

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

In recent years, the dramatic surge in book censorship in the United States has raised concerns about censorship and freedom of expression. This project applies a network science approach to identify genre and author trends in banned books. By constructing and analyzing four networks, genres such as “LGBT”, “Queer”, and “Young Adult” were disproportionately banned despite their lower popularity. Additionally, women tend to be challenged more than men, even though they publish fewer books on average. These results suggest that recent censorship trends may be influenced by social and political agendas.

CODE

<https://github.com/jeremy-piperni/Network-Analysis>

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 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.

5 METHODOLOGY

For the network analysis, multiple networks will be created. This section will explain the network construction, and graph analysis techniques for each network.

5.1 Genre-Book Bipartite Network

5.1.1 Network Construction. The network will be a bipartite graph, where nodes are divided into two distinct sets, and edges can only exist between sets. The first set of nodes will consist of the genres, while the second set of nodes will consist of the books. Edges will connect books to their corresponding genres.

5.1.2 Graph Analysis. The bipartite network provides the opportunity of conducting centrality measure analyses. Degree centrality will be performed on the genre nodes. This will provide information on which genres are banned often. A simple degree distribution will be conducted on the book nodes, to provide insight into whether multi-genre books are more likely to be banned. Eigenvector centrality on the genre nodes will measure a nodes influence based on the node's connections to other nodes in the network, identifying which genres tend to be linked to the most "influential" books.

5.2 Author-Book Bipartite Network

5.2.1 Network Construction. The network will be a bipartite graph where the first set of nodes will consist of the authors, while the second set of nodes will consist of the books. Edges will connect books to their corresponding authors.

5.2.2 Graph Analysis. The bipartite network provides the opportunity of conducting centrality measure analyses. Degree centrality will be performed on the author nodes.

This will provide information on which authors are censored the most often.

5.3 Genre Co-Occurrence Network

5.3.1 Network Construction. The nodes of the network are varying genres. Edges will be added between each node, when a banned book belongs to both genres. The graph is weighted, therefore when a book shares genres, a weight of 1 will be added to the weight of the edge.

5.3.2 Graph Analysis. Clusters will be created between genres that are often banned together. This graph structure permits community detection analysis and centrality measure analysis. For community detection, the Louvain method is utilized; Where communities will be created, demonstrating genre clusters that are frequently challenged together. Certain centrality measures will also provide some insights regarding the relationship between genres. Firstly, Weighted betweenness centrality will provide valuable insight by identifying key genres that link other genres, by measuring how often a genre lies on the shortest path between two other genres. Lastly, weighted closeness centrality will be used to identify the genres that are at the core of censored books.

5.4 Book Similarity Network

5.4.1 Network Construction. Genres don't always capture the similarities between books. This network will aim to provide a different approach to clustering. The nodes of this network will be the book titles. Edges will be constructed when two books share a high similarity. The similarity measure will be calculated by using TD-IDF and Cosine Similarity on the descriptions of the books provided.

5.4.2 Graph Analysis. Clusters will be created between books that share similarities, even if they belong to different genres. The Louvain method will be utilized for community detection, where books with similar themes will be clustered. These books will then be analyzed for their genres, to see if hidden thematic connections were found that the genre analysis was missing.

6 EXPERIMENT SETUP

6.1 Data Extraction / Cleaning

To prepare the network data, the Numpy and Pandas libraries for Python3 were utilized. The raw csv data was extracted, and all non-censored books were removed. An exploratory

analysis of genres was conducted, where 430 unique genres appeared. After an examination of these genres, redundancy was apparent. A mapping file was created to remove some of these redundancies within reason. For example, mapping "nonfiction" to "non-fiction", mapping "humorous" to "humor". Furthermore, children's books had dozens of genres: "children's literature", "children's non-fiction", etc... These were all mapped to "childrens". After the genre cleanup, with entailed turning every character into lowercase, removing out-of-place values, and mapping certain genres, 306 unique genres remained.

The author and description columns were also cleaned; All characters were formatted to lowercase and all non-ascii characters were removed. The author column in particular needed special attention, as author name convention was not uniform. For example, both "cast p. c." and "p. c. cast" appeared, but both represented the same author. Regex and special rules were applied to fix these issues.

