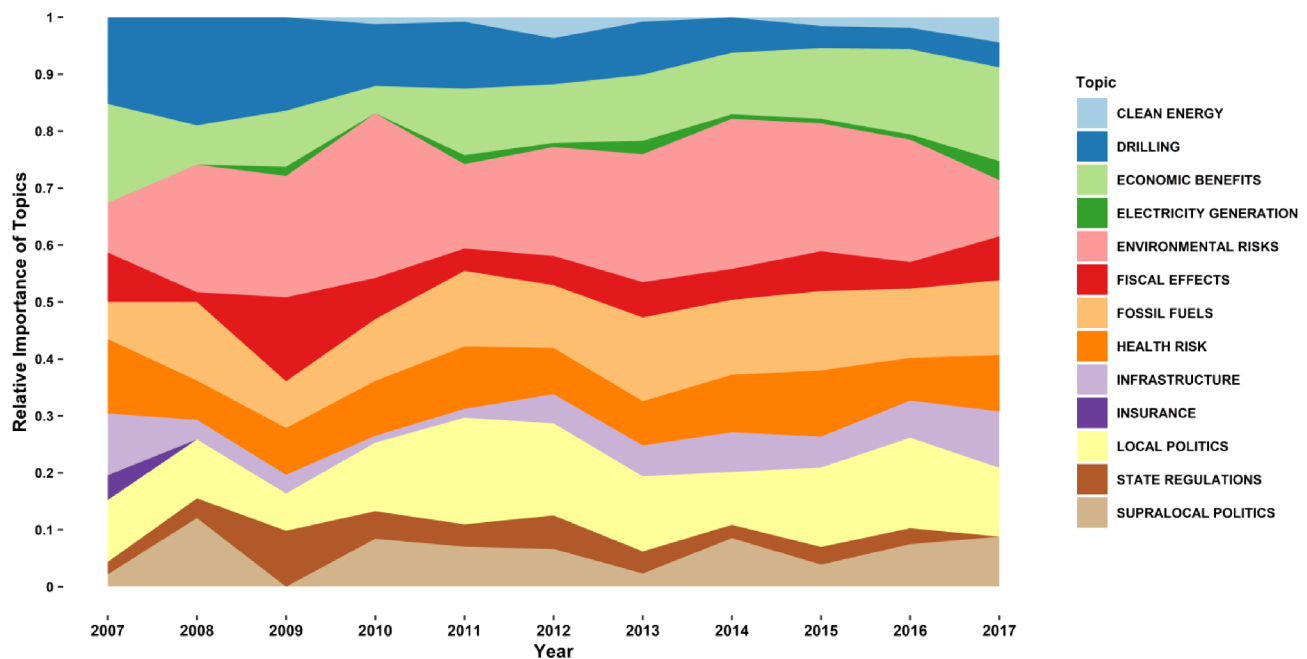


Fig. 4. Relative importance of topics associated with hydraulic fracturing, per year (2007–2017).

“clean energy”, though the absolute increase in their relative frequency is modest (i.e. they do not go over 10%). An updated analysis incorporating data from the last two years would be needed to confirm whether these topics continue to gain momentum. Regardless, these changes, while minor, demonstrate how discussions on hydraulic fracturing have evolved. In the earlier years, we see a heavier emphasis on the development of the practice and its associated environmental costs and economic benefits. In more recent times, debates have increased in topic diversity, including, for instance, the need to pivot to alternative sources of energy that are more environmentally friendly.

We see differences and similarities in the frequency of these topics across states. Individual figures for each state on evolution of topics are

contained in the appendix.

Fig. 7 shows a heat map indicating how much media attention each topic garners in each state, averaged across all years in our study. For each topic, we added the values for each year, which were then divided by the total number of years to obtain the average incidence of the topic. This yielded one unique average value for each topic in every state. The second step was building the heat map and grouping the states based on their similarity. We built a matrix of states and topics in which each cell has the average topic incidence value, and we used the R package *gplot* to build the heat map and cluster the states based on their similarity. We established similarity by estimating the Euclidean distance between each pair of states in the matrix (i.e. each vector of

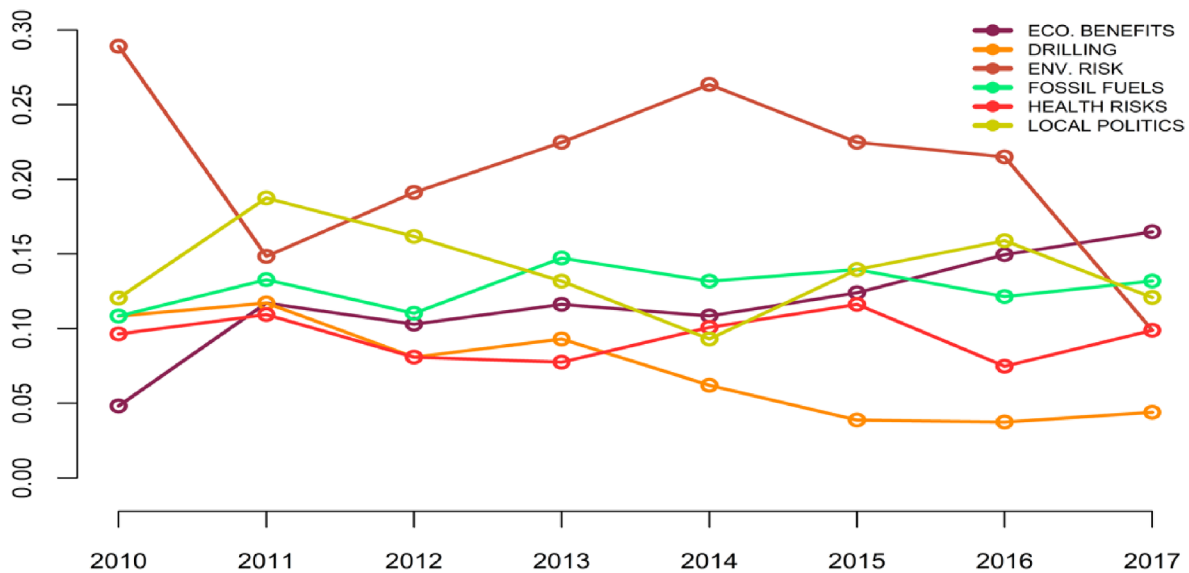


Fig. 5. "Hot" topics.

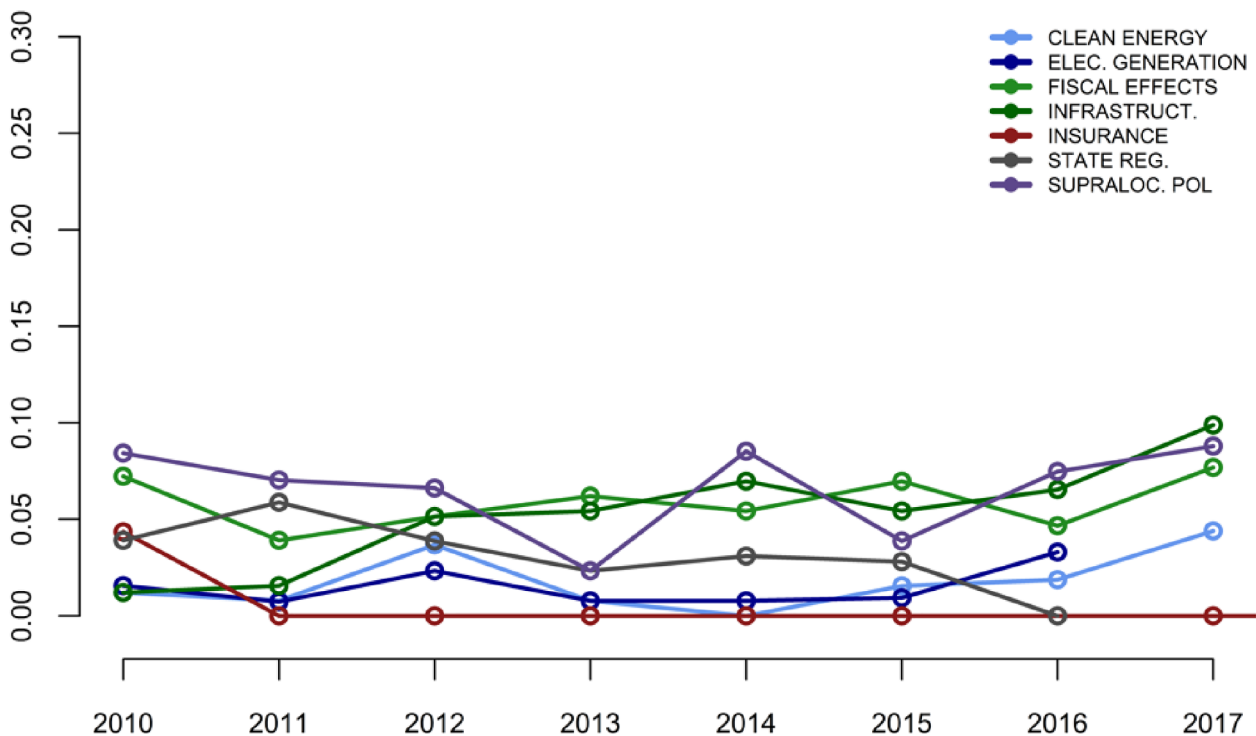


Fig. 6. "Cold" topics.

values of average topic incidence). This grouping is represented by the dendrogram on the left side of the plot. The heat map allows us to uncover patterns of overlap or differences between states based on how topics predominate in media reporting. Red indicates a topic occurs at a high frequency in a state over the time period whereas light yellow indicates a topic occurs at a low frequency.

