Complexities Associated with User-generated Book Reviews in Digital Libraries: Temporal, Cultural, and Political Case Studies

Yuerong Hu
yuerong2@illinois.edu
School of Information Sciences, University of Illinois Urbana - Champaign
Champaign, Illinois, USA

Zoe LeBlanc
zleblanc@illinois.edu
School of Information Sciences, University of Illinois Urbana - Champaign
Champaign, Illinois, USA

Jana Diesner
jdiesner@illinois.edu
School of Information Sciences, University of Illinois Urbana - Champaign
Champaign, Illinois, USA

Ted Underwood
tunder@illinois.edu
School of Information Sciences, University of Illinois Urbana - Champaign
Champaign, Illinois, USA

Glen Layne-Worthey
gworthey@illinois.edu
School of Information Sciences, University of Illinois Urbana - Champaign
Champaign, Illinois, USA

J. Stephen Downie
jdownie@illinois.edu
School of Information Sciences, University of Illinois Urbana - Champaign
Champaign, Illinois, USA

ABSTRACT

While digital libraries (DL) have made large-scale collections of digitized books increasingly available to researchers [31, 67], there remains a dearth of similar data provisions or infrastructure for computational studies of the consumption and reception of books. In the last two decades, user-generated book reviews on social media have opened up unprecedented research possibilities for humanities and social sciences (HSS) scholars who are interested in book reception. However, limitations and gaps have emerged from existing DH research which utilize social media data for answering HSS questions. To shed light on the under-investigated features of user-generated book reviews and the challenges they might pose to scholarly research, we conducted three exemplar cases studies: (1) a longitudinal analysis for profiling the temporal changes of ratings and popularity of 552 books across ten years; (2) a cross-cultural comparison of book ratings of the same 538 books across two platforms; and, (3) a classification experiment on 20,000 sponsored and non-sponsored books reviews. Correspondingly, our research reveals the real-world complexities and under-investigated features of user-generated book reviews in three dimensions: the transience of book ratings and popularity (temporal dimension), the cross-cultural differences in reading interests and book reception (cultural dimension), and the user power dynamics behind the publicly accessible reviews ("political" dimension). Our case studies also demonstrate the challenges posed by user-generated book reviews' real-world complexities to their scholarly usage and propose solutions to these challenges. We conclude that DL stakeholders and scholars working with user-generated book reviews should look into these under-investigated features and real-world challenges to evaluate and improve the scholarly usability and interpretability of their data.

CCS CONCEPTS

• **Information systems → Digital libraries and archives**; • **Applied computing → Digital libraries and archives**.

KEYWORDS

digital libraries; digital humanities; cultural analytics; user-generated content; social media; web archives

ACM Reference Format:
Yuerong Hu, Zoe LeBlanc, Jana Diesner, Ted Underwood, Glen Layne-Worthey, and J. Stephen Downie. 2022. Complexities Associated with User-generated Book Reviews in Digital Libraries: Temporal, Cultural, and Political Case Studies. In *The ACM/IEEE Joint Conference on Digital Libraries in 2022 (JCDL '22), June 20–24, 2022, Cologne, Germany*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3529372.3530930

1 BACKGROUND AND INTRODUCTION

While collections of printed books have been developed and maintained for centuries, archives of ordinary readers' reception of and responses to books have rarely been preserved [9, 52]. As digital historians have argued, *"Historically, ordinary people did not leave behind many records, forcing historians to learn about them from the scant moments when they came into contact with large record-keeping institutions like censuses, churches, poor rolls, or the criminal-justice system"* [52]. Due to a lack of empirical and historical research evidence, many aspects of readership and reception history remained theoretical and/or anecdotal until the 21st century. In the last two decades, researchers have been uncovering historical archives [37] and developing contemporary book reception corpora in support of empirical investigations in reading behavior, reader response, reception, literary appreciation, etc [9, 19, 25]. User-generated book reviews on social reading/reviewing websites such as Amazon and Goodreads meet this need, and have opened up unprecedented research possibilities for studies in digital humanities (DH) and

JCDL '22, June 20–24, 2022, Cologne, Germany

ACM ISBN 978-1-4503-9345-4/22/06…$15.00
https://doi.org/10.1145/3529372.3530930

cultural analytics (CA) [36, 68], library and information sciences [6, 47, 73], literary history [10], computer-supported cooperative work [4, 61], social network analysis [49, 54], etc. However, research limitations and gaps have also emerged from this burgeoning research area, particularly in DH. This presented work was motivated by two emergent gaps and limitations in computational analysis of user-generated book reviews in DH.

First, although DH studies have leveraged datasets from different historical periods and sources, the alignments and comparisons are still largely limited to Anglophone materials and Western perspectives [21, 37, 45, 49, 55, 73]. At the same time, the books that have been intensively studied in DH are mostly books with distinguished popularity and prestige, such as Western classics, popular references in academia, mass market bestsellers, prize winners, etc [10, 10, 38, 49, 58, 68]. All these books are subject to selection bias and historical biases in literary history (e.g., classism, sexism, racism, colonialism), which poses questions about the inclusiveness and representativeness of prior DH studies on book reviews. For instance, in studies of the classics and bestsellers, reciprocal effects and multiplier effects have emerged where various book review platforms (e.g., Goodreads, LibraryThing), booksellers (e.g., Amazon Books, Barnes & Noble), and book impact indexes (e.g., the MLA International Bibliography, the Open Syllabus Project) seem to echo each other's opinions and "endorse" the same groups of books [5, 10, 49]. It is necessary to diversify the DH research datasets of user-generated book reviews for breaking such echo chamber.

Second, existing DH studies often treat user-generated book reviews as if they represent open, honest, independent, and democratic voices from real readers, although fake, paid, biased reviews as well as professionally written reviews have been broadly identified and studied in research areas like computer science and information systems [32, 40, 41, 71, 74]. Likewise, DH scholars rarely examine book reviews' authenticity, usefulness and other aspects of scholarly usability, despite the progress made in fake review detection and review usefulness evaluation by researchers from online marketing, decision support systems, and computer-mediated communication [32, 34, 42, 44, 46, 72]. Such simplified modeling of user-generated book reviews is understandable in some DH studies given the facts that (1) DH scholars have to selectively re-purpose social media data from commercial websites in order to answer specific humanities questions due to lack of alternative historical evidence; and, (2) DH theories, foci, and approaches are significantly different from those in other computing-centered areas. However, real-world contexts associated with review data, when downplayed or even neglected, might lead to biased assumptions and even research fallacies [56]. For instance, both research and news articles have revealed that user-generated book reviews are subject to platform algorithmic moderation, review manipulation, trolls, extortion scams, review bombing, collective fandom action, and other undesirable effects [3, 30, 41, 48, 51, 53, 68]. Nevertheless, few of these problems have been sufficiently acknowledged or addressed in prior DH studies on book reviews. In addition, although online book review platforms like Goodreads have been *expanding the authority of people otherwise marginalized by literary gatekeepers* [24], it would be an oversimplification to assume that user-generated reviews only reflect book opinions from ordinary readers. In reality, this type of publicly accessible data is produced under conditions of (1) wrestling among different book industry stakeholders (e.g., publishers, booksellers, authors, and professional literary critics) [51]; (2) fights and negotiations between the commercial platforms and their users [2]; and, (3) tensions between book authors and readers/reviewers [59, 64]. Lack of research on these real-world issues and features of user-generated book reviews can lead to false assumptions and non-contextual interpretations. Therefore, it is essential to acknowledge and unfold the real-world complexities of user-generated book reviews for more contextualized and responsible DH research.