6.2 Network Creation

To facilitate network creation, 3 .csv files were created from the original file with only the necessary columns needed for each network. The Numpy, Pandas, and NetworkX libraries for Python3 were utilized for creating the networks. The Matplotlib library was used to draw the networks.

6.2.1 Genre-Book Bipartite Network. The genre_book_network csv file was used to create this network. The network contains 2869 book nodes, and 306 genre nodes. 17,976 edges connect the book and genre nodes together. Figure 1 shows the Genre-Book Bipartite Network. Note that a small unconnected subgraph of 1 book and 1 genre node is removed from the visualization. Book nodes are colored light-blue, and genre nodes in red.

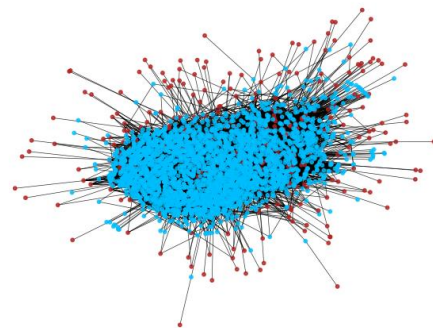


Figure 1: Genre-Book Bipartite Network

6.2.2 Author-Book Bipartite Network. The `author_book_network.csv` file was used to create this network. The network consists of 2876 book nodes, and 2317 author nodes. 3532 edges connect the book and author nodes. Figure 2 shows the Author-Book Bipartite Network. Book nodes are colored light-blue, and author nodes in green.

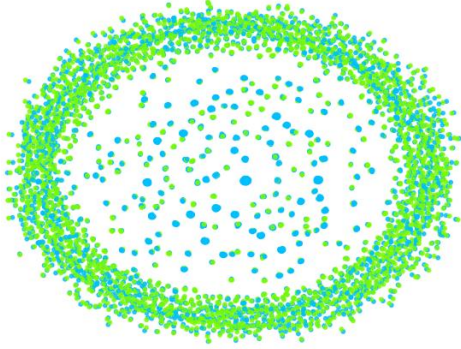


Figure 2: Author-Book Bipartite Network

6.2.3 Genre Co-Occurrence Network. The `genre_book_network.csv` file was used to create this network. The network consists of 306 nodes representing genres. The network holds 4355 weighted edges between genres, representing books that have shared genres. Figure 3 shows the Genre Co-Occurrence Network.

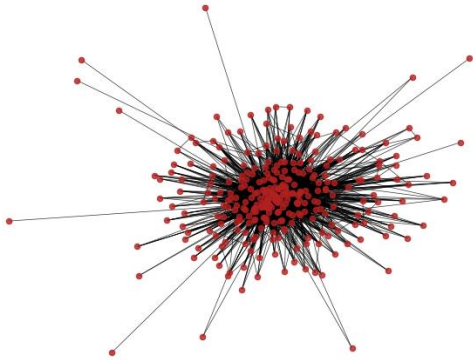


Figure 3: Genre Co-Occurrence Network

6.2.4 Book Similarity Network. The `book_description_network.csv` file will be used to create this network. The network consists of 2875 book nodes, and 2881 weighted edges. The weight of the edges is calculated using TD-IDF on the book description; Then applying cosine similarity to see if two nodes are “similar”. Only edges with

a similarity score of over 0.5 are added to the network. Figure 4 shows the Book Similarity Network.