The visualization produces some noteworthy results. First, the topic "environmental risks" tends to be prevalent across all states that we study, which is not the case for any of the other topics we identified through our modeling exercise. Environmental concerns related to hydraulic fracturing is usually well covered in media sources, and this is

the case even in states where discussions about the economic benefits of hydraulic fracturing tend to receive considerable media attention (North Dakota, Oklahoma, and Texas). This is in line with findings by Habig and Hinojosa [19] and Evensen et al. [22], who identified water pollution as a product of hydraulic fracturing as the main concern portrayed in the smaller set of articles they analyzed in their studies. Second, some topics are not highly prevalent across any of the states (i.e. they do not dominate the discussions), including "clean energy", "electricity generation", "insurance", "state regulation", and "supra-local politics". Even though these topics received attention, they were marginal in comparison to others such as "environmental risks" or

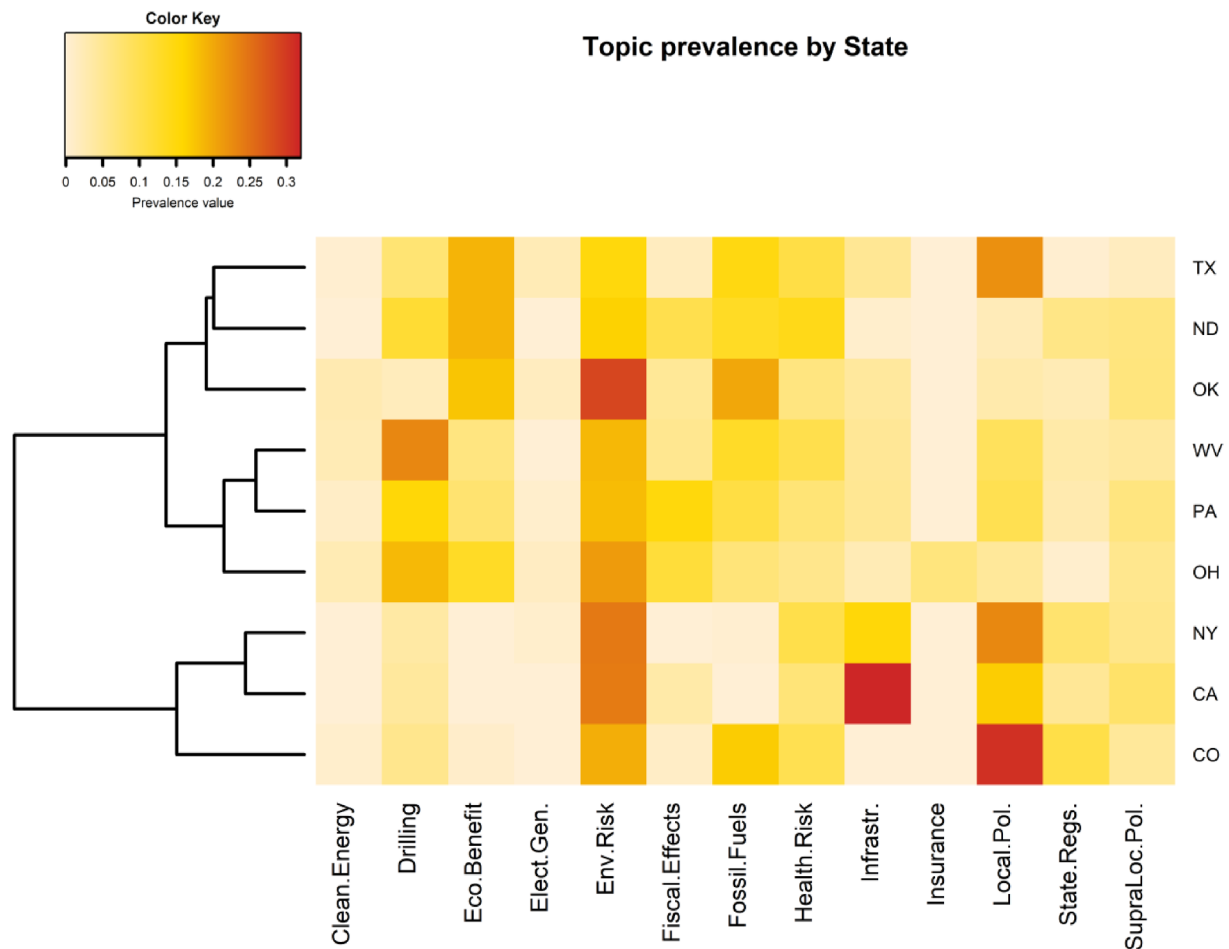


Fig. 7. Heat map of similarity/differences between states by topic.

“local politics”. We note, however, that although topics such as “state regulation” may not be prominent, it is possible that the topic modeling picks up discussions around related issues through other topics (e.g., “environmental risks” may be part of discussions around regulations in news media but the modeling separates these topics).

Also noteworthy is the clustering of states that emerged via visual inspection of common pockets of colors in Fig. 7 and the clustering via the dendrogram on the left side of it. Whereas all the states emphasize environmental risks, states tend to cluster around commonly emergent topics otherwise. The nine states are clustered in three main groups: 1) West Virginia, Ohio and Pennsylvania, 2) North Dakota, Oklahoma, and Texas, and 3) California, Colorado, and New York.

States in the first of these clusters (West Virginia, Ohio, and Pennsylvania) tend to give high coverage to drilling as compared to other states. States in the second cluster (North Dakota, Oklahoma, and Texas) tend to emphasize economic benefits as compared to other states. In the third cluster (New York, California, and Colorado), the topic “local politics” is important. Local politics also appears as important in Texas, but it was likely the high frequency of drilling and relatively low relatively frequency of environmental risks that pulled this state into the other cluster. This is probably a reflection of the growing importance of debates about the capacity of local governments to regulate hydraulic fracturing and the increased level of action by social movements opposing hydraulic fracturing, which often operate at the local level.

4.3. How do the topic frequencies vary across states based on political party governance?

To further disentangle the differences in topics that predominate in the states based on their governing party, we categorized states by political leaning to gauge differences in reporting. California, Colorado, and New York are Democratic states and North Dakota, Ohio, Oklahoma, Texas, and West Virginia are Republican states.⁸ Pennsylvania is split.

Figs. 8 and 9 confirm some of the insights observed in Fig. 7. Reporting on hydraulic fracturing in Republican states tends to emphasize discussions about “fossil fuels” and the “economic benefits” linked to unconventional oil and gas production, while the emphasis in Democratic states seems to be on “environmental risks” and “local politics”. These findings are consistent with research showing that arguments about the economic benefits of fracking tend to be more popular among Republicans than Democrats [10] and that opposition to fracking (which are more likely to appear in the topics “environmental risks” and “local politics”) are strongly related to Democratic Party

⁸ We classified states into three categories: Democratic, Republican or Split, depending on the party that held the majority of governing positions during the time period. We took into account the party of the governor for each electoral cycle as well as which party controlled each chamber of the state legislature. We classified the states as ‘Democratic’ when the Democratic Party controlled the governorship and the state legislature most of the time and ‘Republican’ when the Republican Party dominated. There was only one ‘Split’ (Pennsylvania) in which party control swings from electoral cycle to electoral cycle, without consistent control over the state.

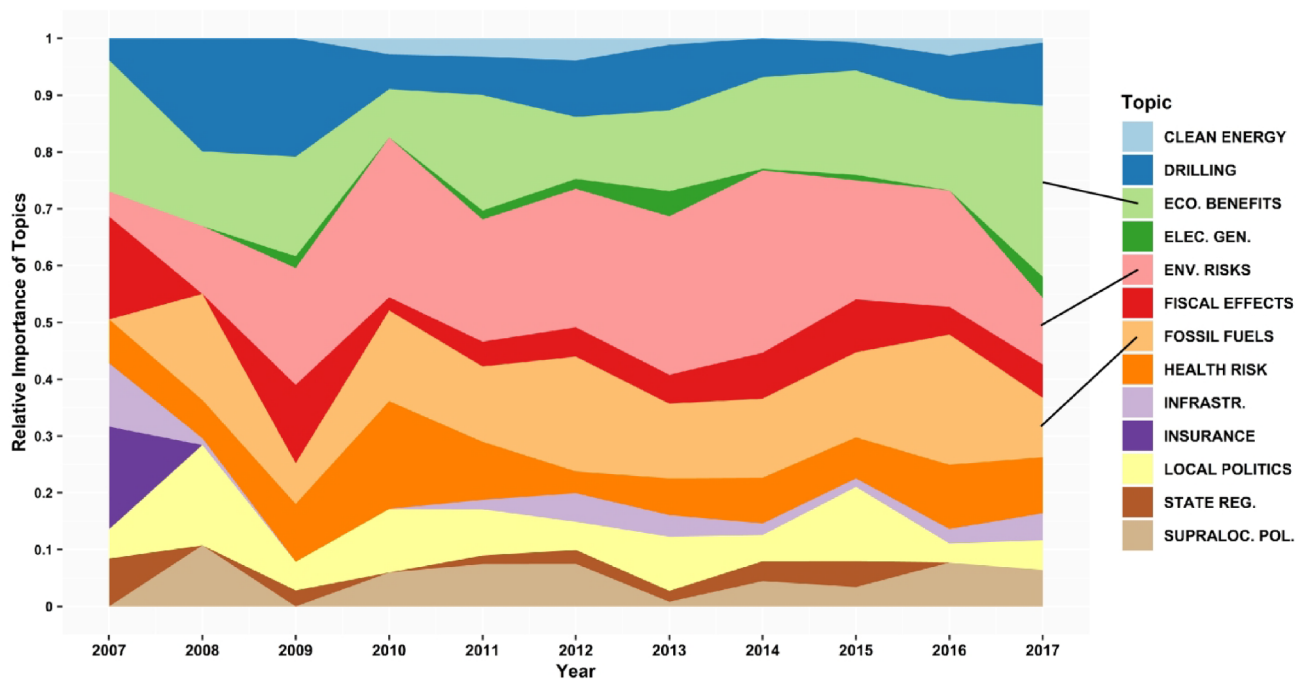


Fig. 8. Relative importance of topics associated with hydraulic fracturing, for Republican states per year (2007–2017).

