Preliminary Results Report

Pedro A. Arroyo, Daejin Kim, Di Tong, Arun Pandian

11 / 5 / 2019

1. Introduction

With the move of political discourse to relatively democratized digital platforms, the United States is increasingly coming into contact with computational propaganda - i.e. the use of algorithms and automation to disseminate false information or to manipulate user activity on social media (Woolley & Howard, 2017). Researchers, politicians, and policy experts have recognized the emergence of these strategies as a threat to democratic institutions. Concerns have grown over the corrosive effects of strategic competitors or malign actors using distorted, misleading, or fabricated information to affect political behavior. For example, the 2016 election saw pro-Trump campaigns leverage computational techniques to generate pro-Trump twitter accounts. By election day, the number of pro-Trump twitter accounts outnumbered pro-Clinton accounts by a factor of 5 to 1 (Kollanyi, Howard, & Woolley, 2016). These techniques threaten the democratic process on two fronts: first, they insert false information into the political consciousness; second, they undermine the franchise of individual voters.

One response to this threat has come from the world of think tanks in the form of working papers and reports attempting to frame and/or outline a counter-strategy. Think tank documents capture only a narrow slice of elite discourse, but they are nonetheless important in so far as they reflect and shape the conversation happening among decision-makers.

To gain a better understanding of how think tanks approach and react to computational propaganda, we propose to analyze published reports from prominent think tanks. We have decided to analyze these texts using topic modeling, a broad class of Bayesian generative models that serve to classify text data (Grimmer & Stewart, 2013). This will help us identify semantically cohesive topics in the corpus and give us insights into how computational disinformation is conceptualized in this discourse.

2. Corpus

2.1 Collection

Our study examines how issues surrounding computational disinformation (or propaganda; we use these terms interchangeably) are conceptualized and understood by experts from think tanks. Since a pre-existing data set isn't available, we created our own corpus of think tank reports. In fact, the bulk of our work so far has been creating a corpus and cleaning it so it can be fed into a topic modelling algorithm. We collected English-language reports from think tanks dealing with computational propaganda (often referred to as digital disinformation, or plainly disinformation), specifically those that are based in the United States and are considered influential.

To ensure good representation of think tanks, the top twenty think tanks were selected from the list entitled "Top Think Tanks in the United States" in the "2018 Global Go To Think Tank Index Report" by the University of Pennsylvania (McGann, 2019). This list includes both neoliberal think tanks like the Council on Foreign Relations as well as libertarian think tanks like the Cato Institute. This report is created by the universities' Think Tanks and Civil Societies program and is the result of "an international survey of over 1,950 scholars, public and private donors, policymakers, and journalists who helped rank more than 6,500 think tanks using a set of 18 criteria developed by the TTCSP". The report demonstrates the rankings of think tanks around the world as well as in the U.S. by their influence on policymaking which is indexed by media reputation, quality and number of publications, ability to recruit scholars, etc. By creating our corpus from these influential think tanks from across the ideological spectrum, we can be confident that the corpus is a fair approximation of the discourse of the policy architects.

Reports were collected from the think tanks using the following method:

For each think tank, we see if their reports are tagged. If they are tagged, we use the reports with tags: “online disinformation”, “computational propaganda”, “influence campaigns”, “disinformation”, “propaganda.” The usage of the last two tags would mean that we have to filter out reports that are not about online disinformation/propaganda but off-line disinformation. Some think tanks (whom we know to be working on online disinformation) clearly classify their reports with tags, and so we can include all reports under specific tags like "Online Disinformation". With sources that do not tag their reports, we use the aforementioned keywords as search terms and use our judgment to decide inclusion based on metadata (title, date published, author names, and table of contents). In short, we took an expansive approach when building our corpus and allowed discontinuities within the discursive space to emerge from the subsequent analysis.

2.2 Meta-information

The corpus consists of 80 reports from 13 think tanks (the other eight think tanks did not have any reports on this issue). Some statistics are reported in Table 1 and Table 2.