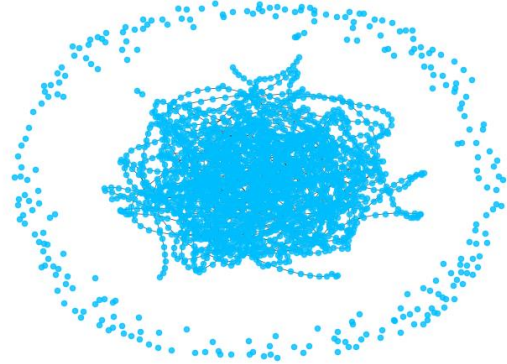


Figure 4: Book Similarity Network

7 RESULTS

7.1 Genre-Book Bipartite Network

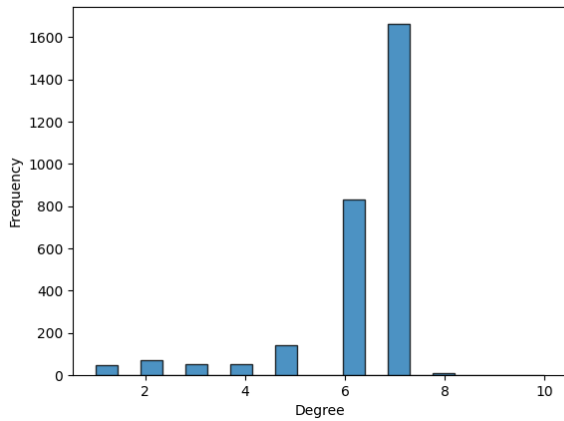
Using the Genre-Book Bipartite Network, degree centrality was conducted on both genre and book nodes. For the genre nodes, a higher degree centrality score demonstrates that the genre was more heavily censored than other genres. Table 1 shows these results; Where “Fiction”, “Young Adult”, “Contemporary”, and “Romance” books had the highest degree centrality scores. Eigenvector centrality values were also calculated on the genre nodes where results can be seen in Table 2. “Fiction”, “Historical”, “Audiobook”, and “Young Adult” books had the highest eigenvector centrality scores. Lastly, the degrees of the book nodes were transformed to a degree distribution graph, as can be seen in Figure 5. Banned books in the dataset mostly contained around 6 to 7 genres.

Table 1: Highest Degree Centrality Scores of Genres

Genre	Score	Genre	Score
Fiction	0.707	Audiobook	0.197
Young Adult	0.526	Children’s	0.197
Contemporary	0.325	Historical	0.187
Romance	0.325	Queer	0.151
Fantasy	0.232	Non-Fiction	0.133
LGBT	0.226	Mystery	0.125
Realistic Fiction	0.226	Picture Books	0.110

Table 2: Highest Eigenvector Centrality Scores of Genres

Genre	Score	Genre	Score
Fiction	0.194	LGBT	0.148
Historical	0.165	Realistic Fiction	0.145
Audiobook	0.162	Non-Fiction	0.145
Young Adult	0.161	Fantasy	0.135
Children's	0.158	Classics	0.132
Contemporary	0.158	Middle Grade	0.130
Romance	0.155	Mystery	0.130

**Figure 5: Degree Distribution of Book Nodes (Genres per Banned Book)**

7.2 Author-Book Bipartite Network

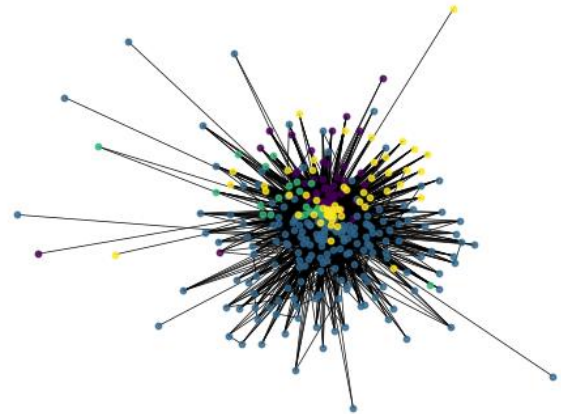
Using the Author-Book Bipartite Network, degree centrality was conducted on the author nodes. Table 3 shows the highest calculated degree centrality scores on the author nodes. The best-selling authors, Stephen King, P. C. Cast, and Ellen Hopkins obtained the highest degree centrality scores. Eigenvector centrality was also conducted on the author nodes, but no significant results were found. More in depth analyses of these authors will be conducted in section 8.