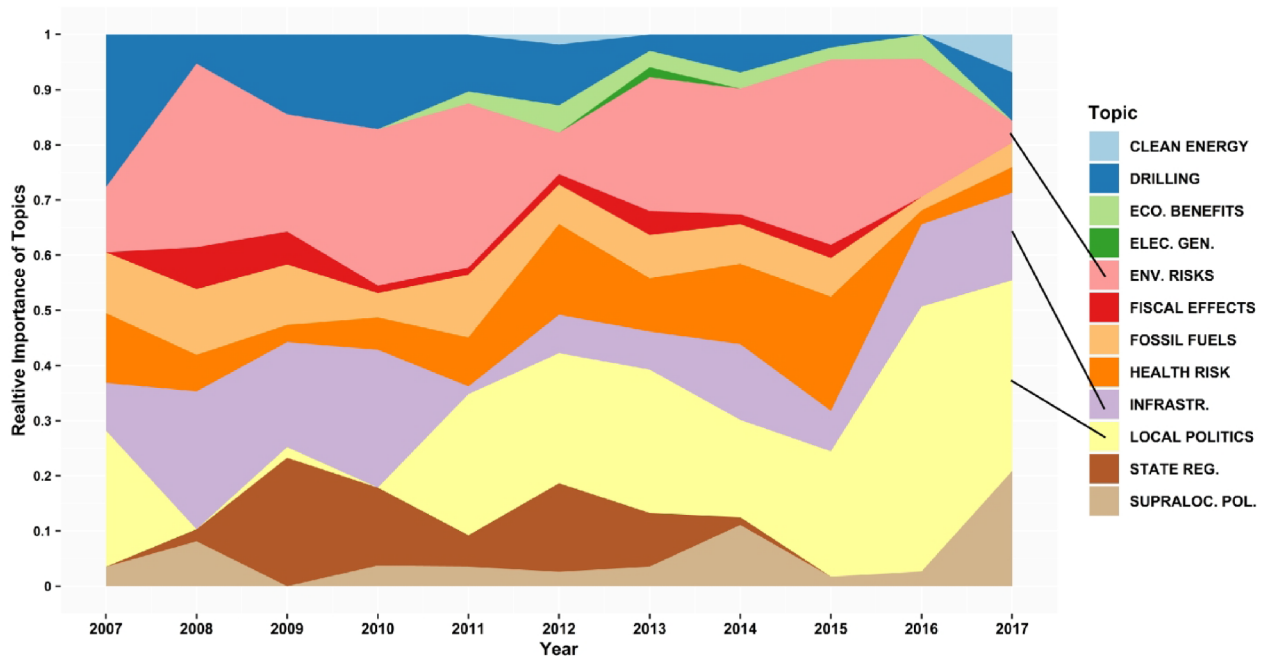


Fig. 9. Relative importance of topics associated with hydraulic fracturing, for Democratic states per year (2007–2017).

identification [57].

In addition to differences in the types of topics that are identified in Democratic versus Republican states, we are also interested in measuring the breadth of discussions – i.e. how much certain topics tend to dominate discussions in reported stories. Measuring attention diversity is important because more diverse discussions can indicate a more nuanced view of the perils and promises of hydraulic fracturing as a technology to increase oil and gas production. To measure breadth, we need to observe how widely media stories are distributed across topics. Boydston et al. [58] compare regular and normalized versions of Shannon's H Information Entropy ("Shannon's H") and Herfindahl

measures, which are used to capture attention diversity in discussions with multiple items and show that the former are a more sensitive measure to capture diversity in discussions. Shannon's H is simply calculated as follows:

$$\text{Shannon's } H = - \sum_{i=1}^n (p(x_i)) * \ln p(x_i) \quad (1)$$

where, in our case, x_i represents a topic associated with hydraulic fracturing, $p(x_i)$ is the proportion of total attention the item receives (i.e. the "relative importance of topics" as we term it in Figs. 8 and 9), and $\ln p(x_i)$ is the natural log of the proportion of attention the topic

Table 1

Testing difference in the means of Shannon's H between Republican and Democratic states.

Group	Obs	Mean	Std. Error	Std. Deviation	95% Conf. Interval
Democratic States	1678	0.249	0.004	0.169	0.240, 0.257
Republican States	4017	0.173	0.003	0.171	0.167, 0.178
Combined	5695	0.195	0.002	0.174	0.190, 0.200
Diff		0.076	0.005		0.066, 0.086
diff = mean (Democratic States) – mean (Republican States) t = 15.336					
Ho: diff = 0 degrees of freedom = 5693					
Ha: diff < 0 Ha: diff = 0 Ha: diff > 0					
Pr(T < t) = 1.0000 Pr(T > t) = 0.0000 Pr(T > t) = 0.0000					

receives. When calculated, the measure can adopt a minimum value of 0 when all the discussion is about only one topic, and then increase as the discussion includes a larger number of topics. An advantage of Shannon's H is that it accounts for the number of topics receiving attention, and as this number grows, the maximum score increases through the $\ln(N)$ portion of the formula (N is the number of topics). For instance, an article that gives equal importance to all five topics that are present in it will produce a *higher* value than one in which one of the five topics has a greater proportion of attention, but it also will produce a *lower* value than an article that assigns equal attention to six topics.

We calculate the Shannon's H for every-one of the articles on which we run our topic models, and then run a difference in means test to find out whether articles in states governed by the Republican Party are more (or less) diverse than articles published in states that are governed by the Democratic Party. We exclude Pennsylvania from this analysis since the state has been governed by both parties in the period we study. The results of the test are shown in Table 1.

The two mean values of Shannon's H for newspaper stories in Democratic and Republican states tend to be relatively low, which indicate that diversity in reporting on hydraulic fracturing is not particularly high. In other words, newspaper stories tend to focus in either a small number of topics, or disproportionately on a particular topic when a story covers multiple ones. Having said this, there is a positive and statistically significant difference in favor of stories published in states where the Democratic Party governs. This means that in the three states where this is the case (California, Colorado, and New York) discussions and debates reported in stories of hydraulic fracturing tend to be more diverse in the kinds of topics they cover. We believe this might be a reflection of the fact that hydraulic fracturing experienced greater opposition in these states by citizens who self-organized in more or less cohesive social movements opposing the widespread use of the technology, which in turn might have resulted in greater media attention to a wider variety of fracking-related topics.

4.4. To what extent do the topics associate with policy conflict?

One of our stated goals in this article is not only to describe the topics that dominate discussions about hydraulic fracturing across the U.S., but also to discern how much conflict is associated with each of the topics. To do this, we develop a new methodological approach for validly and reliably capturing conflict associated with the topics we have identified.

We proceed in the following steps.

1. First, we used a subset of articles from Colorado and Ohio to build a dictionary of conflict words and used an R script to clean the text.⁹

⁹ We developed this dictionary in an earlier stage of our research project, for

We removed punctuation marks and special characters, and removed numbers and common English stop words. In addition, we stemmed the words in the documents to identify all of the related variants of a concept. For example, "complain" and "complaining" are stemmed into "complain". This process reduced each article to a "bag of words" in which relevant words and concepts are kept and which allows for a representation of the entire corpus as a matrix of documents (rows) and the words/terms contained in them (columns) (dtm, or document-term-matrix) [52,59,60].

2. We then used a manual coding approach to develop a list ("dictionary") of conflict words. To be included in the dictionary, words must 1) suggest conflict or concord on their face ("face validity test"), and 2) be assessed by a coder as clearly indicating conflict in at least 80% of their uses ("empirical test").¹⁰ Through this process, we identified a total of 95 words that indicate conflict that met both tests according to at least one coder (see Table A2 in the Appendix for the final list of conflict words).
3. Finally, we counted how many of these conflict words appeared in each document, and then, for each topic, we correlated the number of conflict words with the proportion of that topic present in the document. Thus, more conflict words in the text (generally) and high preponderance of a topic in that text would yield high correlation between conflict and that topic.

We can then gauge how much conflict is associated with the topics in each of the states. We calculated the Pearson correlation between the number of conflict words in each article and the preponderance of each topic present in the same article. For example, to calculate the correlation between conflict and the topic "economic benefits" for West Virginia, we correlated the vector containing the number of conflict words with the vector of the incidence of the topic "economic benefits" in each individual article. This allows us to understand how the preponderance of a given topic appears more or less frequently (and intensely), and which words signal the presence of conflict. Fig. 10 contains the visual representation of the correlations.

Some of the topics are associated with greater occurrence of words indicating conflict. The one topic for which this is the case across all states is "local politics". Warner and Shapiro [61] described how the most controversy over hydraulic fracturing has been generated when local governments clash with their state governments over the need to promote the technology and the associated environmental costs that come with it. Cities across the nation have tried to implement specific requirements to be met by oil and gas companies, and in some cases have sought to ban the practice outright. These controversies are reflected in each one of the states we examine. The other two topics in which a majority of states tend to see a positive correlation with conflict words are "state regulations" and "supralocal politics."

The remaining topics are, on average, associated with lower incidence of conflict words. It is worth noting that the topic "economic benefits" is associated with lower number of conflict words in most, but not all, states. In New York and California, there is a positive association between the topic and the occurrence of words that indicate conflict, which is unsurprising given that these states have seen some of the more consistent opposition to hydraulic fracturing by environmental organizations, social movements, and policy entrepreneurs who advocate for banning the practice [37,54,62–64].

(footnote continued)

which we only had data for these two states. Given that the list of words indicating conflict that resulted from this exercise is very exhaustive, we are confident applying it to the other states we examine here.

¹⁰ Three of the authors reviewed the full list of words to identify those with face validity. Next, we divided the full lists into segments and two authors were assigned to each word to determine whether words with face validity are used in the news articles to refer to conflict or concord.

