

The Power of Paranoia: Why Conspiracy Theories Persist

Luca Di Bartolomeo Florian Hofhammer Philipp Mao

Contents

1	Introduction	1
2	Obtaining the Dataset	2
3	Designing the Visualization	2
4	Implementing the Visualization	3
5	Peer Assessment	3
	References	3

1 Introduction

Conspiracy theories have for a long time already played an important role in the history of mankind. Oftentimes, xenophobic and religious motives played an important role in their spread. For example, the big outbreak of the plague in the 14th century caused serious persecutions of Jews due to conspiracy theories that Jews poisoned wells in order to extinguish the Christian population in Europe.

While a multitude of conspiracy theories kept getting invented throughout the centuries, we focus on conspiracy theories with a significant spread since the turn of the millenia. More specifically, we focus on conspiracy theories playing an important role in modern populist politics in the United States of America.

To this extent, we gather and analyze the dataset of Tweets of the former US President Donald Trump as an important populist politician as well as the transcriptions of the Infowars podcast published by Alex Jones, an US-American right-wing populist political activist.

2 Obtaining the Dataset

Due to Twitter’s API access being severely restricted, we had to rely on previous dumps of Trump’s Tweets. Luckily, the full set of Donald Trump’s Tweets from 2009 until the suspension of his account in early 2021 is readily available on GitHub [1].

The dataset of Trump Tweets has already been the basis of research publications [2]. As such, we consider this dataset not novel enough to constitute the sole source of data for our visualization.

Consequently, we obtained the full dataset of transcripts of Alex Jones’ Infowars podcast only recently published [3]. This dataset has been created by transcribing the podcast episodes with OpenAI’s Whisper speech-to-text engine [4]. The dataset consists of 187’748’262 words of text, adding up to 1.2 GB of textual data. To the best of our knowledge, we are the first project to conduct extensive analysis on this dataset due to its recency.

During the implementation phase of our project, we found the Trump Tweet dataset to be too sparse for meaningful analysis and comparisons with relation to conspiracy theories. As a result, we shifted the focus of our visualization from a comparison of Alex Jones’ and Donald Trump’s views on conspiracy theories to a comparison between Alex Jones’ views reported in the Infowars podcast and news covered in different US-American media outlets.

This latter dataset provides us with 2.7 million news articles from the years 2016-2020, amounting to roughly 8.8 GB of textual data [5].

In order to prepare the textual data for further analysis, we trained a Word2Vec model [6], [7] on the datasets. This preprocessing step allows us to extract connections between conspiracy theories in our visualization.

A copy of our underlying dataset including the Word2Vec model we trained can be obtained at <https://go.epfl.ch/com480-conspiracies>.

3 Designing the Visualization

Our visualization focuses on how conspiracy theories evolve over time and how the views of different sources on the theories differ. For the former purpose, we highlight semantic connections between conspiracy theories and their development over time. Thus, a consumer of our visualization can for each conspiracy theory we focus on deduce what other conspiracy theories developed in parallel and what causal relationships between conspiracy theories may exist.

To achieve visualizing these relationships, a timeline which shows the frequency of specific terms in our dataset per conspiracy theory is combined with a chord chart highlighting the connections between theories. The width of the chords in the graph determines the contextual closeness of two conspiracy theories. This contextual closeness implies a correlation: either the conspiracy theories exhibit

an overlap in affected topics and likely stem from the same motives behind them (xenophobia, homophobia, antisemitism, etc.) or that one of them developed into the other over time. The type of this correlation can be determined by observing the development over time of the theories in the aforementioned timeline.

Given that we cannot assume that every page visitor is familiar with the conspiracy theories presented, hovering over a conspiracy theory also provides the reader with contextual information.

In order to visualize how different sources (i.e., Alex Jones and different classic US-American media outlets and news agencies report about conspiracy theories, we once again leverage a timeline. In this graph, a page visitor can select and deselect data sources for a given conspiracy theory to plot the development over time across sources. This provides the viewer with an insight into the temporal relationship of references to a conspiracy theory across classic media outlets and Alex Jones’ Infowars podcast.

On the other hand, we refrain from providing a wordcloud with conspiracy theories as we had originally intended. During the implementation of our visualization, we encountered issues with the sparsity of the wordcloud. We consequently decided to remove the wordcloud from our final visualization and focus on the timelines and chord charts instead.

4 Implementing the Visualization

An example for code blocks:

Listing 1: test

```
1 hello world
2 test
```

References work like this: Lst. 1

5 Peer Assessment

References

- [1] M. Hershey, “Complete Donald Trump Tweets Archive.” Jan. 10, 2021. Available: <https://github.com/MarkHershey/CompleteTrumpTweetsArchive>
- [2] E. A. Morales, C. J. P. Schultz, and K. D. Landreville, “The Impact of 280 Characters: An Analysis of Trump’s Tweets and Television News Through the Lens of Agenda Building,” *Electronic News*, vol. 15, no. 1–2, pp. 21–37, Mar. 2021, doi: 10.1177/19312431211028610.
- [3] E. Simonsen, “Infowars.” Apr. 18, 2023. Available: <https://github.com/Fudge/infowars>

- [4] E. Simonsen, “4682 episodes of The Alex Jones Show (15875 hours) transcribed,” Mar. 22, 2023. https://www.reddit.com/r/datasets/comments/11yyoth/4682_episodes_of_the_alex_jones_show_15875_hours/
- [5] A. Thompson, “All the News 2.0 – 2.7 million news articles and essays from 27 American publications.” Jul. 09, 2022. Available: <https://components.one/datasets/all-the-news-2-news-articles-dataset/>
- [6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Sep. 06, 2013. <http://arxiv.org/abs/1301.3781> (accessed Jun. 02, 2023).
- [7] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” Oct. 16, 2013. <http://arxiv.org/abs/1310.4546> (accessed Jun. 02, 2023).