To uncover a more in-depth understanding of the real-world challenges and nuances associated with deploying user-generated book reviews for DH research, we investigated user-generated book reviews through the lens of (1) temporal changes; (2) cross-cultural divergence; and, (3) user power dynamics. To better align our study with ongoing DH research interests and gaps, we collected data from two large-scale, frequently studied, and cross-cultural online reader communities: Goodreads based in the U.S. [27] and Douban based in China [22]. We collected the books' ratings, numbers of ratings, ranking, and textual reviews from Goodreads and Douban to explore three research questions: (1) how do book ratings and the popularity of books change over time, particularly how durable or ephemeral are ratings and popularity; (2) do users' interests in and opinions about books differ by their cultural background; and if so, what are the cross-cultural differences and divergences; and, (3) to which extent do reviews reflect an open, transparent, and democratic paradigm? In the following paragraphs, we first introduce the data sources we used and how we collected the data. Next, we present our approach to and findings from each of the three case studies. Then, we summarize our findings across the case studies and draw overarching conclusions. Finally, we discuss our research limitations and directions for future work.

2 DATA SOURCES AND ACCESS

2.1 Data Source

We surveyed prior research to identify frequently leveraged and high quality data sources for our exemplar studies. Table 1 provides a quick summary of three of the most frequently leveraged resources for research on online book reviews in English: Amazon.com: Books (Hereafter, Amazon Books), Goodreads, and LibraryThing [14, 16, 17, 27, 43, 68]. Table 1 summarizes and compares their history, user base, and prominent features. Note that in this paper, "Amazon Books" only refer to the online book selling department of Amazon.com [14], not the Amazon Books retail bookstores [13]. All three platforms allow readers to post their numerical ratings of the books on a 1-to-5-star scale along with optional textual and graphic reviews. Goodreads and LibraryThing also allow readers to add their own tags to the books, build custom virtual book collections, vote for books, join online reading discussions, etc. We chose to use Goodreads data because it is more "bookish" compared to Amazon Books [1] and has a larger user base

[1]By more "bookish" we mean that Goodreads users and their reviews are more devoted to books and reading activities. By contrast, many Amazon book ratings and reviews are based on customer services and product quality instead of the book content or the reviewers' reading experience.

Name	Founding	User Base	Selected features and activities
Amazon.com: Books	2005, US	Controls over 50% of all book distribution in the US	- The world's largest online bookseller - Multiple boycott campaigns and lawsuits regarding price-fixing
Goodreads	2007, US	90 million registered members as of July 2019	- Dominant position among US digital reading platforms - Become a subsidiary of Amazon in 2013
LibraryThing	2005, US	2.6 million users as of February 2021	- With library metadata imported and library cataloging rules applied. - Crowdsourced tags from readers, bookstore owners, and librarians - Became a subsidiary of Amazon in 2008

Table 1: Profiles of Amazon Books, Goodreads, and LibraryThing

	Founding Year	User Base	Coverage of Items	Commercial Dependency	Languages
Goodreads	2007	90 million registered members as of July 2019	Primarily books	Acquired by Amazon	Multiple languages, primarily English
Douban	2005	220 million registered members as of 2020	Heterogenous cultural products and activities	Independent	Multiple languages, primarily Chinese

Table 2: Comparison of Goodreads and Douban

compared to LibraryThing. To break the existing Anglophone "echoing chamber" in DH datasets and enable cross-cultural comparisons, we used Douban Books as the second data source. Douban is the largest social platform based in China for reviewing all categories of cultural products and activities, including books, movies, TV episodes, albums, concerts, museum exhibitions, and so on [15, 22]. With about 220 million registered users [75], Douban functions as a combination of Goodreads, LibraryThing, Rotten Tomato, and IMDb. Douban Books is the division of Douban that devotes to book reviews. Douban was chosen for two reasons. First, it is similar and comparable to Goodreads in terms of functions, size of the user base, and impact in the online book review communities, which is summarized in Table 2. As we can tell from Table 2, the main differences between Goodreads and Douban lie in their coverage of items (books only vs. books included), user communities (Western vs.non-Western) and commercial dependency (affiliated with Amazon vs. independent). Second, the user communities of Douban and Goodreads are significantly different from each other for cross-cultural comparisons. Douban users are based in *"the national boundaries of China as well as the Greater China region"* [35]. According to the profiles of a million Goodreads users [63], users based in Mainland China, Hong Kong, Macao and Taiwan only make up 0.7%, while users from the United States, United Kingdom, Canada and Australia make up 49%. Therefore, it is reasonable to assume that Douban and Goodreads data were collected from two user communities with little overlap, which makes the cross-cultural comparisons valid and meaningful.

2.2 Legal and Ethical Use of Data

Given the fact that both Douban and Goodreads have suspended their APIs, we decided to only collect a small amount of their publicly accessible data for our studies. The data we collected (1) were available to any online user without registration or logging in required; and, (2) involved no identifiable information about or collected from individual users. While we only collected the present book rating data from Douban and Goodreads, we reused several existing book review datasets. For instance, in order to recover the historical book lists on Douban, we gathered book lists archived and shared by Douban official accounts and other Douban users [7, 8, 62, 65, 66, 77, 78]. Meanwhile, to build models on the Goodreads book reviews, we reused existing open-access Goodreads datasets [69, 70]. More details about these datasets are provided in the following sections. Due to legal concerns (e.g., risks of copyright infringement) and ethical considerations (e.g., protecting the readers' privacy and their intellectual freedom from unwanted attention and surveillance), we decided to not publicly share the raw data we scraped or republish any data extracted from existing datasets. Instead, we decided to share the Douban and Goodreads URLs used as handles for retrieving the data collected. These URLs would enable other researchers to examine the data we used and conduct their own research on the same books. Sharing the URLs instead of the data we scraped would also allow the copyright owners (e.g., the platforms, the users, the dataset curators) to update and/or delete the content they created at any time. With these decisions made, our access to and use of the user-generated data should (1) be small-scale, research-only, transformative, and non-consumptive [1, 11, 20]; and (2) have minor impact on the potential market for or value of the copyrighted materials held by Goodreads/Douban. Therefore, our case studies should not violate (1) the copyright laws or the Computer Fraud and Abuse Act; (2) the platforms' terms of services or instructions specified in their robots.txt files; nor (3) the privacy or rights of the users.

3 DESIGNS AND WORKFLOWS OF THE THREE CASE STUDIES

To realize our overarching research goals, we designed three case studies to explore the temporal, cultural, and political dimensions of user-generated book reviews, respectively. For each case study, we first introduce the specific research questions and the research design for answering them. Next, we describe how we collected the datasets needed, particularly how we leveraged our domain knowledge in book history and readership to clean, correct, align,

and annotate the dataset. Finally, we combined a variety of computational techniques to analyze the dataset we built. In particular, we adopted the measurements and models that have been frequently used in recent DH and CA research to engage with the latest discussions.