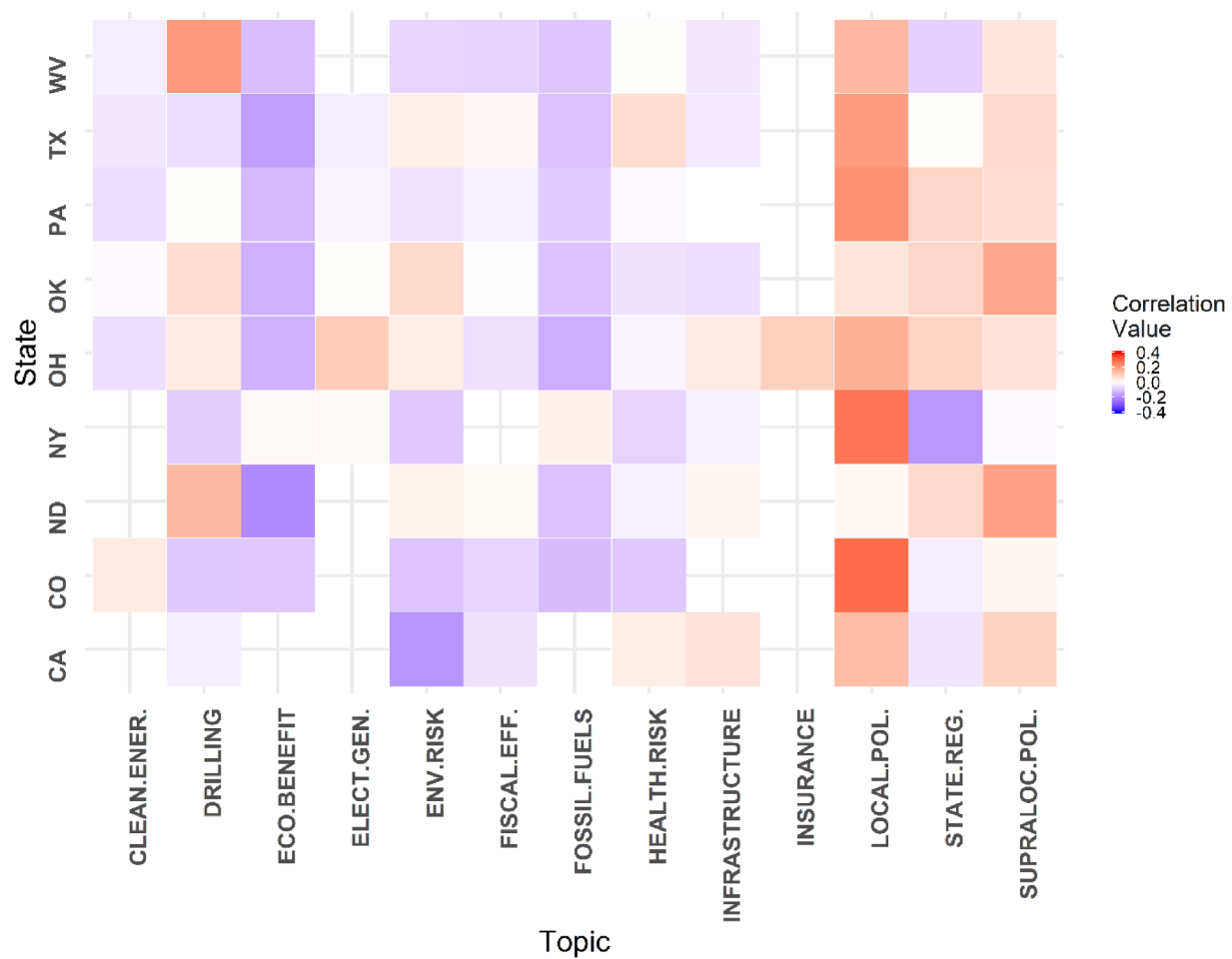


Fig. 10. Conflict per individual topic by state.

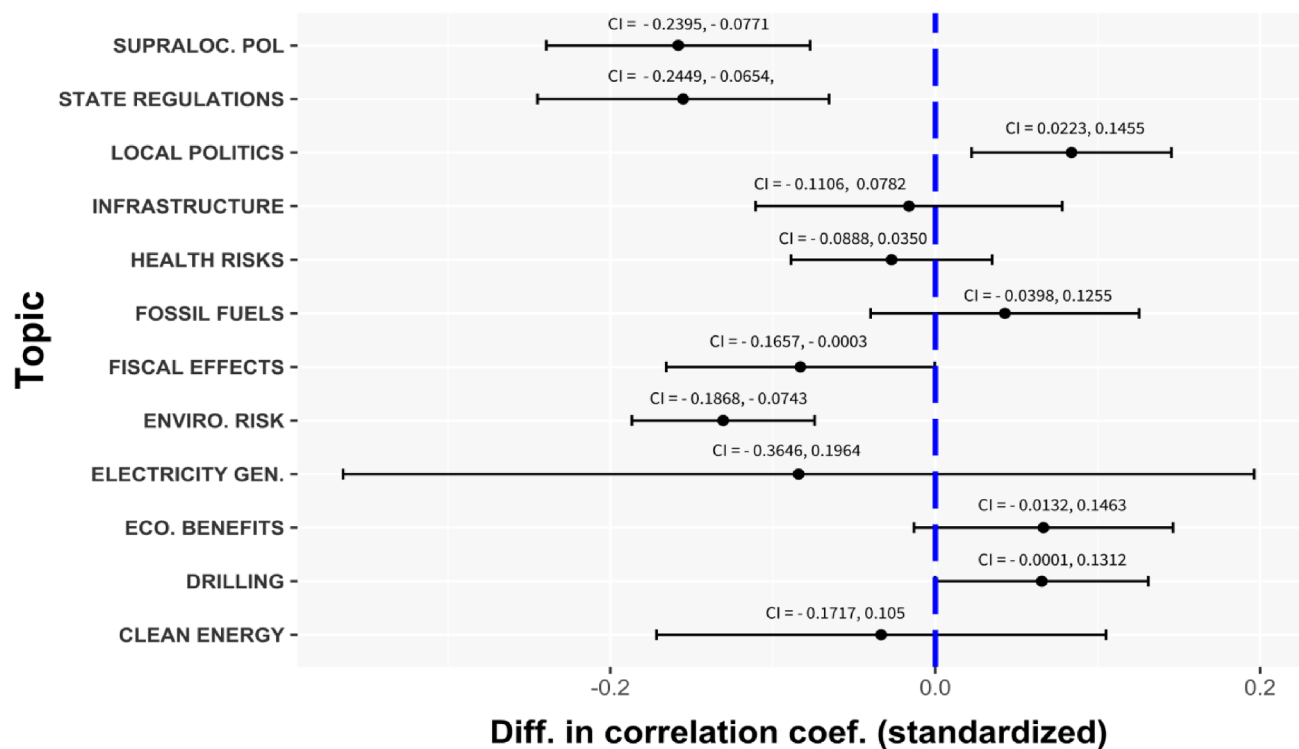


Fig. 11. Conflict and Party Affiliation of State Governments.

Finally, to assess – as we did with the diversity of topics – whether there are differences in the level of conflict associated with topics between states governed by the Democratic Party and states governed by the Republican Party, we calculated the difference of the correlation coefficients between each topic and the prevalence of conflict for each group of states. To do this, we use Fisher's *r-to-Z* transformation [65]. The appendix contains a description of the steps for calculating this measure.

Fig. 11 presents the results. Values located to the left of the blue dashed line represent the presence of higher level of conflict associated with the topic in Republican-governed states. Values to the right of the line represent greater conflict associated with the topic in Democratic-governed states. (Lines represent 95% confidence intervals.)

The results confirm inferences from the previous figures. Four topics are associated with greater levels of conflict in Republican states (“supralocal politics”, “state regulations”, “fiscal effects”, and “environmental risks”). The association between conflict and “supralocal politics” and “state regulations” is not surprising in Republican states as they tend to be more resistant to policies that might restrict the growth of industry. Yet, these states also tend to be the most active states for drilling, which likely prompt questions over fiscal issues such as severance taxes from drilling. Additionally, we speculate that conflicts over environmental risks are a function of the ability of a well-established industry in Republican states to resist such claims. Conversely, there is one topic that is associated with greater levels of conflict in Democratic states (“local politics”), with two others’ confidence intervals barely straddling the blue dashed line (“drilling” and “economic benefits”). In Democratic states, opposition to hydraulic fracturing often emerges at the local level. In New York and Colorado, for instance, local communities have often attempted to ban or pass moratoria on hydraulic fracturing.¹¹ Organized opposition to hydraulic fracturing in Democratic states have pushed for both local and state-wide policies to strictly regulate drilling practices (e.g., well siting, groundwater monitoring). This may explain the association between conflict and topics associated with drilling and the economic benefits of hydraulic fracturing that are often touted by proponents of the practice. There are several topics that show no clear association with the political leaning of any state, including clean energy, electricity generation, infrastructure, health risks, and fossil fuels.

5. Conclusion

We have introduced a new approach to measuring attention to and conflict over hydraulic fracturing in nine states across the U.S. Knowing how the media reports on this technology is particularly important given that media coverage can drive public attention to an issue [56]. While scholars have explored media coverage over hydraulic fracturing across different locations and time periods, there remains a lack of cross-state comparative research that explores media coverage longitudinally. This article offers a methodological approach that can be leveraged by scholars to access and examine large quantities of textual data in a variety of policy arenas comparatively and longitudinally.

We have uncovered notable patterns and trends in the types of topics that drive media coverage on hydraulic fracturing across the nine states we analyzed. First, we found that several topics have staying power and tend to maintain regular media coverage both over time and across states. One possible explanation for these trends may be that media outlets emphasize simple messages about the benefits and costs of hydraulic fracturing that defenders and opponents aim to showcase. However, even in the presence of this relative stability, certain topics

seem to be gaining greater amounts of attention, and others decreasing in importance in the last two years of our analysis. Future research should aim to confirm whether these changes are sustained as discussions over hydraulic fracturing mature, or whether they are simple temporal deviations from the rather stable discussions we have observed Table A3.

There are significant differences in the conflict levels associated with individual topics based on whether the states in which the news are produced are politically dominated by the Republican or the Democratic party, which is a reliable proxy for the ideological leanings of most of the voting population. For instance, in the three more Democratic states in our sample (California, Colorado, and New York), reports of hydraulic fracturing tend to cover a wider range of topics, which might result from more nuanced discussions about the technology that have sprung from the heavier opposition from non-governmental actors in these states. Regarding levels of conflict associated with topics, we describe how certain topics (“supralocal politics”, “state regulations”, “fiscal effects”, and “environmental risks”) are linked to greater conflict in Republican-governed states, while others (“local politics”, and to a lesser extent “drilling”, and “economic benefits”) see more conflict in Democratic states. These patterns are somewhat expected given the differing views about the costs and benefits of hydraulic fracturing that are likely to exist in states where one party has political dominance.

We introduce a new approach to examining the content of reporting about hydraulic fracturing that can speed up comparative analysis in this subject-area. Yet, interpreting the results from this approach and drawing implications should be done with caution, given certain limitations of our work. For example, previous analyses of newspaper coverage of hydraulic fracturing have demonstrated that media outlets often emphasize the negative impacts of the technology, such as the environmental risks associated with it [22,66]. This potentially tempers our finding of more conflicts around the topic of environmental risks of hydraulic fracturing. Future research is needed to offer better estimates of the variation in risks and benefits of hydraulic fracturing via different data sources. Additionally, all complex issues can be partitioned into simplified depictions to aid understanding and communication; such partitioning reflects both the analytical lens of the approach itself and interpretation of the researcher. We used a transparent and systematic methodological approach to identify topics; yet, our approach also reflects the built-in assumptions of LDA modeling as well as our interpretations of the results of this modeling exercise. Our concern about this limitation can be lessened only if readers of our work find our decisions reasonable.

One of the contributions of this study is its extensive coverage of news media across time and space. That is, we collected data from two newspapers in nine states with eleven years of coverage. This overcomes the limitations of past research that focused mostly on local media [18]. Although no single study can do it all, additional research should compare the newspapers across states and local areas where hydraulic fracturing is being discussed or implemented. Similarly, our goal in this article was to understand and describe patterns in how public discussions about hydraulic fracturing take place, and how those patterns evolved. More research is needed to explain variation in the way news are reported using more explanatory variables than we incorporated in this study. For instance, it may be interesting to know how the ideological leanings of corporations that own the newspapers affect the tenor of the stories that are reported.