Table 1 Meta-information by year

Year	Number of reports
Pre-2016	2
2016	3

2017	15
2018	34
2019	26
Total	80

Table 2 Meta-information by the ranking of think tanks and the number of reports

Think Tank rank in US	Think Tank	Number of reports
1	Brookings Institution	15
2	Center for Strategic and International Studies	3
3	Carnegie Endowment for International Peace	1
4	Heritage Foundation	1
6	RAND Corporation	11
7	Peterson Institute for International Economics	2
8	Center for American Progress	16
10	Atlantic Council	8
11	Council on Foreign Relations	6
14	Belfer Institute	13
15	Hudson Institute	2
16	American Enterprise Institute	1
19	Stimson Center	1

2.3 Pre-processing

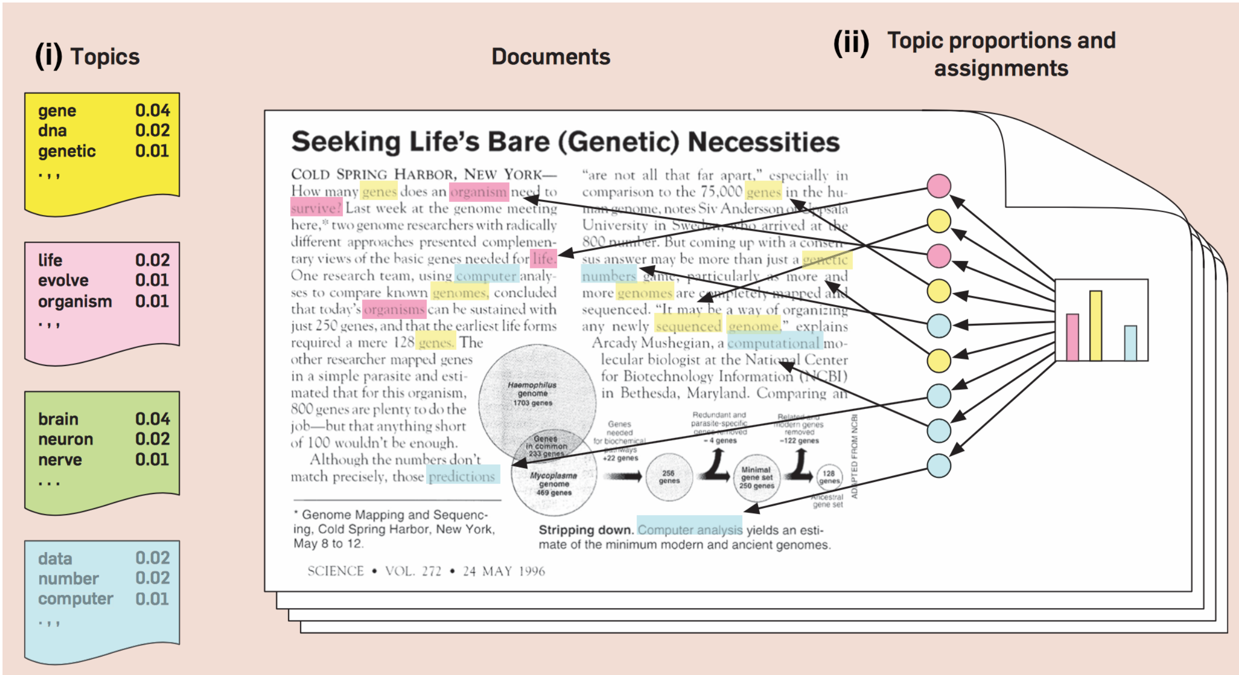
Each report was converted to a text file and saved in a folder. Naming conventions for files recorded metadata for later analysis (see below). If the report was already available on a webpage, we just extracted the text by hand. If the report was available in PDF form, we used the python library PDFMiner to convert PDF files to text files. This choice to automate conversion was necessary to handle the amount of text we obtained. We treated the text files as bags of words, i.e., no information about page or section breaks was taken into account, because the documents are treated as bags of words by the downstream topic modeling algorithm. For the cleaning process, we tokenized texts by spaces and removed stopwords, numbers, words with numbers, and words with non-English characters as well as some frequently occurring words that do not add meaning to the documents. We also had to go through the text files manually to find words that were part of the document syntax (e.g. section names like Introduction, Conclusion, etc.) Although numbers could reveal some information (such as important years as it pertains to the discourse), it is not unusual in text analysis to get rid of numbers as they might just add noise and a lot of them might be part of the document syntax (and therefore not contribute to the discourse).