3.1 Temporal Dimension: Longitudinal Analysis

This case study was driven by three research questions: (1) how do book ratings change over time; (2) how durable is the popularity of books; and, (3) under which circumstances are the ratings and popularity of books most transient or durable. These questions were prompted by how ranked lists of cultural products are used in research for mapping cultural evolution and trends. For instance, in DH studies, scholars take ranked lists of well-received and/or high-rated cultural products as credible representations of ordinary people's cultural interests and the products' social-cultural impacts, such as the Billboard Hot 100 (for popular music) and Amazon Bestsellers (for books) [49, 50]. However, impacts, popularity and commercial success of books were not enduring or universal. Readers' feelings and opinions about books change over time, and yesterday's banned books can be today's classics [26]. Therefore, when it comes to studying user-generated ranked book lists, we want to start from their temporal contexts: how durable is the ranked list?

We decided to use Douban Top 250 Books List for this case study on durability. Douban Top 250 Books List is considered one of the most frequently cited book index curated on Douban Books[29]. It presents real-time rankings, average ratings, and numbers of ratings of the 250 most popular and highly-rated books on Douban. The list was automatically generated and updated based on cumulative and crowdsourced ratings and reviews contributed by Douban users. According to Douban's official statements, their algorithms have been updated several times for (1) filtering out fake and suspicious reviews; and, (2) for re-balancing the impact of heterogeneous factors (commercial promotion, participatory culture, fandom activities, etc.) [7, 8]. This list was selected because it was the only "recoverable" Douban book list with sufficient historical data archived. The other Douban book lists were either (1) discrete and non-overlapping (e.g., the Most Popular Books in 2021); or, (2) too small for systematic analysis (e.g., the Top Ten Most Popular American Literature).

Since Douban only shows real-time data and does not keep track of the historical changes, we recovered the historical Douban Top 250 Books lists from 2011 to 2021 by consulting as many existing references as possible, including Douban users' book lists, scraped datasets of the Douban Top 250 Books, news articles in different years with the lists, etc. We collected the list in 2021 by scraping a limited amount of publicly accessible data. Through rigorous comparison and multiple rounds of data cleaning (deduplication of volumes, normalization of book titles, alignment of different text encodings, etc.), we curated a longitudinal dataset of Douban Top 250 Books lists with eight "snapshots" of the list in 2011, 2013, 2016, 2017, 2018, 2019, 2020, and 2021 [2]. The items we collected included book metadata, rankings, ratings, numbers of ratings, numbers of reviews, and crowdsourced tags. Multiple versions/editions of the same book were counted as one unique book in the aggregated book list. Through longitudinal analysis of the dataset, we found that among all of the 552 unique books that had been on the Douban Top 250 Books List at least once from 2011 to 2021, 58% (n=321) showed up less than three times. Figure 1 presents more information about the occurrence of each book over time. According to Figure 1, 76 (14%) books appeared on the list once and 215 twice (39%). It is to be noted that there were 22 books being on the list over eight times in the eight years. This happened because more than one edition/volume of the same book was on the list in the same year. Douban Books did not distinguish or deduplicate such occurrences.

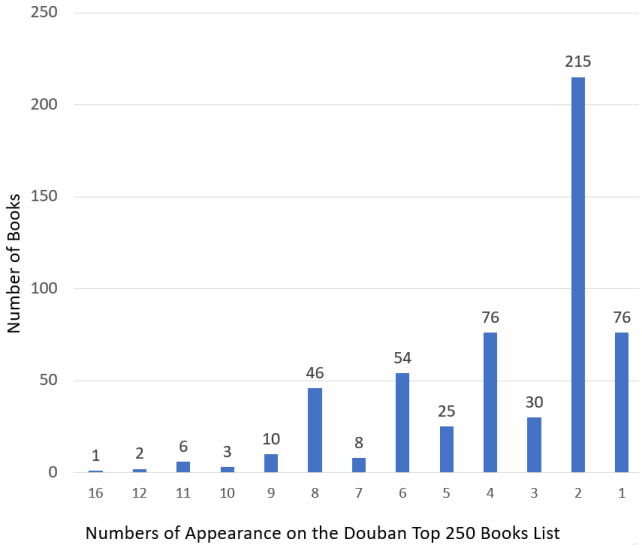

Figure 1: Numbers of Appearance on the Lists of Five Different Years.

Item	Change of Ranking	Change of Rating	Change of Number of Ratings
Mean	85.28	0.081	114282.50
Min	9.00	0.000	16707.00
Max	214.00	0.500	574277.00
Std	50.35	0.089	108148.23

Table 3: Changes in Rankings, Ratings, and Numbers of Ratings (based on data about the 64 unique books that stayed on the Douban Top 250 Books List from 2013 to 2021).

Since the lists were "crowdsourced", the items collected varied individually and were not always comparable. To enable an analysis of continuous and comparable data, We selected five of the eight lists (2013, 2016, 2017, 2018, 2021) that recorded continuous data of book ratings, and the numbers of ratings. Figure 2 visualizes the overlap of the five lists. In Figure 2, each colored oval represents a set of 250 listed books in one particular year. The number in each colored block (with one color or overlapping colors) represents the number of overlapping books. For instance, the five ovals overlap with each other at the center of Figure 2. The "73" in the center means that there are 73 books that stayed on the Douban Top 250 Books List from 2013 to 2021. After deduplication, the 73 volumes pointed us to 64 unique books. This happened because Douban

[2] we did not find any complete book lists for the years of 2012, 2014 and 2015

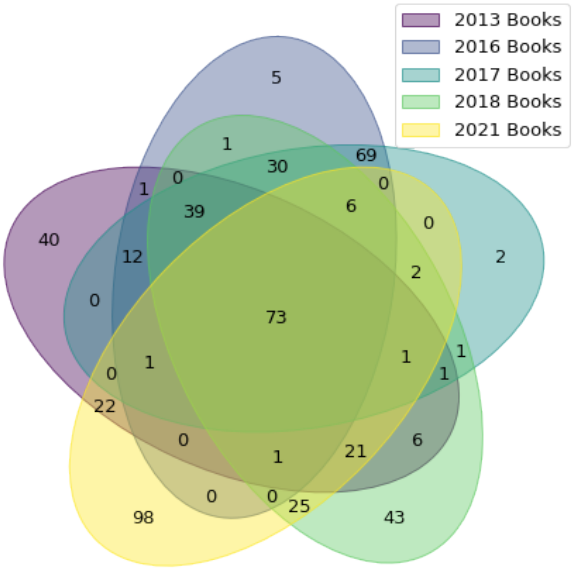

Figure 2: Overlap of the Douban Top 250 Books List across Time (based on the lists collected in 2013, 2016, 2017, 2018, and 2021).