This article shows that the insights from an automated analysis of media content can inform our understanding of policy-relevant discussions and societal conflicts. Harnessing this power can help researchers provide fuller descriptions about the trends across space and time in these discussions.

¹¹ New York eventually banned hydraulic fracturing at the state level, with the state’s Supreme Court voting to uphold local bans in 2014; Colorado’s Supreme Court ruled in favor of state preemption in 2016, thus rendering local bans unconstitutional.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Acknowledgements

This work was supported by the United States National Science Foundation, Decision, Risk, and Management Sciences Program (grant SES-1734310 and SES-1734294).

Appendix

Fisher's r -to- Z transformation

To obtain a score for the transformation, we first calculated

$$Z = \frac{1}{2}(\ln(1+r) - \ln(1-r)) \quad (A1)$$

where r is the observed correlation coefficient between two variables. Next, we performed a test of significance for the difference between two correlation coefficients in which the test statistic z [67] is calculated as follows:

$$z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \quad (A2)$$

where Z_1 and Z_2 are the two Z transformed correlations that are being compared and n_1 and n_2 are the size of the two groups the correlations are based on.

To calculate the confidence intervals of the difference, we used Zou's procedure [68]. This produces the confidence interval of the difference between the two correlation coefficients r_1 and r_2 . The lower and upper bound for the interval (L and U , respectively) is given by:

$$L = r_1 - r_2 - \sqrt{(r_1 - l_1)^2 + (u_2 - r_2)^2}, \text{ and} \quad (A3)$$

$$U = r_1 - r_2 + \sqrt{(u_1 - r_1)^2 + (r_2 - l_2)^2} \quad (A4)$$

where the lower and upper bound for the confidence interval of r_1 (l_1 and u_1) and r_2 (l_2 and u_2) are calculated as:

$$l = \frac{\exp(2l') - 1}{\exp(2l') + 1}, \text{ and} \quad (A6)$$

$$u = \frac{\exp(2u') - 1}{\exp(2u') + 1}$$

And where

$$l', u' = Z \pm z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n-3}} \quad (A7)$$

The α simply denotes the desired alpha level of the confidence interval, whereas n denotes the size of the group the correlation is based on. We calculate the CI with an α of 95%.

Figs. A1–A9.

Table A1

Number of articles per state per year.

		Year										
		2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
State	CA	5	11	18	15	7	19	51	52	48	24	16
	CO	54	47	39	23	67	158	128	156	89	79	105
	NY	0	22	29	59	120	96	51	44	21	19	6
	ND	15	19	10	52	101	119	138	153	159	102	55
	OK	25	17	20	39	64	79	78	96	107	67	48
	OH	3	11	9	11	77	167	79	115	78	35	34
	PA	7	33	74	258	331	331	157	173	293	194	65
	TX	44	42	49	96	125	95	144	166	124	110	106
	WV	14	19	20	70	212	115	102	62	50	23	45

Table A2

Description of topics and most common terms in them.

Clean energy	Discussions associated to zero or low-carbon emission energy sources (solar, wind), industries and energy development; or alternative and low-carbon emission transportation systems.	Common terms/phrases: wind energy, solar energy, low carbon, green energy, renewable energy, solar credits, renewable mandates, alternative energy.
Drilling	Issues around drilling activity and access to the land to drill. This topic captures discussions around drilling activity without further treatment of specific externalities or concerns.	Common terms/phrases: drilling in forest, national forest, public lands, mineral rights, land property, drilling in property, conservation fund, state park, park service, private owners.
Economic benefits	Discussions associated to economic development or positive economic externalities due to fracking activity. This category does not include fiscal benefits due to fracking, as these topics were aggregated under different labels	Common terms/phrases: economic boom, economic growth, jobs, million dollars profit, steel plants, industrial development, cubic feet, million barrels, oil and gas prices.
Electricity generation	Issues associated with electricity generation, its cost, sources and availability. This topic was largely influenced by a seasonality effect, especially in those state characterized by colder climates.	Common terms/phrases: electricity generation, power plants, nuclear power, nuclear plants, power prices, energy committee, utilities, electric bill, winter, electricity availability, network resilience.
Environmental risks	Issues associated to the environmental risks and potential consequences of fracking activities and fossil fuels extraction and use, beyond just hydraulic fracturing. The discussions went from land, air and water contamination, wildlife loss and impacts over the ecosystem, to greenhouse gas emissions and climate change.	Common terms/phrases: environmental protection, environmental degradation, environmental consequences, wildlife, greenhouse gases, emissions, climate change, environmental groups, water quality.
Fiscal effects	Discussions about budget (at a local or state level) and taxes.	Common terms/phrases: taxes, budget, severance tax, oil and gas tax, tax collection, impact fees, sales tax, tax rate
Fossil fuels	Discussions about fossil fuels development and price, but not related to corporate or ownership profits. This label is associated to technical aspects in development, oil/ gas prices and their impact on power and consumers and coal (as a non-hydraulic fracturing related industry)	Common terms/phrases: oil and gas development, coal production, coal mines, gas prices, energy industry, wells drilled, petroleum association, million cubic feet, processing plants, million barrels, production
Health risks	Issues associated to the health risks and potential consequences of fracking activities and fossil fuels extraction and use, beyond just hydraulic fracturing.	Common terms/phrases: chemical disclosure, drinking water, proximity to fracking site, health concerns, wastewater management, brine, radioactive waste, waste management, health effects, water quality, air quality.
Infrastructure	Discussions associated to infrastructural developments related to the hydraulic fracturing process (in general), pipelines and offshore drilling	Common terms/phrases: construction of pipeline, offshore drilling, regulatory commission, pipeline company, offshore platform.
Local politics	Issues and topics at the local level. Most of the discussions were about the extent to which local governments could regulate hydraulic fracturing within their jurisdiction, whether to pause or ban fracking-related activities to take place in their jurisdiction and how to cope with local externalities of fracking.	Common terms/phrases: local control, local ban, referendum, local impacts, city permits, drilling permits, local drilling
State regulations	Discussions associated to the development and effects of state regulations over fracking activities, and oil and gas in general. This set of regulations covered oil and gas development, drilling regulations, safety, environmental protections and standards, emissions, among others.	Common terms/phrases: oil and gas regulations, drilling regulations, drilling standards, state standards, state control, disposal regulations, waste management, drilling guidelines
Supralocal politics	Discussions associated to policy issues and topics at the state and federal level. This label gathers a wide range of issues, from polling during state and federal elections in which fracking was part of the policy discussions, state and federal energy policy, among others.	Common terms/phrases: state elections, state committee, federal elections, federal policy, state policy, energy policy, state permits, FERC, legislature, senate, house of representatives, senator, representative

Table A3

List of conflict words.

Accus	Disappoint	Rebuk
Adversari	Disput	Rebut
Anger	Fear	Refus
Argu	Frustrat	Reject
Argument	Heighten	Resist
Arrest	Inadequ	Retribut
Assert	Inappropri	Revers
Attack	Injunct	Rival
Bash	Intimid	Sanction
Battl	Invad	Scare
Blame	Judgment	Schism
Blindsid	Lawsuit	Scof
Challeng	Litig	Scoff
Claptrap	Mislead	Scream
Clash	Moratorium	Scrutin
Complain	Neglig	Shirk
Complaint	Object	Shout
Conflict	Oppon	Sue
Contend	Oppos	Suit
Contenti	Opposit	Threat
Contest	Outcri	Threaten
Contradict	Overturn	Trespass
Controversi	Petition	Unaccept
Countermeasur	Pitfal	Unfair
Counterpoint	Plaintiff	Unlaw
Court	Plead	Upset
Critic	Pounc	Vandal

Table A3 (continued)

Accus	Disappoint	Rebuk
Debat	Pretext	Veto
Defeat	Problematic	Violat
Defend	Protest	Vocal
Denounc	Provok	Wrongdo
Disagre	Radic	

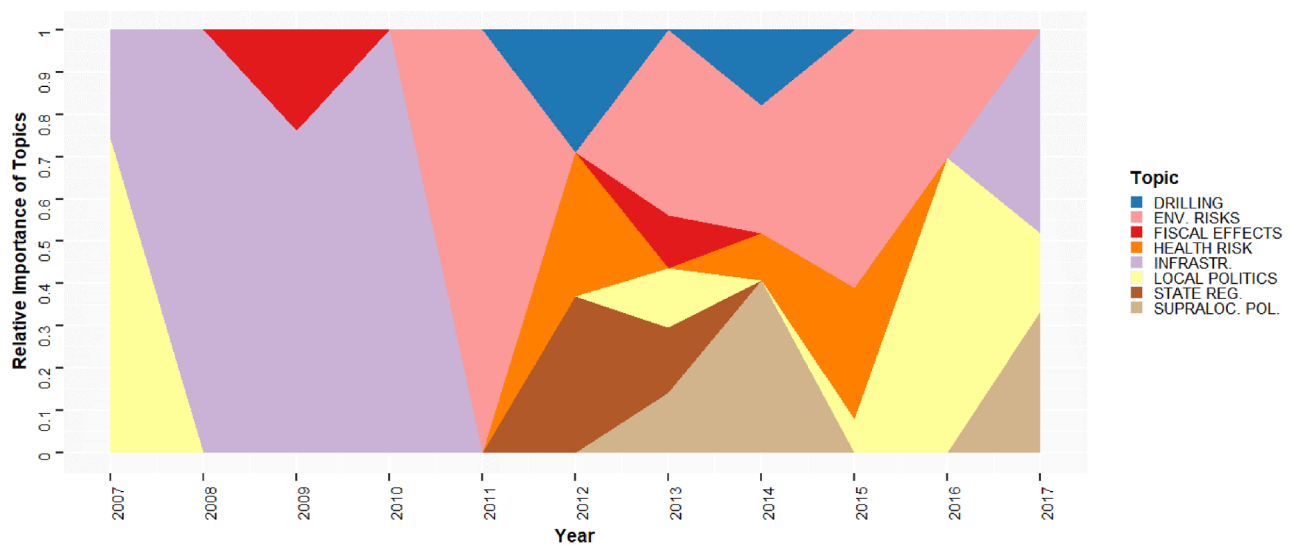


Fig. A1. Relative importance of topics associated with hydraulic fracturing in California.

