A Network Analysis of Censored Books

Identifying Genre and Author Trends in Banned and Challenged Books

Jeremy Piperni  
 Computer Science  
 McGill University  
 Montreal, QC, Canada  
jeremy.piperni@mail.mcgill.ca

ABSTRACT

This progress report aims to update the reader on the progress made on the network analysis of censored books. The project aims to identify genre and author trends in banned and challenged books. By pursuing this project, we could identify if some genres and/or authors are more likely to be censored. These trends could lead to understanding if political or ideological factors play a role in the rise of book bans.

KEYWORDS

Networks, Censorship, Books, Banned Books, Challenged Books, Genre Analysis, Author Analysis, Education, Politics

1 INTRODUCTION AND MOTIVATION

Censorship and freedom of speech have been at the forefront of politics and public discourse in North America. While freedom of speech tends to be important to a lot of individuals, the banning of books has surged in the US, with an increase of 65% in 2023 compared to 2022 [1]. In the past, certain books were associated with events like assassinations, which lead to justifying certain book bans. Famously, the man who killed John Lennon was found with “The Catcher in the Rye” on his person [2]. The book is often criticized for its vulgarity, dishonesty, and use of alcohol [3], which led to it being heavily censored. The assassination reinforced the beliefs of many, that the book could lead to violence, leading to them justifying the ban. Nowadays, book censorship is more frequent, although justifications and reasonings are often lacking.

This prompts several questions: What/Who is leading this increase in censorship? What political and ideological factors are maybe driving this censorship? Some articles have been written with the goal of answering these questions, but none have utilized network science as their methodology. This project aims to identify genre and author trends in banned books by utilizing network models. By analyzing these networks, some trends might lead us to answers on the possible agendas of the individuals or groups pushing for these book bans. A network analysis approach can also uncover hidden relationships and patterns, by revealing clusters of nodes that might be easily missed with other approaches.

2 RELATED WORK

There is a lack of research for network analyses that focuses on censored books. This project aims to bridge this gap and provide some insights to future work. The following articles provide relevant information on both network approaches, and book censorship context.

2.1 Detecting Network-based Internet Censorship via Latent Feature Representation Learning [4]

The article written by Shawn P. Duncan and Hui Chen designed a classification model that detects network-based internet censorship. The authors propose a sequence-to-sequence autoencoder to capture the structure data [4]. They then utilize a densely connected multi-layer neural network model to estimate the probability of censorship events [4]. The authors also created a second model, that uses network reachability data for an image-based classification model [4]. Both models were able to successfully detect network-based internet censorship [4]. While this article focuses on internet censorship and not book censorship, it can provide interesting network-based approaches that can be applied in this project.

2.2 Book bans in political context: Evidence from US schools [5]

This article written by Langrock et al. focuses on the rise in book bans in the US. Using data from PEN America of 2,532 book bans during the 2021-2022 school year, the authors analyze the types of banned books and authors, the socio-political environments of book bans, and the interest in book bans. The authors discover that banned books feature disproportionately characters of color in children’s books, and that banned books are disproportionately written by people of color [5]. They also discovered that counties that are right-leaning but have become less conservative are more likely to ban books [5]. This article provides a useful analysis of both genres and authors of banned books, which we can compare to the results of this project. However, the authors take a more direct approach to their research, and do not utilize network models.

2.3 A History of Censorship in the United States [6]

This essay written by Jennifer Elaine Steele provides a history of censorship in the US. The author delves into the censorship in public libraries and the censorship in schools. They speak about many types of censorship, like the censorship of religion, comic books, communist texts, and many more. This essay provides a necessary context of censorship, which can be used to explain some of the censorship trends that we will find in this project.

3 PROBLEM DEFINITION

The censorship of books can often appear to be subjective, where political and ideological factors often drive these decisions. This project aims to bring an understanding of whether certain genres or authors are targeted disproportionately to censorship. Several network models will be created using the dataset mentioned in section 4. These models will be analyzed, with the goal of answering multiple censorship questions like the following: Do specific authors face a higher likelihood of censorship? Do specific genres tend to be censored more often? Do other hidden censorship patterns emerge from analyzing thematic connections between banned books? Answering these questions would provide insights into who and what is censored more often, and if there could be political or ideological factors responsible.

4 DATASET

A dataset compiled by Chieler Li [7], gathered book information during the DotData Hackathon in 2025. The dataset contains the title, author, book description, and genre of around 17,000 books, which of these around 7,700 books are labelled as censored, challenged, or banned. All descriptions and genres were gathered from Goodreads, and all banned/challenged books were gathered using ALA, Wiki, and PEN America’s Index. All uncensored books will be ignored, as they do not pertain to this project.

430 –> 306 genres

REFERENCES

[1] *American Library Association*. 2024. Book Ban Data. Retrieved March 18, 2025, from <https://www.ala.org/bbooks/book-ban-data>.

[2] Carroll, S. 2020. The role of Catcher in the Rye in John Lennon’s assassination. Medium. Retrieved March 28, 2025, from <https://medium.com/@sec220/the-role-of-catcher-in-the-rye-in-john-lennons-assassination-4216033a75ed>

[3] Frangedis, H. (1988). Dealing with the Controversial Elements in “The Catcher in the Rye”. The English Journal, Vol. 77, No.7, 72-75, <https://doi.org/10.2307/818945>

[4] Duncan, S.P., Chen, H. (2023). Detecting Network-based Internet Censorship via Latent Feature Representation Learning. Computers & Security, Vol. 128, https://doi.org/10.48550/arXiv.2209.05152

[5] Langrock, I., LaViolette, J., Goncalves, M., Sargsyan, K., Carter, A. M., & Sherrill-Mix, S. (2024). Book bans in political context: Evidence from US schools. PNAS Nexus, 3(6), pgae197. <https://doi.org/10.1093/pnasnexus/pgae197>

[6] Steele, J.E. (2020). A History of Censorship in the United States. Journal of Intellectual Freedom and Privacy, 5(1), 6–19. https://doi.org/10.5860/jifp.v5i1.7208

[7] Chieler Li. 2024. *Banned Book Dataset*. Kaggle, from https://www.kaggle.com/datasets/chielerli/banned-book-dataset

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00