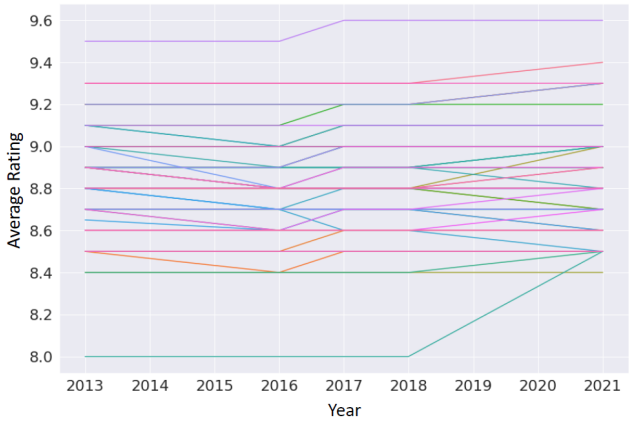

Figure 3: Changes in Ratings across Time (based on data about the 64 unique books that stayed on the Douban Top 250 Books List from 2013 to 2021).

used to have different versions/editions/volumes of the same book on the list in the same year. For a book in this case, we chose to study its most rated version/edition/volume. We considered these 64 books a collection of books with durable popularity and quantified their temporal changes. Figure 3 visualizes the changes in the 64 books' overall ratings and Figure 4 visualizes the changes in their numbers of ratings. Both visualizations were based on real-world data from 2013, 2016, 2017, 2018, 2021, with the data generated by normalization for the years in between. Please note that multiple lines overlapped in Figure 3 and Figure 4, which makes the data look sparse. To fill the visual gaps, Table 3 provides a brief statistical summary of the changes. We can tell from figure 4 and Table 3 that the ratings of these books remain constantly high with little

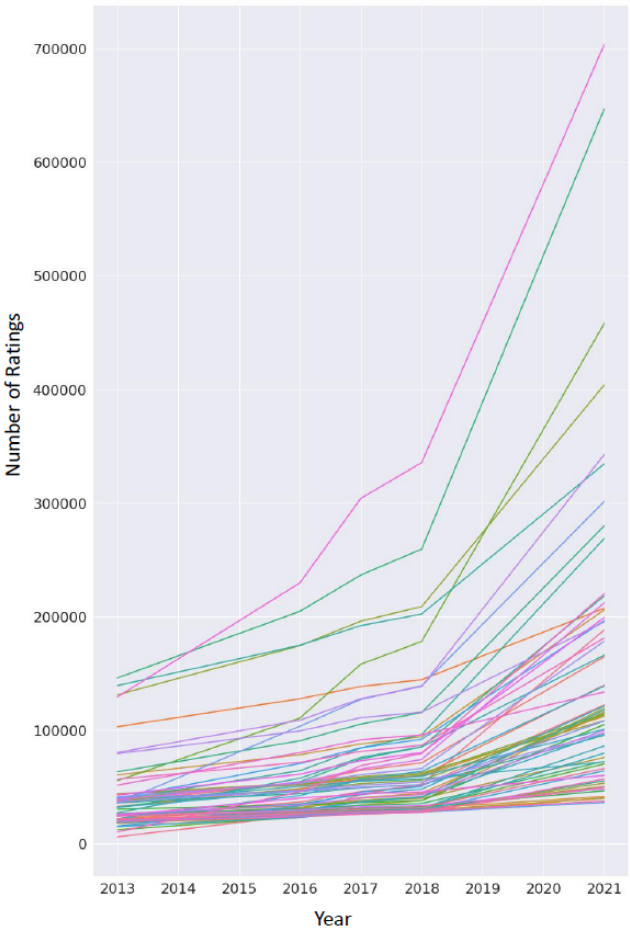

Figure 4: Changes in Numbers of Ratings across Time (based on data about the 64 unique books that stayed on the Douban Top 250 Books List from 2013 to 2021).

fluctuations (the mean value of changes in ratings in 0.081/10 with an overall standard deviation of 0.089/10), which means their high ratings have been very stable. Meanwhile, the ratings these books received have been increasing with varying speeds. According to Figure 4 and Table 3, some of the books that started with larger numbers of ratings have been gaining ratings faster than the other books with accelerated speeds. Table 3 shows that on average, each book gained 114,283 more ratings from 2013 to 2021, with a maximum increase of 574,277 ratings and a minimum increase of 16,707. Table 3 also summarizes the changes in rankings. The biggest change in a book's position on the ranked list is 214 while the smallest change is 9. The average change in ranking is 85.28, with a standard deviation of 50.35.

3.2 Cultural Dimension: Comparative Analysis

This case study is motivated by the gap in existing research where only Western books and readership have been studied. We wanted to know how users' reading interests and opinions might differ by cultural background. For comparison, we built a parallel dataset of book ratings of the same books across Douban and Goodreads.

Given our aims of expanding the existing research scope and breaking the anglophone echoing chamber, we focused on the 552 books on the Douban Top 250 Books List from 2011 to 2021. The next thing was to find the web pages for these books on Goodreads for comparison. Through our preliminary investigation into the settings of Douban and Goodreads, we found that both platforms were not able to correctly identify or link all editions/volumes of the same books, which disabled comparisons of "complete sets" of the same books. Therefore, we decided to use the most-rated/most-reviewed edition of the same book on each platform for comparison.

In terms of measurements, we focused on differences and divergences of the ratings across cultures. We quantified the differences by subtracting the Goodreads ratings from their corresponding Douban ratings. For divergences, we calculated the Kullback–Leibler divergence (K-L divergence) between the parallel distributions of 1-star to 5-star ratings on Douban and Goodreads [12, 39]. In our research contexts, we considered the distances of the distributions of ratings an important factor, second to the differences in the overall ratings. Figure 5 explains why rating divergence matters through an example of *Lolita* (snapshots accessed from Goodreads and Douban on Jan 17, 2022) [23, 28]. This book had similar overall ratings on the two platforms, but their percentages of 1-star to 5-star ratings were quite different. On Goodreads (see the upper snapshot), *Lolita* got a 3.88 out of 5 with nearly 70% of positive reviews: 35% of 5-star ratings and 32% 4-star ratings. On Douban (see the lower snapshot), it got a 7.7 out of 10 (3.85 out of 5), however, it only had 25% of 5-star reviews. Most raters on Douban gave it 4 stars (43.8%) and 3 stars (27.4%). While the overall ratings on Goodreads and Douban were very close (a difference of 0.03 out of 5), the different distributions of ratings suggested that the two groups of readers "embraced" the book differently. Given such cases, we believed the full distribution of ratings was just as important as its mean value for our comparisons.

Figure 5: *Lolita*'s Rating Pages on Douban and Goodreads.

To build the dataset, we manually searched for the Goodreads matches for the 552 selected books. We did that manually because the matching processing involved a number of difficult cases in multiple languages. For instance, there are different books written by different authors with the same book title, whereas some books have several different titles in one language due to translation differences. We had to manually differentiate such cases from the "same book, varied version/edition" cases with our library sciences expertise. Another reason for manual processing was that we were not able to generate Goodreads URLs based on English book titles as other researchers did [76]. We had experimented with this method, which, however, only worked for the books with especially well-accepted versions/editions where the best match would pop up as the first item retrieved. In our tests, the Goodreads URLs generated with book titles were not consistent. Then, With the parallel Goodreads and Douban URLs we manually paired, we scraped a small amount of public data in need for comparison, including book titles, author names, languages, ratings, numbers of ratings, etc. We found some inaccurate metadata in the scraped datasets, so we manually corrected and cleaned them. Additionally, We converted the Douban overall ratings on a 2-10 scale to ratings on a 1-5 scale. While both Douban and Goodreads allow their users to rate the books on a 1-to-5-star basis, the overall ratings generated on Douban were on a 2-10 scale while those on Goodreads were on a 1-5 scale. Therefore, to align and compare the overall ratings across platforms, we have to "normalize" the ratings. We verified that both the Goodreads and Douban overall ratings were the weighted means of the 1-to-5 score ratings. For all the books in this parallel dataset, the correlation between Goodreads overall 1-5 ratings and the weighted average values is 0.9996, and the correlation between Goodreads overall 2-10 ratings and the weighted average values is 0.9982. Therefore, the Douban ratings on a 2-10 scale can be converted to ratings on a 1-5 scale without distorting the real-world distributions of the ratings the books got (for instance, from 6/10 to 3/5). Similar conversion was also adopted in prior research hong2017empirical.