3. Method

Our analysis will rely on topic modeling, a broad class of Bayesian generative models that serve to classify text data (Grimmer & Stewart, 2013). As an unsupervised machine learning technique, topic modeling has the potential to either confirm existing theories or discover unknown categories and patterns not immediately apparent to human readers (Blei, Ng, & Jordan, 2003; Evans & Aceves, 2016; Nelson et al., 2018). Comparing with the more basic form of text-classification method, Latent Semantic Analysis (LSA), which mainly serves to reduce the dimensions of the corpus through building a space with stable semantic distances, topic modeling is better at discovering "semantically cohesive topics and their combination across document collections (Evans & Aceves, 2016)." Latent Dirichlet Allocation (LDA) is the most prevalently-used algorithm for topic modeling. Assuming that each document contains a mixture of topics, some of which occur throughout the entire corpus, LDA uncovers hidden thematic structures both in the corpus and in individual documents. Specifically, it can identify (i) the topics defined by a group of words with different weights in a given corpus and (ii) the distribution of topics in the corpus as well as in an individual document (see Figure 1 for visualization).

Specifically, we will use the recently-developed Structural Topic Modeling (STM), an extension of LDA that takes into account the document metadata, such as the publication year, the author, the region focus, etc. when identifying topics and determining their prevalence (Roberts et al., 2014; Roberts, Stewart, & Tingley, 2014; Farrell, 2016). STM would result in more accurate classification for a corpus with a great deal of variation in its meta-information, such as a corpus of documents produced by a lot of heterogeneous institutions or covering a broad time period. STM is also a convenient tool for us to explore the inter-topic relations and the relations between topics and the meta-information of documents.

Our approach is largely exploratory, especially so at the outset. To begin, we will use STM to establish the shared dominant topics for all texts in our corpus to get a general sense of the major concerns regarding disinformation and related campaigns in the United States. Moreover, we will investigate how the generated topics are related to each other by comparing the frequency distribution of words in each topic. Also, we will examine the distribution of topics along the meta-categories of documents, such as the institutions and authors that produce the documents, their party affiliation, the publication time of the documents and the regional focus of the document. Figure 2 demonstrates the general flow of the topic modeling analysis.



Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77. doi: 10.1145/2133806.2133826

Figure 1 (i) Topics defined by a group of words and (ii) the distribution of the topics in the corpus as well as in an individual document