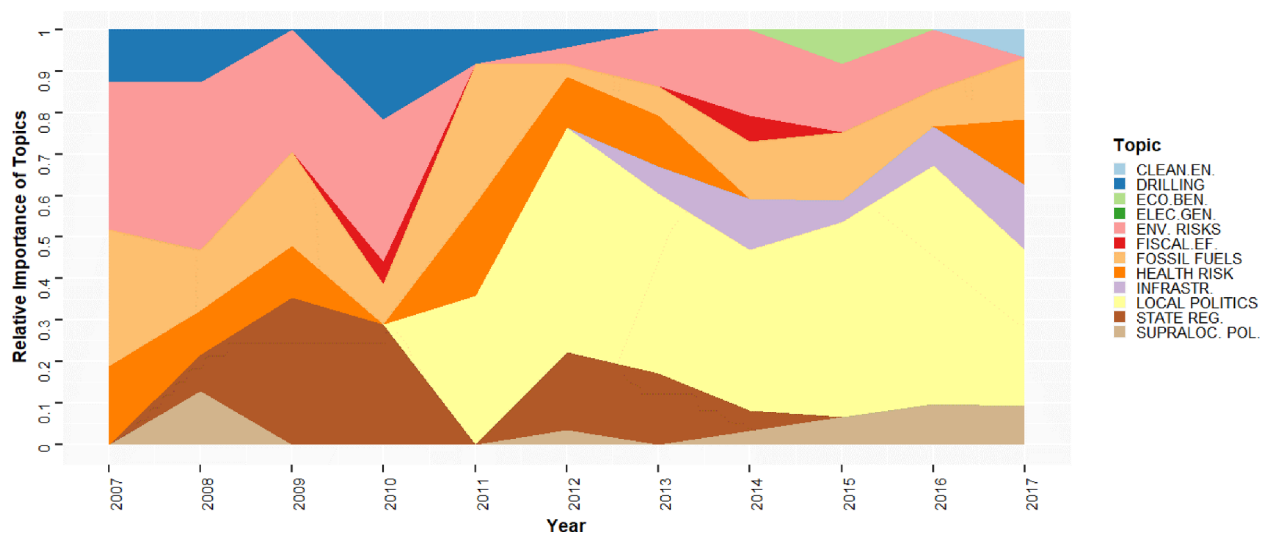


Fig. A2. Relative importance of topics associated with hydraulic fracturing in Colorado.

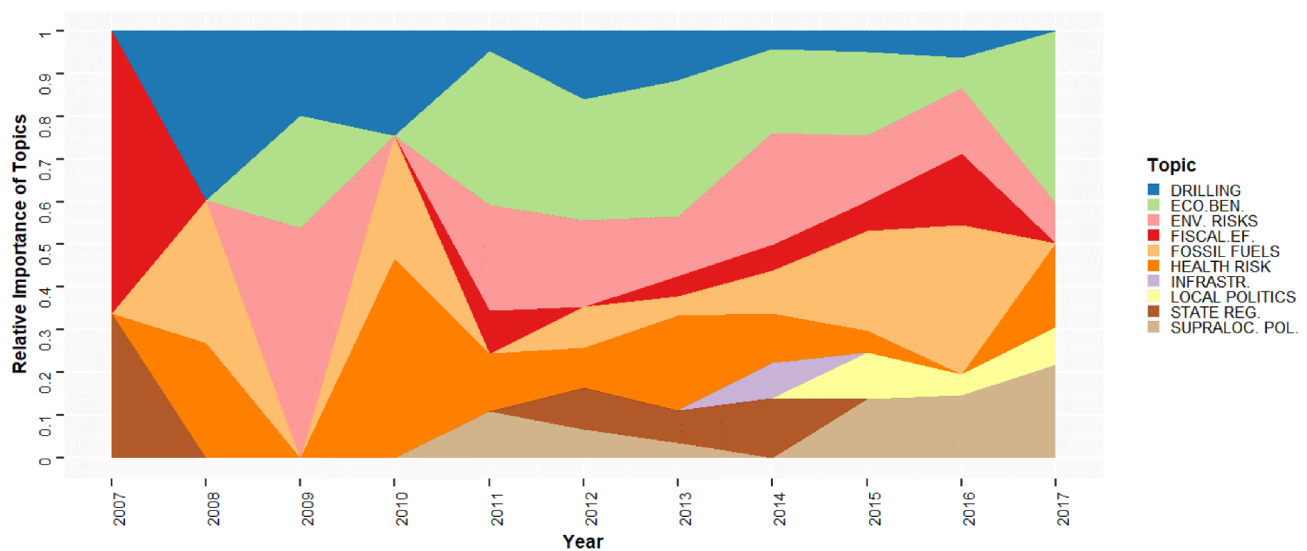


Fig. A3. Relative importance of topics associated with hydraulic fracturing in N. Dakota.

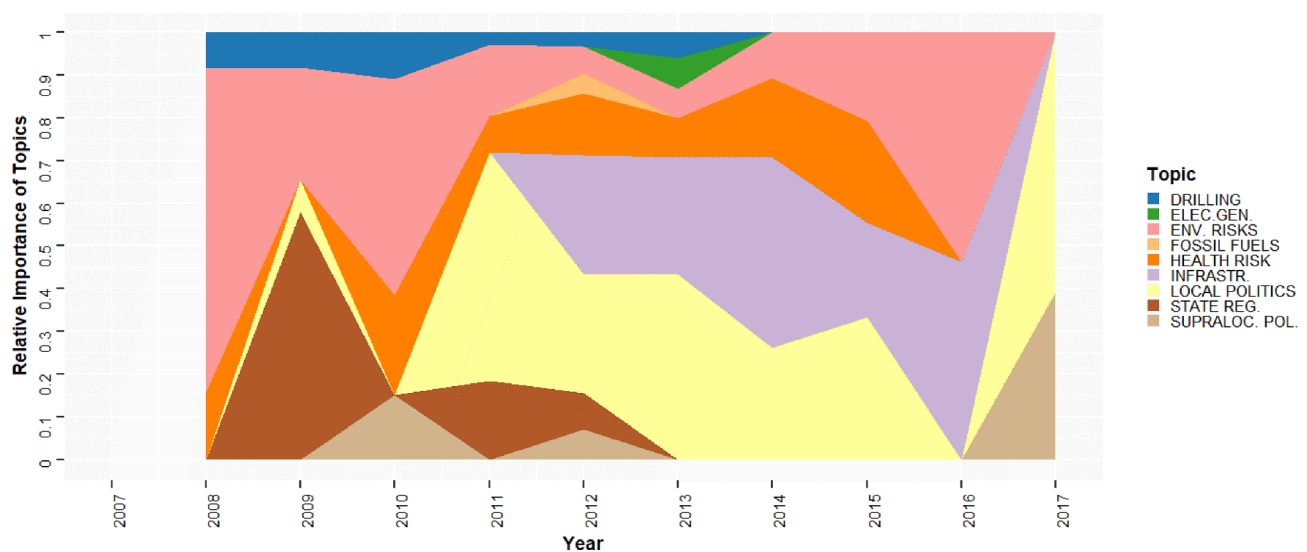


Fig. A4. Relative importance of topics associated with hydraulic fracturing in New York.

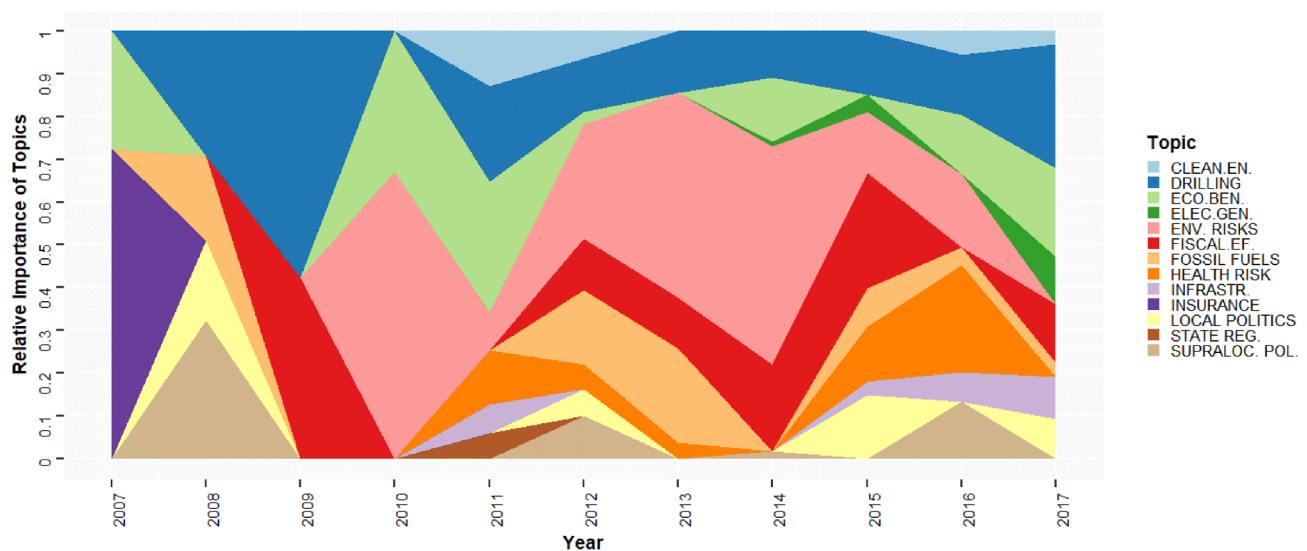


Fig. A5. Relative importance of topics associated with hydraulic fracturing in Ohio.

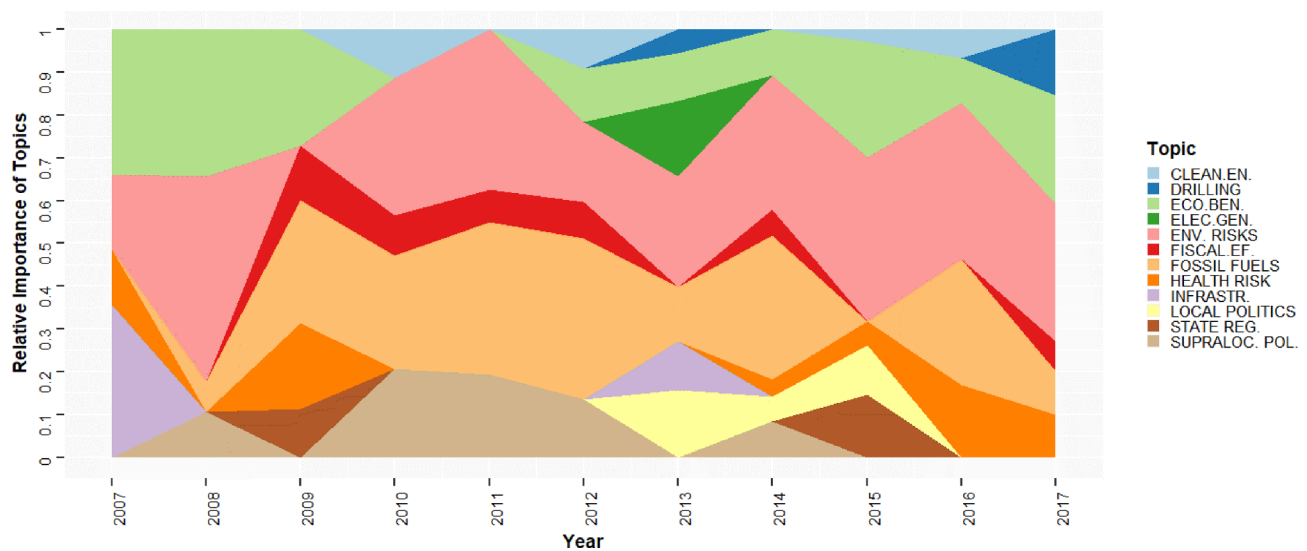


Fig. A6. Relative importance of topics associated with hydraulic fracturing in Oklahoma.