Next, we manually added cultural tags to the books since this parallel dataset was created for studying cross-cultural reception. In this case study, cultural identity was defined as "the language of the first publication of the book" which pointed to its earliest and primary group of readers. We did not use (1) the authors' motherland; (2) the author's country of citizenship; (3) the language used for writing the books; or, (4) the language of the book on the scraped page. The scraped language tag was not reliable for determining the work's cultural identity because it only described the language of the specific edition/volume. For authors, their cultural and political identities can be dynamic and controversial, and do not always align with the language of their publications. For instance, Milan Kundera, the famous Czech writer, became a French citizen in 1981 after his Czechoslovak citizenship was revoked in 1979. Later in 2019, he got Czech citizenship. However, most of his well-known books gain popularity as works published in French and he personally *"sees himself as a French writer and insists his work should be studied as French literature and classified as such in book stores"* [18]. Such examples show how political and cultural identities can be complex for authors. Since our research was focused on book reception rather than authorship, we stuck to the language of the first publication of each book for it is more likely to be aligned with the cultural identities of the first intended readership. For instance, Milan Kundera's *The Unbearable Lightness of Being*, which

was written in Czech but first published in French, was tagged as a "French" book in our dataset.

A few of the selected books were not available on both platforms, so we ended up with a parallel dataset of 538 pairs of books. We calculated the differences and divergences between the ratings, and compared their numbers of ratings. Table 4 presents the differences in overall ratings, aggregated by each book's language of the first publication and sorted from the largest to the smallest. For instance, according to the first line, this dataset had only one book in Greek and there was a difference of 0.45/5 between its overall rating on Goodreads and that on Douban. According to Table 4, the fewer samples provided in a language, the larger the average difference. The maximum differences came from books in Japanese (1.9 out of 5) and Chinese (1.15 out of 5). Table 5 shows the divergences of the ratings aggregated by each book's language of the first publication, sorted from the largest to the smallest. For example, according to the first role of Table 5, there were 30 books in Japanese and their mean divergence across the two platforms was around 0.78. Overall, the distributions in Table 5 diverged the most on the Japanese and Chinese works. The maximum divergence values also emerged from Japanese and Chinese books (19.27 for Japanese and 13.69 for Chinese), which were higher than the third-largest divergence value for English (0.48 for English). Table 6 summarizes the differences in the number of ratings a book got across the two platforms, aggregated by each book's language of the first publication and sorted by the number of books in each language. Positive values mean that Douban has larger numbers of ratings than Goodreads, while negative values point to the opposite. All the numbers in Table 6 are absolute numbers of ratings, which means the differences in Douban and Goodreads' use bases should be considered in data interpretation. Even though Douban has a smaller user based than Goodreads, the books that were first published in Japanese, Chinese, Bengali, and Danish still have received more ratings on Douban than Goodreads. Meanwhile, the books published in other languages have accumulated far more ratings on Goodreads than on Douban. Such contrast suggests differences in readers' cultural interests across the two platforms.

First Publication Language	Count	Mean	Std	Min	Max
Greek	1	0.45	nan	0.45	0.45
Portuguese	2	0.38	0.02828	0.36	0.4
Czech	1	0.37	nan	0.37	0.37
Swedish	1	0.37	nan	0.37	0.37
Spanish	5	0.336	0.24996	0.02	0.58
Arabic	1	0.33	nan	0.33	0.33
Italian	9	0.32556	0.16372	0.08	0.52
Danish	3	0.32333	0.25502	0.07	0.58
Norwegian	1	0.31	nan	0.31	0.31
German	13	0.3	0.1193	0.11	0.46
French	20	0.272	0.15178	-0.06	0.62
Japanese	80	0.23962	0.29469	-0.2	1.9
Russian	6	0.225	0.24664	-0.06	0.54
Bengali	2	0.21	0.08485	0.15	0.27
English	128	0.19219	0.20595	-0.54	0.85
Chinese	263	0.11875	0.27098	-0.8	1.15
Hebrew	2	0.075	0.12021	-0.01	0.16

Table 4: Differences in Book Ratings across Douban and Goodreads

First Publication Language	Count	Mean	Std	Min	Max
Japanese	80	0.77858	2.80639	0.0048	19.27238
Chinese	263	0.56508	1.77565	0.00715	13.68772
Greek	1	0.18964	nan	0.18964	0.18964
Danish	3	0.17926	0.1099	0.08914	0.3017
Swedish	1	0.16512	nan	0.16512	0.16512
Arabic	1	0.16036	nan	0.16036	0.16036
Italian	9	0.15999	0.12171	0.02022	0.33054
Portuguese	2	0.15568	0.01681	0.1438	0.16757
French	20	0.15271	0.09523	0.03796	0.42416
Czech	1	0.14566	nan	0.14566	0.14566
Spanish	5	0.14004	0.10393	0.02125	0.24761
German	13	0.13019	0.06945	0.02022	0.24213
Russian	6	0.12565	0.08051	0.01648	0.24249
Norwegian	1	0.11668	nan	0.11668	0.11668
English	128	0.10964	0.08504	0.01127	0.48123
Bengali	2	0.07955	0.02483	0.062	0.09711
Hebrew	2	0.03283	0.01493	0.02227	0.04339

Table 5: Divergences of Book Ratings across Douban and Goodreads

3.3 Political Dimension: Text Classification

Our question for the third case study was whether user-generated reviews reflected an open, transparent, and democratic paradigm. More specifically, we wanted to explore if sponsorship would change the nature of the reviews. Sponsorship comes from various stakeholders of the book industry, such as publishers, book review platforms, booksellers, authors, etc. Sponsorship is often detectable in the review texts. For instance, many readers who accepted free and/or advanced copies from a third party had to write reviews in return for the sponsorship. Typical sponsored reviews come with explicit claims like "*"I received this book from the X via Y [anonymized] to read and review*" and "*I received this book in exchange for an honest review, this has not altered my opinion*". While these reviewers frankly acknowledged the existence of sponsorship, they also denied the sponsorship's impacts on their reviews. We intended to investigate the power dynamics and relationships between sponsored and non-sponsored reviews through two questions: (1) if sponsored reviews would gain more visibility than the non-sponsored ones; and, (2) if the sponsored reviews significantly differed from the non-sponsored ones in terms of content. To answer the first question, we quantified the proportions of the two types of reviews. The second question was investigated through text classification. We experimented with a variety of text classifiers to see if they could easily tell the difference between the two types of reviews [33].