Table 3: Highest Degree Centrality Scores of Authors

Author	Score	Author	Score
Stephen King	0.0156	James Patterson	0.0049
P. C. Cast	0.0080	Anh Do	0.0045
Ellen Hopkins	0.0063	Cassandra Clare	0.0045
Sarah J. Maas	0.0056	Jacqueline Jules	0.0045
Saadia Faruqi	0.0049	Rick Riordan	0.0042
Cecily von Ziegler	0.0049	Lauren Myracle	0.0042

7.3 Genre Co-Occurrence Network

Using the Genre Co-Occurrence Network, the Louvain community detection algorithm was conducted on the nodes. The algorithm found 4 distinct communities as can be seen in Figure 6. Community 1 contained 43 genres and can be seen as the “Fantasy” community, containing genres like “Fantasy”, “Vampires”, and “Cyberpunk”. Community 2 contained 55 genres and can be classified as the “Controversial” community, containing genres that can be perceived as contentious, such as “LGBT”, “Mental Health”, and “Erotica”. The third community would be classified as the “Mystery” community, with only 22 genres such as “Mystery”, “Thriller”, and “Dark”. The last community incorporates 186 genres with no discernable theme, including genres like “Soccer”, “Games”, “Geography”, and “Science”. Every genre in the “Controversial” community can be found in Appendix A; Further analyses will be made in Section 8.

**Figure 6: Louvain Communities of Genre Nodes**

Weighted betweenness and closeness centrality scores were also calculated on the genre nodes. Table 4 and 5 present the highest scores respectively. “Reference”, “Fiction”, and “Parenting” books appeared in the top 4 for both weighted betweenness and closeness centrality.

Table 4: Highest Weighted Betweenness Centrality Scores of Genres

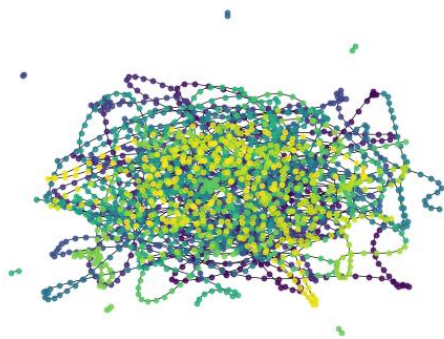
Genre	Score	Genre	Score
Non-Fiction	0.091	Contemporary	0.045
Reference	0.086	Historical	0.042
Fiction	0.084	Autobiography	0.041
Parenting	0.064	Classics	0.040
Young Adult	0.059	Audiobook	0.037
Picture Books	0.059	Diary	0.036
Children's	0.048	Coming of Age	0.036

Table 5: Highest Weighted Closeness Centrality Scores of Genres

Genre	Score	Genre	Score
Reference	0.263	Autobiography	0.246
Parenting	0.257	Sexuality	0.245
Fiction	0.254	Politics	0.245
Diary	0.254	Erotica	0.243
Coming of Age	0.252	Theology	0.242
Non-Fiction	0.251	Survival	0.242
Picture Books	0.247	Horses	0.241

7.4 Book Similarity Network

Using the Book Similarity Network, the Louvain community detection algorithm was conducted on the nodes. All unconnected nodes were removed from the graph as they do not belong to any community. 62 communities were found with the algorithm as seen in Figure 7. Genre information for each book was gathered and a genre distribution for each community was conducted. Appendix B shows each community with a size of 50 or more nodes; Only genres with a distribution of more than 5% in a community are shown. Percentages are rounded to the nearest whole number. Most communities were led in distribution by “Fiction” and “Young Adult” genres, with distributions never exceeding 14%.

**Figure 7: Louvain Communities of Book Nodes based on Book Descriptions**

8 EVALUATION

8.1 Genre Analysis

Table 1 in section 7.1 provided insight on which book genres appeared frequently among banned books. While this information is pertinent, we cannot yet say that these book genres are disproportionately targeted for bans. Some book genres are more popular than others; For example, fiction books that encompass many other genres are more popular than autobiographies. The “Fiction” genre appears with the highest degree centrality score, but it wouldn’t be fair to conclude that fiction books are disproportionately challenged. Unfortunately, book sales data are not readily available for precise calculations of genre popularity. Nevertheless, OpenLibrary.org is an online library catalogue that aims to have every book ever published on its platform [8]. By searching genres manually on the website, the number of results for that genre can be used to analyze if books are disproportionately banned. Table 6 shows the genres gathered from Table 1, sorted by a bias index (Degree Centrality Score / (Books / 10000)).