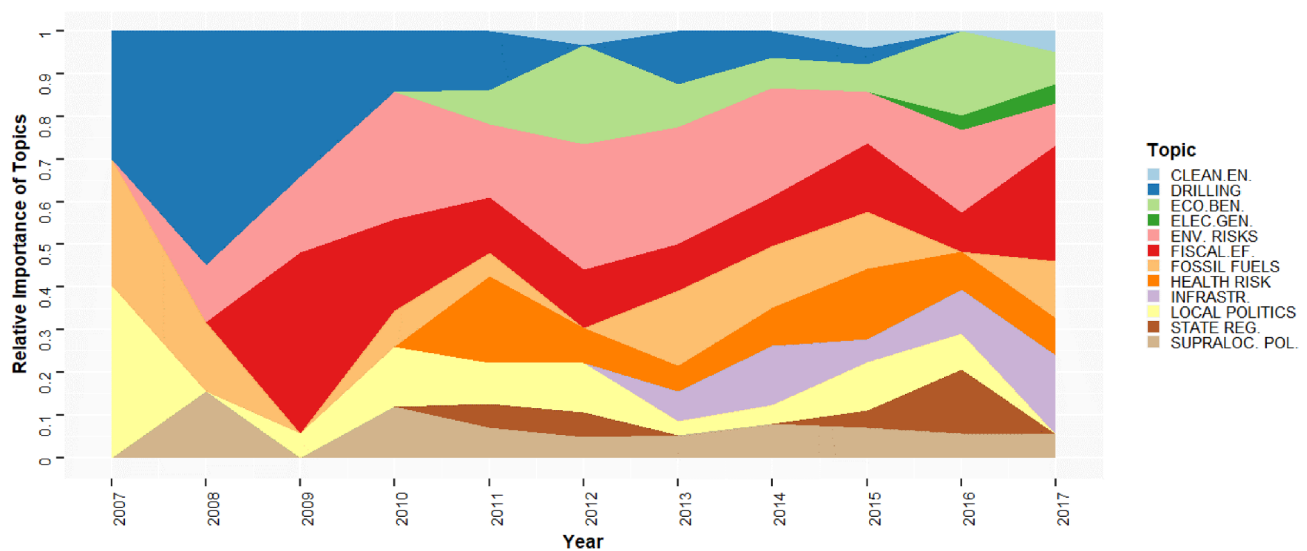


Fig. A7. Relative importance of topics associated with hydraulic fracturing in Pennsylvania.

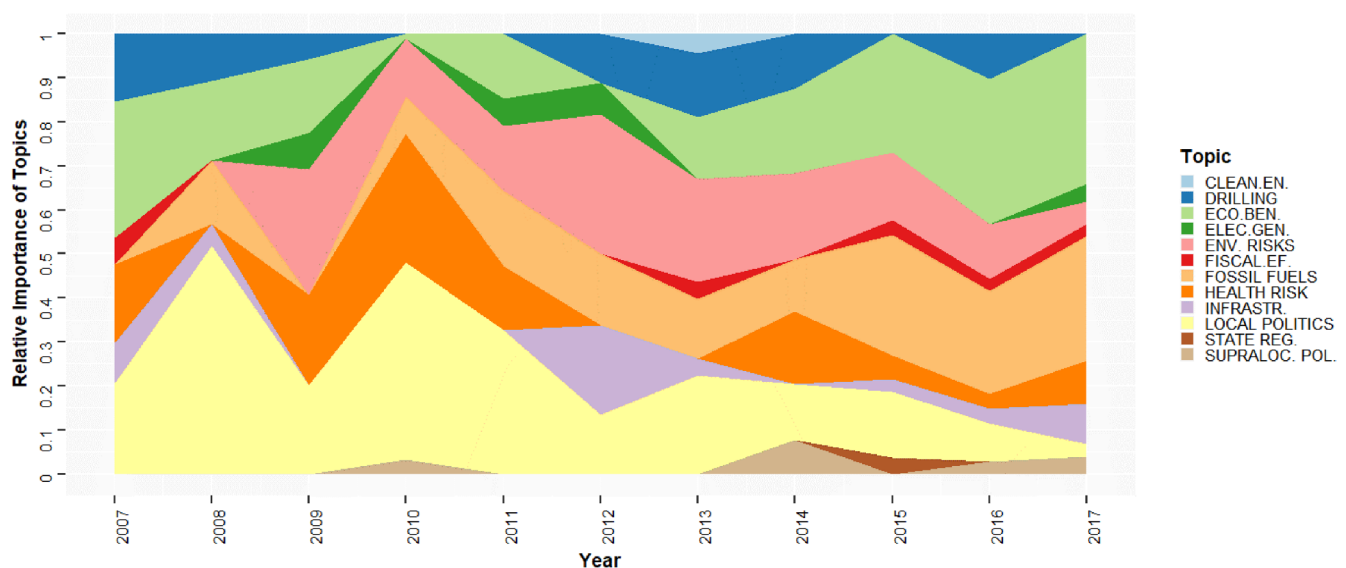


Fig. A8. Relative importance of topics associated with hydraulic fracturing in Texas.

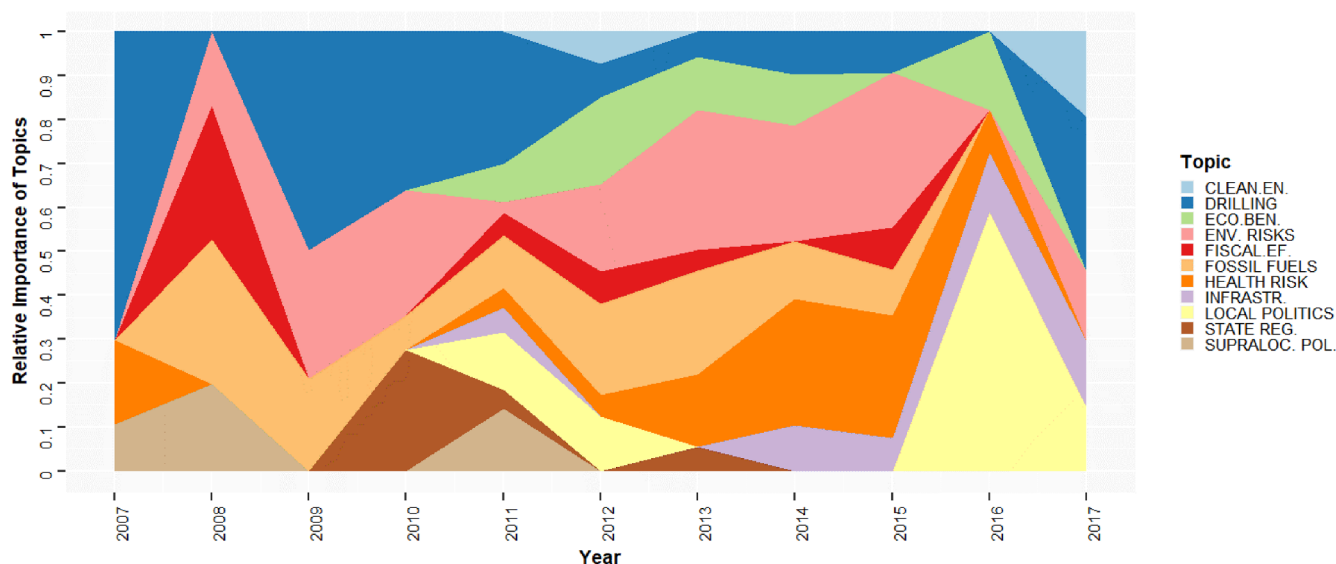


Fig. A9. Relative importance of topics associated with hydraulic fracturing in West Virginia.