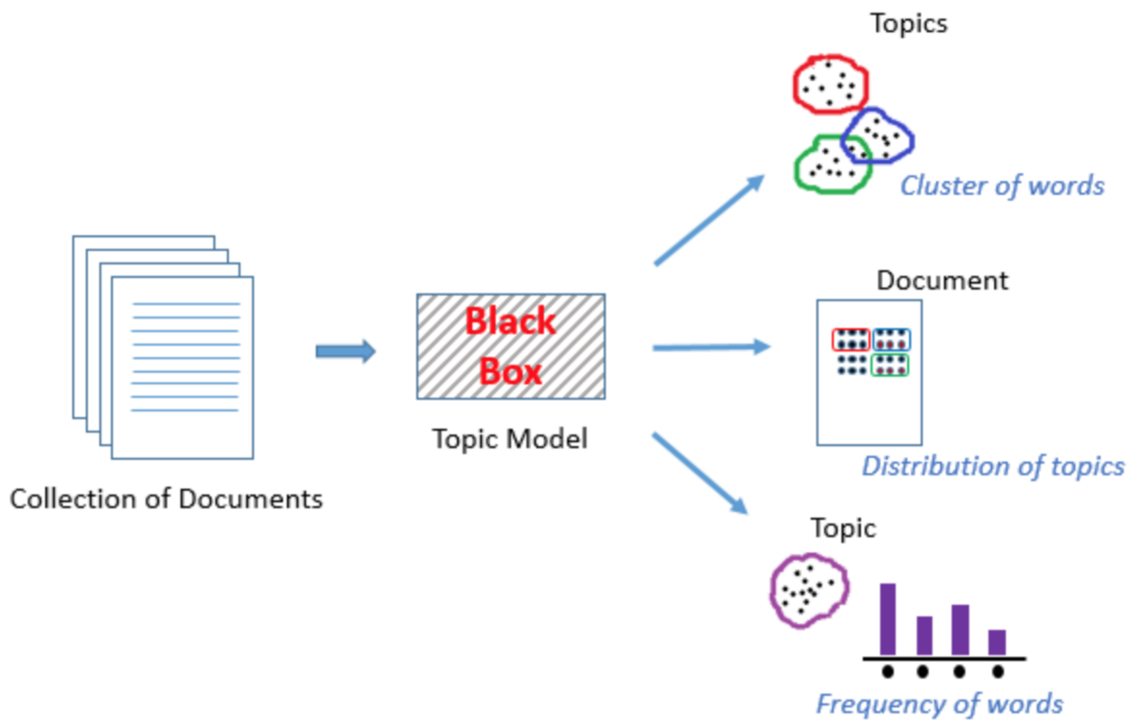


Figure 2 https://www.analyticsvidhya.com/blog/2018/10/stepwise-guide-topic-modeling-latent-semantic-analysis/

4. Preliminary Results

We used a word cloud generated using the "wordcloud" package on Python to know what we can expect once we finish the topic modeling analysis. We used all the texts after pre-processing. In word clouds, the bigger the text the more frequently the word appears. Thus word clouds visually represent frequency distribution of the top k (for some k) most occurring words in the text. As seen in Figure 3, certain themes are prominent. As can be expected, Russia and the United States occur very prominently reflecting the nature of the discourse post-2016 from the US-based think tanks we have looked at. Moreover, many political terms can be found, such as "campaign," "election," "government," "public," demonstrating the strong connection between

American politics and computational disinformation in recent years. These political words may appear as a general topic when the data are fit in the topic modeling.

Also, some words may be able to indicate in what tone experts are talking about computational disinformation. For example, negatively charged words such as “attack,” “threat,” and “risk” in the word cloud may suggest that some reports describe the disinformation as an alarming issue in U.S. politics. In the topic modeling, these words may further form a topic distribution to represent how aggressively or negatively some reports view the issue and these sentiments might change across time.

4.1 Expectations

Since we use a model that incorporates meta-data like time, and publisher, we will be able to see how the discourse changes between 2016 to 2019, i.e., how the problem and its solution is conceptualized in different time periods. Perhaps in 2016, we will see topics that are indicative of a new type of discourse being formed (indicated by perhaps, topics related to “newness”) which will cease to be found in later years. We also would like to see how the discourse changes by the 12 think tanks we have used. It will be interesting to see where the overlaps and the differences lie, or if they do at all.

Figure 3 Word cloud of the corpus with eighty reports from thirteen think tanks

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Evans, J. A., & Aceves, P. (2016). Machine translation: Mining text for social theory. *Annual Review of Sociology*, *42*.

Farrell, J. (2016). Corporate funding and ideological polarization about climate change. *Proceedings of the National Academy of Sciences*, *113*(1), 92-97.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21*(3), 267-297.

McGann, J. G. (2019). 2018 Global Go To Think Tank Index Report.

Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2018). The future of coding: a comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods & Research*, 0049124118769114.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). stm: R package for structural topic models. *Journal of Statistical Software*, *10*(2), 1-40.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064-1082.