

# The Power of Paranoia: Why Conspiracy Theories Persist

Luca Di Bartolomeo      Florian Hofhammer      Philipp Mao

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Obtaining the Dataset</b>	<b>2</b>
<b>3</b>	<b>Designing the Visualization</b>	<b>2</b>
<b>4</b>	<b>Implementing the Visualization</b>	<b>2</b>
<b>5</b>	<b>Peer Assessment</b>	<b>2</b>
	<b>References</b>	<b>2</b>

## 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.

In order to prepare the textual data for further analysis, we trained a Word2Vec model [5], [6] on the datasets. This preprocessing step allows us to extract connections between conspiracy theories in our visualization.

## 3 Designing the Visualization

## 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/](https://www.reddit.com/r/datasets/comments/11yyoth/4682_episodes_of_the_alex_jones_show_15875_hours/)
- [5] 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).
- [6] 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).