To build the training and test sets for classification, we reused a public research dataset of millions of Goodreads book reviews called the UCSD Book Graph [69, 70]. It contains (1) eight genres of book reviews scraped from Goodreads, which were posted from

First Publishing Language	Count	Mean	Std	Min	Max
Chinese	263	60,035.36	74,306.79	-372,200	620,027
English	128	-720,383	1,216,113	-7,835,500	166,919
Japanese	80	28,546.65	119,798.4	-264,932	683,579
French	20	-154,812	314,847.1	-954,624	100,070
German	13	-154,178	249,309.8	-702,545	94,463
Italian	9	-41,134.7	54,274.76	-166,168	3,005
Russian	6	-269,755	351,413.1	-710,927	61,232
Spanish	5	-160,009	199,780.6	-473,618	23,827
Danish	3	2,286	9,550.524	-8,420	9,930
Bengali	2	32,765	33,716.27	8,924	56,606
Hebrew	2	-337,822	272,715.5	-530,661	-144,983
Portuguese	2	-1,234,104	1,444,424	-2,255,466	-212,742
Arabic	1	-24,113	nan	-24,113	-24,113
Czech	1	-4,337	nan	-4,337	-4,337
Greek	1	-156,954	nan	-156,954	-156,954
Norwegian	1	-116,813	nan	-116,813	-116,813
Swedish	1	-170,649	nan	-170,649	-170,649

Table 6: Differences in Numbers of Ratings across Douban and Goodreads

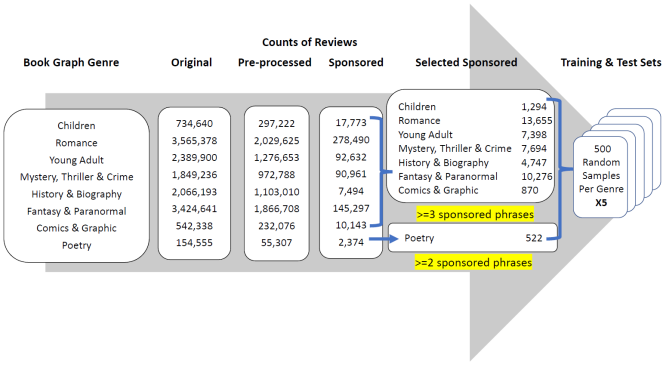

Figure 6: Data Processing Workflow and Counts of Reviews.

2006 to 2017; and, (2) book metadata and author information. First, we aligned all reviews with the book metadata and the author information for validation. Then, we filtered the aligned reviews by dropping non-English reviews and the reviews that were too short (less than 50 characters) to be semantically meaningful. Next, based on close reading and manual annotation of a sample of sponsored reviews, we developed a small dictionary of typical "sponsored reviews" phrases. We used this dictionary to computationally identify all the "candidates" of sponsored reviews who had at least one phrase from the dictionary. To find the "true positive" sponsored reviews, we further filtered the candidate sponsored reviews to find the ones containing no less than three sponsorship phrases. This threshold was chosen based on our experiment with a small sample of hundreds of reviews. We manually identify the true positive "sponsored" reviews among all the potential ones, and we found that three sponsorship phrases were stronger signals for positive true cases. The only exception was poetry reviews: we kept the poetry reviews with no less than two phrases from the dictionary because (1) there were only a small number of poetry reviews available for filtering; and, (2) poetry reviews were less likely to involve sponsorship phrases. Figure 6 shows our workflow for processing the existing dataset, the counts of reviews after each step, and the distributions of reviews by genre. According to our observation, around 8% of the scraped reviews were potentially sponsored.

Finally, we randomly sampled 500 potentially sponsored reviews from each genre and randomly sampled another potentially 500 non-sponsored reviews from each genre to construct a balanced training and test dataset for classification. We conducted the sampling five times, which produced 5 training and test datasets, each with 8000 reviews. We built a series of classifiers that were commonly used in text classification tasks with features generated from the review texts. The models we used were Decision Tree, KNeighbors, Naïve Bayes, Logistic Regression, Support Vector Machine, Random Forest, and Xtereme Gradient Boosting classifiers imported from scikit-learn (sklearn), a machine learning library in Python [57]. The features included counts of word vectors, word-level TF-IDF, ngram-level TF-IDF and character-level TF-IDF generated with sklearn vectorizers [57, 60].For comparison, we removed the sentences that contained the "sponsored review" phrases from the review texts to build another 5 training and test datasets. We called the second 5 sets of book reviews without the explicit "sponsored review" sentences "filtered reviews". Table 7 summarizes the classifiers' performances built on each combination of features and model on the two sets of training and test sets. According to Table 7, most classifiers (except for the KNeighbors models) can easily distinguish the "sponsored" reviews from the non-sponsored ones (average accuracy over 90%, up to 99.7 %). However, once we removed the explicit "sponsorship" claims, the overall accuracy of the models drastically dropped to around 50%, which suggests that our models could no longer differentiate the two types of reviews.

4 FINDINGS AND DISCUSSIONS

4.1 Temporal Dimension: Durability of Book Ratings and Popularity

Our longitudinal analysis of the 552 books on the Douban Top 250 Books List indicates that popularity does not last for most books and the ranked lists based on crowdsourced ratings change constantly. For the small group of 64 books with durable popularity, while their ratings remain high and stable, the attention they get vary drastically across time. Therefore, digital librarians and researchers who work with such datasets should specify what snapshots are used and which periods the datasets truly represent. They should also provide more information about the historical contexts and temporal durability of the ranked lists for potential users of the datasets. Developers of such lists should (1) clearly communicate the timeliness of their datasets; and, (2) consider providing longitudinal archives of the changing lists for better data transparency and interpretability.

4.2 Cultural Dimension: Differences and Divergences in Interests and Opinions

With parallel data about 538 pairs of books collected from Douban and Goodreads, we quantified the differences and divergences in cross-cultural ratings of the same books. As far as we know, this is so far the first computational and comparative case study on Western

Features	Model	Average Accuracy (original reviews)	Standard Deviation of Accuracy (original reviews)	Average Accuracy (filtered reviews)	Standard Deviation of Accuracy (filtered reviews)	Average Change of Accuracy (original reviews VS filtered reviews
Count Vector	Decision Tree	0.974	0.001	0.500	0.011	-0.475
	KNeighbors	0.819	0.005	0.503	0.008	-0.315
	Naïve Bayes	0.788	0.014	0.504	0.014	-0.284
	Logistic Regression	0.978	0.005	0.509	0.013	-0.469
	Support Vector Machine	0.956	0.003	0.504	0.005	-0.452
	Random Forest	0.941	0.005	0.503	0.009	-0.438
	Xtereme Gradient Boosting	0.988	0.002	0.503	0.008	-0.486
Word-level TF-IDF	Decision Tree	0.969	0.004	0.504	0.016	-0.465
	KNeighbors	0.578	0.029	0.502	0.012	-0.075
	Naïve Bayes	0.865	0.006	0.503	0.011	-0.361
	Logistic Regression	0.943	0.007	0.504	0.012	-0.439
	Support Vector Machine	0.958	0.005	0.502	0.010	-0.456
	Random Forest	0.979	0.002	0.504	0.006	-0.476
	Xtereme Gradient Boosting	0.989	0.003	0.505	0.008	-0.484
Ngram-level TF-IDF	Decision Tree	0.990	0.003	0.502	0.003	-0.488
	KNeighbors	0.575	0.146	0.505	0.013	-0.070
	Naïve Bayes	0.952	0.003	0.507	0.016	-0.446
	Logistic Regression	0.986	0.003	0.510	0.011	-0.475
	Support Vector Machine	0.988	0.002	0.511	0.010	-0.477
	Random Forest	0.992	0.001	0.504	0.017	-0.489
	Xtereme Gradient Boosting	0.997	0.001	0.507	0.013	-0.490
Character-level TF-IDF	Decision Tree	0.955	0.007	0.500	0.011	-0.455
	KNeighbors	0.539	0.016	0.502	0.012	-0.037
	Naïve Bayes	0.892	0.036	0.500	0.014	-0.392
	Logistic Regression	0.942	0.004	0.499	0.014	-0.444
	Support Vector Machine	0.957	0.003	0.498	0.010	-0.458
	Random Forest	0.973	0.003	0.497	0.011	-0.477
	Xtereme Gradient Boosting	0.986	0.003	0.498	0.013	-0.488
Average	-	0.909	0.011	0.503	0.011	-0.406