Table 6: Highest Bias Index of Genres

Genre	Books	Deg. Centrality	Bias Index
Realistic Fiction	114	0.226	19.825
LGBT	389	0.226	5.810
Audiobook	417	0.197	4.724
Queer	1096	0.151	1.378
Young Adult	8052	0.526	0.653
Non-Fiction	2147	0.133	0.620
Contemporary	23676	0.325	0.137
Picture Books	23572	0.110	0.047
Fantasy	83560	0.232	0.028
Romance	143082	0.325	0.023
Historical	85176	0.187	0.022
Mystery	91583	0.125	0.014
Fiction	619942	0.707	0.011
Children's	384002	0.197	0.005

From Table 6, the most disproportionately challenged genres are “Realistic Fiction”, “LGBT”, “Audiobook”, “Queer”, and “Young Adult. These calculations can be heavily influenced by the OpenLibrary.org database, which is further examined in section 9.3. Notably, 4 out of these 5 genres appeared in the Louvain “Controversial” community (as seen in Appendix A) created in section 7.3; that is the “Realistic Fiction”, “LGBT”, “Queer”, and “Young Adult” genres appear in the “Controversial” community.

The Genre Co-Occurrence Network brought some additional insights into the genre analysis that the Genre-Book Bipartite Network did not provide. Weighted betweenness centrality was used to see if any genres were central to the network’s structural flow, while helping tie together other genres. Weighted closeness centrality was also used to find genres that frequently co-occur, demonstrating a core genre that isn’t isolated. Notably, two genres stood out that did not appear in the degree centrality and eigenvector centrality experiments from section 7.1. Both the “Parenting” and “Reference” genres had top 3 weighted closeness and betweenness scores. These genres can be classified as having a small popularity (less than 50,000 books published) [8] but are frequently co-occurring and very important to the structure of the Genre Co-Occurrence Network. Because of their low popularity and significance in the mentioned network, “Parenting” and “Reference” books can be seen as being disproportionately targeted to bans. Notably, both these genres are not part of the Louvain “Controversial” community (as seen in Appendix A).

8.2 Author Analysis

Table 3 in section 7.2 provided insight into which authors had the most books challenged or banned. The dataset used for this project does not provide any author information, so using Goodreads, some general information from the authors in Table 3 was gathered (as seen in Table 7 below) [9].

Table 7: Author General Information

Author	Books Written	Sex	Ethnicity/Nationality
Stephen King	65	Male	American
P. C. Cast	50	Female	American
Ellen Hopkins	23	Female	American
Sarah J. Maas	14	Female	American
Saadia Faruqi	39	Female	Pakistani-American
Cecily von Zeigesar	30	Female	American
James Patterson	200+	Male	American
Anh Do	98	Male	Vietnamese-Australian
Cassandra Clare	62	Female	American
Jacqueline Jules	22	Female	American
Rick Riordan	31	Male	American
Lauren Myracle	20	Female	American

11 out of the 12 most challenged authors were American, with the other being Australian. This makes sense as most

books in the dataset were English written books, and the dataset is focused on challenged books in the USA. 8 out of the 12 most challenged authors are Female which is interesting. Especially that on average, the 8 female authors had around 33 written books, while the male authors averaged around 99 written works. Therefore, the female authors wrote less books but had higher degree centrality scores, which tends to signify that women are being disproportionately targeted for book censorship. The 12 most challenged authors wrote many books with varying genres. Although the “Fiction”, “Young Adult”, “Children’s”, and “Fantasy” genres appeared the most as common genres between these authors [9].

8.3 Description Analysis

In section 7.4, Louvain community detection was applied to the Book Similarity Network to find if any thematic connections would create clusters to find hidden patterns in banned books. 62 communities were found, and Appendix B shows the genre distributions of every community with more than 50 nodes. Sadly, the genre distribution does not provide many insights. The “Fiction”, “Young Adult”, and “Contemporary” genres lead in almost every community, which makes sense as the most banned books in the dataset are from those genres (as seen in Table 1). Some communities had a significant (over 5%) distribution of “Controversial” genres found in section 7.3, such as “LGBT” and “Queer”. Since the number of books in the dataset is quite vast, a manual analysis to uncover hidden themes of these 62 communities was not possible. This analysis can be pursued in future work.