References

- [1] H. Boudet, et al., "Fracking" controversy and communication: using national survey data to understand public perceptions of hydraulic fracturing, *Energy Policy* 65 (2014) 57–67.
- [2] J.B. Jacquet, et al., A decade of Marcellus Shale: Impacts to people, policy, and culture from 2008 to 2018 in the Greater Mid-Atlantic region of the United States, *Extr. Ind. Soc.* 5 (2018) 596–609.
- [3] J.L. Adgate, B.D. Goldstein, L.M. McKenzie, Potential public health hazards, exposures and health effects from unconventional natural gas development, *Environ. Sci. Technol.* 48 (2014) 8307–8320.
- [4] K.J. Brasier, et al., Residents' perceptions of community and environmental impacts from development of natural gas in the Marcellus Shale: A comparison of Pennsylvania and New York cases, *J. Rural Soc. Sci.* 26 (2011) 32–61.
- [5] C. Davis, J.M. Fisk, Mitigating risks from fracking-related earthquakes: Assessing state regulatory decisions, *Soc. Nat. Resour.* 30 (2017) 1009–1025.
- [6] A.R. Ingraffea, M.T. Wells, R.L. Santoro, S.B. Shonkoff, Assessment and risk analysis of casing and cement impairment in oil and gas wells in Pennsylvania, 2000–2012, *Proc. Natl. Acad. Sci.* 111 (2014) 10955–10960.
- [7] S.C. Latta, L.C. Marshall, M.W. Frantz, J.D. Toms, Evidence from two shale regions that a riparian songbird accumulates metals associated with hydraulic fracturing, *Ecosphere* 6 (2015) art144.
- [8] S.G. Osborn, A. Vengosh, N.R. Warner, R.B. Jackson, Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing, *Proc. Natl. Acad. Sci.* 108 (2011) 8172–8176.
- [9] T. Sangaramoorthy, et al., Place-based perceptions of the impacts of fracking along the Marcellus Shale, *Soc. Sci. Med.* 151 (2016) 27–37.
- [10] D.P. Christenson, J.L. Goldfarb, D.L. Kriner, Costs, benefits, and the malleability of public support for "Fracking", *Energy Policy* 105 (2017) 407–417.
- [11] A. Mazur, How did the fracking controversy emerge in the period 2010–2012? *Public Underst. Sci.* 25 (2016) 207–222.
- [12] Energy Information Administration. U.S. Crude Oil and Natural Gas Proved Reserves, Year-End 2018. 51 <https://www.eia.gov/naturalgas/crudeoilreserves/pdf/usreserves.pdf> (2019).
- [13] E.E. Schattschneider, *The Semi-Sovereign People: A Realist backlash's View of Democracy in America*, Holt, Rinehart and Winston, 1960.
- [14] G. Arnold, L.A. Nguyen Long, M. Gottlieb, Social networks and policy entrepreneurship: How relationships shape municipal decision making about high-volume hydraulic fracturing, *Policy Stud. J.* 45 (2017) 414–441.
- [15] H. Yi, J.T. Scholz, Policy networks in complex governance subsystems: Observing and comparing hyperlink, media, and partnership networks, *Policy Stud. J.* 44 (2016) 248–279.
- [16] T. Heikkilä, R. Berardo, C.M. Weible, H. Yi, A comparative view of advocacy coalitions: Exploring shale development politics in the United States, Argentina, and China, *J. Comp. Policy Anal. Res. Pract.* 21 (2019) 151–166.
- [17] B.D. Blair, L. McCormack, Applying the narrative policy framework to the issues surrounding hydraulic fracturing within the news media: A research note, *Res. Polit.* 3 (2016) 2053168016628334.
- [18] M. Gottlieb, E. Bertone Oehninger, G. Arnold, "No Fracking Way" vs. "Drill Baby Drill": A restructuring of who is pitted against whom in the narrative policy framework, *Policy Stud. J.* 46 (2018) 798–827.
- [19] S. Habib, M.S. Hinojosa, Representation of fracking in mainstream American newspapers, *Environ. Pract.* 18 (2016) 83–93.
- [20] A. Olive, A.B. Delshad, Fracking and framing: A comparative analysis of media coverage of hydraulic fracturing in Canadian and US newspapers, *Environ. Commun.* 11 (2017) 784–799.
- [21] O. Ashmoore, D. Evensen, C. Clarke, J. Krakower, J. Simon, Regional newspaper coverage of shale gas development across Ohio, New York, and Pennsylvania: Similarities, differences, and lessons, *Energy Res. Soc. Sci.* 11 (2016) 119–132.
- [22] D.T. Evensen, C.E. Clarke, R.C. Stedman, A New York or Pennsylvania state of mind: Social representations in newspaper coverage of gas development in the Marcellus Shale, *J. Environ. Stud. Sci.* 4 (2014) 65–77.
- [23] A. Downs, Up and down with ecology-the issue-attention cycle, *Public Int.* 38 (1972).
- [24] F.R. Baumgartner, B.D. Jones, Agenda dynamics and policy subsystems, *J. Polit.* 53 (1991) 1044–1074.
- [25] B.D. Jones, F.R. Baumgartner, *The politics of attention: How government prioritizes problems*, University of Chicago Press, 2005.
- [26] C.M. Weible, An advocacy coalition framework approach to stakeholder analysis: Understanding the political context of California marine protected area policy, *J. Public Adm. Res. Theory* 17 (2007) 95–117.
- [27] P.A. Sabatier, An advocacy coalition framework of policy change and the role of policy-oriented learning therein, *Policy Sci.* 21 (1988) 129–168.
- [28] C.M. Weible, P.A. Sabatier, Coalitions, science, and belief change: Comparing adversarial and collaborative policy subsystems, *Policy Stud. J.* 37 (2009) 195–212.
- [29] S. Iyengar, G. Sood, Y. Lelkes, Affect, not ideology a social identity perspective on polarization, *Public Opin. Q.* 76 (2012) 405–431.
- [30] R.A. Clawson, Z.M. Oxley, *Public Opinion: Democratic Ideals, Democratic Practice*, CQ Press, 2017.
- [31] H. Jenkins-Smith, D. Nohrstedt, C. Weible, P.A. Sabatier, The advocacy coalition framework: Foundations, evolution, and ongoing research, *Theor. Policy Process* 3 (2014).
- [32] C.M. Weible, T. Heikkilä, Policy conflict framework, *Policy Sci.* 50 (2017) 23–40.
- [33] R. Berardo, M. Lubell, The ecology of games as a theory of polycentricity: Recent advances and future challenges, *Policy Stud. J.* 47 (2019).
- [34] M. Nie, Drivers of natural resource-based political conflict, *Policy Sci.* 36 (2003) 307–341.
- [35] M. Lubell, J. Mewhirter, R. Berardo, The origins of conflict in polycentric governance systems, *Public Adm. Rev.* 80 (2020) 222–233.
- [36] J.B. Jacquet, Review of risks to communities from shale energy development, *Environ. Sci. Technol.* 48 (2014) 8321–8333.
- [37] T. Heikkilä, C.M. Weible, Unpacking the intensity of policy conflict: A study of Colorado's oil and gas subsystem, *Policy Sci.* 50 (2017) 179–193.
- [38] S. Lilliox, J.M. Smith, Colorado's fracking debates: Citizen science, Conflict and Collaboration, *Sci. Cult.* 27 (2018) 221–241.
- [39] D. Costie, F. Holm, R. Berardo, Hydraulic fracturing, coalition activity and shock: Assessing the potential for coalition-based collective action in Argentina's Vaca Muerta formation, *Extr. Ind. Soc.* 5 (2018) 499–507.
- [40] L. Rizzo, Fracking and Resistance in the Land of Fire: Struggles over fracking in Northern Patagonia, Argentina, highlight the need to decommodify and democratize energy resources and seek alternatives, *NACLA Rep. Am.* 49 (2017) 470–475.
- [41] A. Olive, What is the fracking story in Canada? *Can. Geogr. Géographie Can.* 60 (2016) 32–45.
- [42] J. Finkeldey, Unconventionally contentious: Frack Free South Africa's challenge to the oil and gas industry, *Extr. Ind. Soc.* 5 (2018) 461–468.
- [43] M. Espig, K. de Rijke, Unconventional gas developments and the politics of risk and knowledge in Australia, *Energy Res. Soc. Sci.* 20 (2016) 82–90.
- [44] S. Chailleux, J. Merlin, Y. Gunzburger, Unconventional oil and gas in France: From popular distrust to politicization of the underground, *Extr. Ind. Soc.* 5 (2018)

- 682–690.
- [45] L.L. Benites-Lazaro, L. Giatti, A. Giarolla, Topic modeling method for analyzing social actor discourses on climate change, energy and food security, *Energy Res. Soc. Sci.* 45 (2018) 318–330.
 - [46] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, *J. Mach. Learn. Res.* 3 (2003) 993–1022.
 - [47] T. Heikkilä, et al., Understanding a period of policy change: The case of hydraulic fracturing disclosure policy in Colorado, *Rev. Policy Res.* 31 (2014) 65–87.
 - [48] D.M. Blei, J.D. Lafferty, Topic models. in *Text Mining* 101–124, (Chapman and Hall/CRC (2009).
 - [49] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, R. Harshman, Indexing by latent semantic analysis, *J. Am. Soc. Inf. Sci.* 41 (1990) 391–407.
 - [50] S.T. Dumais, Latent semantic analysis, *Annu. Rev. Inf. Sci. Technol.* 38 (2004) 188–230.
 - [51] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, *Mach. Learn.* 42 (2001) 177–196.
 - [52] D.M. Blei, Probabilistic topic models, *Commun. ACM* 55 (2012) 77–84.
 - [53] H.M. Wallach, D.M. Mimno, A. McCallum, Rethinking LDA: Why priors matter, *Adv. Neural Inf. Process. Syst.* 1973–1981 (2009).
 - [54] K.J. Hedding, Sources and framing of fracking: A content analysis of newspaper coverage in North Carolina, New York, and Pennsylvania, *Environ. Commun.* 11 (2017) 370–385.
 - [55] McCallum, A.K. Mallet, A machine learning for language toolkit, *Httpmallet Cs Umass Edu* (2002).
 - [56] A.E. Boydston, A. Hardy, S. Walgrave, Two faces of media attention: Media storm versus non-storm coverage, *Polit. Commun.* 31 (2014) 509–531.
 - [57] C. Davis, J.M. Fisk, Energy abundance or environmental worries? Analyzing public support for fracking in the United States, *Rev. Policy Res.* 31 (2014) 1–16.
 - [58] A.E. Boydston, S. Bevan, H.F. Thomas III, The importance of attention diversity and how to measure it, *Policy Stud. J.* 42 (2014) 173–196.
 - [59] A.L. Maas et al. Learning word vectors for sentiment analysis. in 1 2011 (Association for Computational Linguistics 142 150.
 - [60] Phan, X.-H., Nguyen, L.-M. & Horiguchi, S. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. in *Proceedings of the 17th international conference on World Wide Web* 91–100 (2008).
 - [61] B. Warner, J. Shapiro, Fractured, fragmented federalism: A study in fracking regulatory policy, *Publius J. Fed.* 43 (2013) 474–496.
 - [62] R.C. Stedman, et al., Environmental reviews and case studies: Marcellus shale gas development and new boomtown research: Views of New York and Pennsylvania residents, *Environ. Pract.* 14 (2012) 382–393.
 - [63] C. Grosse, Grassroots vs. big oil: Measure P and the fight to ban fracking in Santa Barbara County California, *Case Stud. Environ.* (2017).
 - [64] G. Arnold, Does entrepreneurship work? Understanding what policy entrepreneurs do and whether it matters, *Policy Stud. J.* (2020), <https://doi.org/10.1111/psj.12388>.
 - [65] R.A. Fisher, On the 'probable error' of a coefficient of correlation deduced from a small sample, *Metron* 1 (1921) 1–32.
 - [66] C. Davis, K. Hoffer, Federalizing energy? Agenda change and the politics of fracking, *Policy Sci.* 45 (2012) 221–241.
 - [67] R.A. Fisher, *Statistical methods for research workers*, Genesis Publishing Pvt Ltd, 2006.
 - [68] G.Y. Zou, Toward using confidence intervals to compare correlations, *Psychol. Methods* 12 (2007) 399.