Table 7: Performances of the Classifiers on Book Reviews with and without Explicit Sponsorship Claims

and non-Western readership based on user-generated data. In terms of reading interests, we found that the Goodreads users have been less interested in non-Western books compared to reviewers on Douban. As for readers' opinions about the same books, the parallel ratings diverge most for books that Goodreads users rated less frequently, particularly the Japanese and Chinese works. Many non-Western canonical books and/or books popular with Douban users are seldom rated on Goodreads. Given these differences, we urge future DH and CA studies to (1) specify culture-dependency of user-generated book reviews when making claims; and, (2) diversify data sources for more inclusiveness and broader representativeness. Meanwhile, DL stakeholders are encouraged to collect data from (1) the communities they are most familiar with for more cultural contexts and better interpretability; and/or (2) a wide variety of communities for comprehensive perspectives.

4.3 Political Dimension: User Power Dynamics

Through text classification of sponsored and non-sponsored book reviews (with and without explicit sponsorship claims), we found that around 8% of the reviews scraped from Goodreads are potentially sponsored. However, to the classifiers, the main differences between the sponsored and the non-sponsored reviews are sentences that explicitly claim the existence of the sponsorship. Once such explicit claims were removed, the two types of reviews became hard to differentiate (a drastic drop of 40% in the average classification accuracy). This finding suggests that for our case study, sponsored reviews were not significantly affected by the sponsorship. While our findings are subject to further examination with respect to other features and factors (e.g., ratings of the reviews, the books that got sponsored), they remind us that user-generated book reviews come from reviewers with varying motivations, backgrounds, platform-wise activities, etc. Certain voices might be sponsored by less visible agents, even though such voices might not be prioritized on a given platform. For future research, power dynamics and relationships between different groups of users should be investigated to detect and address any potential biases in the data.

4.4 Discussions and Limitations

Our empirical investigations illuminated under-examined features of user-generated book reviews, however, our exploration and findings are preliminary. We would like to clearly communicate our research limitations to our readers as well as future developers/users of similar datasets. First, all presented case studies were based on data collected from Goodreads and Douban, which might entail platform-based biases, limitations, and data peculiarities to our research. For instance, both platforms lacked high-quality bibliographic control, which disabled the aggregation of book rating data over all editions or versions of the same book. Meanwhile, the data were not always comparable due to the platform-wise differences in terms of user base, items provided, etc. Therefore, the parallel dataset we built only covered a small amount of data from the two platforms. Besides, both Goodreads and Douban are commercial websites. When re-purposing their data for our research, we did not aim to identify or remove the impact of commercial factors that are

embedded in the platforms' designs (e.g., how did advertisements on the platforms affect the book ratings). Second, our research methods and findings need further improvement and verification. Taking the third case study as an example, all reviews were selected from an existing dataset, which was a convenient and imbalanced sample of Goodreads reviews. Given the scraped reviews' imbalanced distribution in the eight genres, the 522 sponsored poetry reviews had to be reused in the five random samples. In addition, our lexicon-based approach for identifying the sponsored reviews was deliberately streamlined, and our feature engineering might not be the best choice for this dataset. These might be the causes of both the particularly high and low performances before and after removing the explicit sponsorship sentences. As a result, although clear patterns emerged from our text classification, further experiments and error analysis are in need for verifying and interpreting these patterns.

However, the point of our research was neither to find "ubiquitous patterns" with representative samples nor to draw "globally true" conclusions about book reception. Instead, we aimed to empirically explore the understudied complexities of user-generated book reviews through preliminary and exemplar studies. For instance, our second case study revealed the differences in Goodreads and Douban readers' interests and reception of the same books. Although it didn't tell us whether such differences would remain the same across other book collections and/or readerships, it demonstrated that user-generated reviews diverged in cultural dimensions, and enriched the social-cultural perspectives of current DH research on book reviews. We expect more case studies with other datasets to cross-examine the findings of our case studies.

5 CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Through three case studies, our research empirically explored the under-examined features of user-generated book reviews through the lens of temporal changes, cross-cultural divergence, and user power dynamics. Our case studies also demonstrate the challenges posed by user-generated book reviews' real-world complexities to their scholarly usage, and propose solutions to these challenges. With data collected from Goodreads, which is based in U.S., and Douban, which is based in China, we investigated (1) the durability of book ratings and their popularity; (2) the cross-cultural differences and divergences in users' reading interests and opinions; and, (3) the user power dynamics reflected in sponsored and non-sponsored reviews. Our findings shed light on the emergent gaps between (1) user-generated book reviews scraped from social reading platforms; and, (2) curated web archives ready for scholarly use in DH. We urge future researchers to examine the longitudinal durability, culture dependency, user power dynamics and other features of their research data obtained from social reader platforms for a more critical and contextualized understanding of the dataset. Meanwhile, we suggest that digital library stakeholders who curate such datasets should specify the temporal contexts and cultural representativeness of their data provisions. We also recommend collecting user-generated data from diverse sources for better inclusiveness and broader representativeness in DH research. Last but not least, our strategies and workflows for developing the datasets demonstrated how digital library professionals' domain expertise in web archives, book history, readership, bibliographic control, etc., help with curating and utilizing user-generated book reviews. With joint efforts from library professionals and humanities/social sciences scholars, we will be able to critically address the complexities of user-generated book reviews and improve their usability in future research.

5.2 Future Work

Regarding future work, we plan to further the three case studies by addressing the aforementioned limitations, optimizing their research designs, and scrutinizing our existing findings. For instance, we can examine (1) if the books' ratings, numbers of ratings, and durability of their popularity are correlated; and, (2) how the durability of ratings and reading interests vary across readerships and platforms. With supplementary data provisions from other sources, we might be able to identify some external factors (e.g., film adaptions, celebrity recommendations) and industry-wise mechanisms (e.g., the long tail effects, online word of mouth) behind the fluctuations in book ratings and popularity. Second, we shall develop more use cases to deepen and expand our cross-cultural comparisons. For instance, we can (1) use book lists on Goodreads to create more parallel instances; and, (2) look into the books that were only available on Douban or Goodreads but not both to understand why certain books turned out to be exclusively available/popular among certain reading communities. We can leverage data obtained from more platforms and readership communities, especially the ones that are not primarily in English or Chinese, to reflect on the patterns identified and enrich our research agenda. Third, we should further our analysis of sponsored and non-sponsored reviews. For instance, we can inspect the books associated with large numbers of sponsored reviews (e.g., What are their genres? Which publishers the sponsored reviews are most frequently associated with?) to illuminate any implicit relations between the review sponsorship and the book industry. At the same time, we are still working on the datasets we've created and reused. As aforementioned, our presented research is preliminary, and we are still enriching and analyzing the datasets. For the sake of research transparency, we've (1) shared all the resources we leveraged; and, (2) released the Goodreads and Douban URLs we curated on our own as data retrieval handles for other researchers to look into [3]. We will keep releasing and updating our datasets and scripts in this repository.