9 DISCUSSION

9.1 Political / Ideological Factors

The findings of this study suggest that certain book genres and authors are disproportionately banned, where the censorship examined often revolves around marginalized identities and social issues. Women tend to be censored more than men despite writing fewer books. Genres such as “LGBT”, “Queer”, and “Young Adult” appear often in censoring, while being less popular genres. These results “Parenting” and “Reference” books are also overrepresented in censorship. These books tend to be important literature for knowledge and for the growth of children. These results reflect the current reality of a growing right-wing movement, especially prevalent in the United States [10]. In 2025, right-

wing groups have been pushing for more censorship in school libraries, often targeting books focused on LGBTQ+ issues and race [10]. The reasoning behind this censorship is that parents should be the ones to decide what books their children can read, and to limit the school's power in educating their students [10]. These censorship decisions are being backed by Republican lawmakers, religious groups that believe that publicly funded schools should reflect their ideological views.

“LGBT”, “Queer”, and “Young Adult” are also disproportionately banned by having a high degree centrality score while being genres that are less popular. These results align with the current reality of an increase in book censorship [1], often led by right-wing leaning groups [10].

9.2 Relevant Literature

The article written by Langrock et al. revealed that banned books disproportionately featured characters of color in children's books, and that banned books are disproportionately written by people of color [5]. This study did not reflect these findings which could be explained by several reasons. Firstly, the dataset used for our analysis favored English written texts and was not a complete list of banned books in the US. Secondly, information about the race or identity of the book's characters was not available. It is quite possible that a more thorough examination of the existing themes in the literature would show that many of the challenged books featured characters of color.

9.3 Limitations

9.3.1 Dataset. The dataset used for this analysis is quite rudimentary. The dataset favors English written texts and does not contain every censored book. A more developed dataset with additional author information for example could lead to better and more involved author analysis.

9.3.2 Lack of Book Sales Data. Clusters No book sales data or scholarly book popularity information is readily available. Therefore, precise calculations on the disproportionality of banned books are not possible.

9.3.3 OpenLibrary.org. The website's genre classifications can be confusing and do not match the genre's provided by the dataset. This makes exact book counts for each genre impossible, and exact calculations on the disproportionality of banned books are not possible.

10 CONCLUSION

This study sets out to find genre and author trends in banned and challenged books, and to confirm if certain genres or authors are disproportionately banned using a network analysis methodology. The experiments revealed that women authors tend to be challenged more than men, while writing less books on average. Contentious genres like

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>
- [8] Open Library. 2025. Open Library: An open, editable library catalog. Retrieved April 7, 2025, from <https://openlibrary.org/>
- [9] Goodreads. 2025. Goodreads: Meet your next favorite book. Retrieved April 8, 2025, from <https://www.goodreads.com/>
- [10] Berger, E. 2025. Rightwing groups across US push new bans to limit 'obscene' books in libraries. *The Guardian*. Retrieved April 10, 2025, from <https://www.theguardian.com/us-news/2025/apr/01/new-book-bans-library-schools>

APPENDIX A

'Controversial' Community Genres:

Abuse, Adult, Australia, Basketball, BDSM, Bisexual, Boys Love, Boarding School, Canada, Chick Lit, College, Coming of Age, Contemporary, Cult Classics, Death, Disease, Drama, Enemies to Lovers, Erotica, Fandom, Gay, Gender, Harlequin, High School, Hockey, Illness, Intersex, Ireland, Israel, Lesbian, LGBT, Fae, Fiction, Film, Football, Friends to Lovers, Mental Health, M. M. Romance, New Adult, New York, Novel in Verse, Poetry, Polyamory, Prostitution, Queer, Realistic Fiction, Relationships, Road Trip, Romance, Small Town Romance, Sports, Teen, Tragedy, Transgender, Young Adult

APPENDIX B**Genre Distribution of Communities based on Book Description:**

Community	Genre Distribution
0	Fiction: 12% YA: 9% Fantasy: 6% Contemporary: 6% Romance: 5%
1	Fiction: 10% YA: 6% Children's: 6%
2	Children's: 7% Fiction: 7% Non-Fiction: 7% Historical: 7% YA: 6% Biography: 6%
3	Fiction: 12% YA: 8% Romance: 7% Fantasy: 6% Contemporary: 5%
4	Fiction: 11% YA: 9% Contemporary: 5%
5	Fiction: 13% YA: 8% Romance: 5%
6	Fiction: 11% YA: 9% Contemporary: 8% Romance: 6% Children's: 6% R. Fiction: 6%
7	Fiction: 11% YA: 8% Romance: 7% Contemporary: 6%
8	Fiction: 13% YA: 8%
9	Fiction: 10% YA: 9% Romance: 5% Contemporary: 5%
10	Fiction: 12% YA: 10% Romance: 6% Contemporary: 6% Fantasy: 6%
11	Fiction: 10% YA: 9% Fantasy: 7%
12	Fiction: 13% YA: 9% Romance: 6%
13	Fiction: 12% YA: 9% Fantasy: 7%
14	Fiction: 12% Contemporary: 9% YA: 7% Romance: 6%
15	Fiction: 10% YA: 8% Romance: 7% Contemporary: 6%
17	Fiction: 12% YA: 9% Romance: 7% Contemporary: 7%
18	Fiction: 12% YA: 10% Romance: 8% LGBT: 7% Contemporary: 6% Queer: 5%
19	Fiction: 11% YA: 9% Romance: 7% Contemporary: 5%
20	Fiction: 12% YA: 7% Contemporary: 6% R. Fiction: 5%
21	Fiction: 12% YA: 8% Audiobook: 6% Romance: 5% Contemporary: 5%
22	Fiction: 11% YA: 8% Contemporary: 5%
23	Fiction: 12% YA: 10% Romance: 6% Contemporary: 6% LGBT: 5%
24	Fiction: 10% YA: 8% Contemporary: 7% Romance: 8%
25	Fiction: 12% YA: 9% Fantasy: 7%
26	Fiction: 12% YA: 11% Romance: 6% Contemporary: 6%
27	YA: 10% Fiction: 10% Contemporary: 8% Romance: 8% LGBT: 7% Queer: 7%
28	Fiction: 13% YA: 10% Fantasy: 6% Contemporary: 6% Romance: 5%
29	Fiction: 12% YA: 8% Fantasy: 7%

30	Fiction: 12% YA: 10% Romance: 7% Fantasy: 7%
31	Fiction: 11% YA: 6%
32	Fiction: 12% Contemporary: 7% YA: 7% Romance: 8%
33	Fiction: 13% YA: 8% Contemporary: 7% R. Fiction: 7% Children's: 6%
34	Fiction: 14% YA: 9% Romance: 6% Contemporary: 6%
35	Fiction: 12% YA: 9% Historical: 7%
36	Fiction: 13% YA: 9% Romance: 7% Fantasy: 6%
37	Fiction: 11% YA: 9% Contemporary: 5%
38	Fiction: 11% YA: 9% Contemporary: 6% Romance: 6%
39	Fiction: 12% YA: 9%
40	Fiction: 13% YA: 8% Romance: 5%
41	Fiction: 11% YA: 6% Historical: 5%
42	Fiction: 12% YA: 7% Contemporary: 6%
43	Fiction: 12% YA: 9% Children's: 5%
44	Fiction: 13% YA: 12% Contemporary: 10% Romance: 7% LGBT: 5%
46	Fiction: 11% YA: 10% Romance: 8% Fantasy: 6%
47	YA: 10% Fiction: 9% Romance: 8% Fantasy: 5%
48	Fiction: 9% Children's: 8%
50	Fiction: 10% YA: 9% Romance: 6% Fantasy: 5% Contemporary: 5%
54	Fiction: 12% YA: 7% LGBT: 6% Contemporary: 6%
55	Fiction: 13% YA: 13% Romance: 8% Fantasy: 6% Contemporary: 6%
56	Fiction: 11% YA: 11% Contemporary: 6%
57	Fiction: 11% YA: 10% Romance: 6% Contemporary: 5%
58	Fiction: 13% YA: 12% Contemporary: 10% Romance: 7% LGBT: 5%

*YA: Young Adult

*R. Fiction: Realistic Fiction