REFERENCES

[1] Parsers "admin". 2020. US court fully legalized website scraping and technically prohibited it. https://parsers.me/us-court-fully-legalized-website-scraping-and-technically-prohibited-it/
[2] Anne-Mette Bech Albrechtslund. 2017. Negotiating ownership and agency in social media: Community reactions to Amazon's acquisition of Goodreads. *First Monday* (2017).
[3] Maria Antoniak and Melanie Walsh. 2020. The Crowdsourced "Classics" and the Revealing Limits of Goodreads Data. http://dx.doi.org/10.17613/7k61-eg23
[4] Maria Antoniak, Melanie Walsh, and David Mimno. 2021. Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.
[5] Tong Bao and Tung-lung Steven Chang. 2014. Why Amazon uses both the New York Times Best Seller List and customer reviews: An empirical study of multiplier effects on product sales from multiple earned media. *Decision Support Systems* 67 (2014), 1–8.

[3] https://github.com/Yuerong2/JCDL2022ResearchPaperData

[6] Peishan Bartley. 2009. Book tagging on LibraryThing: how, why, and what are in the tags? *Proceedings of the American Society for Information Science and Technology* 46, 1 (2009), 1–22.
[7] Douban Books. 2019. How many Douban Top 250 Books have you read? https://mp.weixin.qq.com/s?__biz=MzAwNzYyNDMyMA==&mid=2651117440&idx=1&sn=86f24dcbc54b18c40978ce325fbefb08
[8] Douban Books. 2020. Big changes to Douban Top 250 Books: 107 new books on the list for the first time. https://mp.weixin.qq.com/s/iYCf7lGdLkgNurzv_HNa-Q
[9] Peter Boot. 2013. The desirability of a corpus of online book responses. In *Proceedings of the Workshop on Computational Linguistics for Literature*. Association for Computational Linguistics (ACL), 32–40.
[10] Karen Bourrier and Mike Thelwall. 2020. The social lives of books: Reading Victorian literature on Goodreads. *Journal of Cultural Analytics* 1, 1 (2020), 12049.
[11] HathiTrust Research Center. 2017. HathiTrust Research Center Non-Consumptive Use Policy. https://www.hathitrust.org/htrc_ncup
[12] Kent K Chang and Simon DeDeo. 2020. Divergence and the Complexity of Difference in Text and Culture. *Journal of Cultural Analytics* 4, 11 (2020), 1–36.
[13] Wikipedia contributors. 2021. Amazon Books. https://en.wikipedia.org/wiki/Amazon_Books
[14] Wikipedia contributors. 2021. Amazon (company. https://en.wikipedia.org/wiki/Amazon_(company)
[15] Wikipedia contributors. 2021. Douban. https://en.wikipedia.org/wiki/Douban
[16] Wikipedia contributors. 2021. Goodreads. https://en.wikipedia.org/wiki/Goodreads
[17] Wikipedia contributors. 2021. LibraryThing. https://en.wikipedia.org/wiki/LibraryThing
[18] Wikipedia contributors. 2021. Milan Kundera. https://en.wikipedia.org/wiki/Milan_Kundera
[19] Lianbin Dai. 2017. *From history of the book to history of reading: theories and methods for historical studies of reading*. Xinxing.
[20] Pat Deely. 1975. Copyright: Limitation on Exclusive Rights, Fair Use. *Hous. L. Rev.* 13 (1975), 1041.
[21] Stefan Dimitrov, Faiyaz Zamal, Andrew Piper, and Derek Ruths. 2015. Goodreads versus Amazon: the effect of decoupling book reviewing and book selling. In *Ninth international AAAI conference on web and social media*.
[22] Douban. 2021. About Douban. https://www.douban.com/about
[23] Douban. 2022. Lolita (webpage for the book). https://book.douban.com/subject/1465324/
[24] Beth Driscoll. 2021. How goodreads is changing book culture. *Kill Your Darlings* (2021), 213–216. https://search.informit.org/doi/10.3316/INFORMIT.112742441519686
[25] James F English. 2021. A Future for Empirical Reader Studies. https://culturalanalytics.org/post/1208-a-future-for-empirical-reader-studies
[26] RA Gekoski. 2004. *Tolkien's Gown: And Other Stories of Great Authors and Rare Books*. Constable.
[27] Goodreads. 2021. About Goodreads. https://www.goodreads.com/about/us
[28] Goodreads. 2022. Lolita (webpage for the book). https://www.goodreads.com/book/show/7604.Lolita
[29] Zhu guang shi (on Zuoshu2013). 2020. Big Changes to the Douban Books Top 250 List, The Kite Runner Is No Longer Ranked No. 1. https://post.smzdm.com/p/a830r7gq/
[30] Nan Hu, Indranil Bose, Noi Sian Koh, and Ling Liu. 2012. Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision support systems* 52, 3 (2012), 674–684.
[31] Jacob Jett, Boris Capitanu, Deren Kudeki, Timothy Cole, Yuerong Hu, Peter Organisciak, Ted Underwood, Eleanor Dickson Koehl, Ryan Dubnicek, and J. Stephen Downie. 2020. The HathiTrust Research Center Extracted Features Dataset (2.0). https://doi.org/10.13012/R2TE-C227
[32] Ming Jiang and Jana Diesner. 2016. Issue-focused documentaries versus other films: Rating and type prediction based on user-authored reviews. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*. 225–230.
[33] Ming Jiang and Jana Diesner. 2016. Says who…? Identification of expert versus layman critics' reviews of documentary films. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2122–2132.
[34] Tomoya Kambara, Shohei Okamoto, Yuka Teramoto, Kazuma Kusu, and Kenji Hatano. 2018. Evaluating usefulness of reviews based on evaluation standpoints of consumers. In *Proceedings of the 10th International Conference on Management of Digital EcoSystems*. 110–117.
[35] Ho Kiu-Chor et al. 2007. A Case Study of Douban: Social Network Communities. *Masaryk University Journal of Law and Technology* 1, 2 (2007), 43–56.
[36] Marijn Koolena, Peter Bootb, and Joris J van Zundertb. 2020. Online Book Reviews and the Computational Modelling of Reading Impact. *Proceedings http://ceur-ws.org ISSN* 1613 (2020), 0073.
[37] Joshua Kotin, Rebecca Sutton Koeser, Carl Adair, Serena Alagappan, Paige Allen, Jean Bauer, Oliver J Browne, Nick Budak, Harriet Calver, Jin Yun Chow, et al. 2021. Shakespeare and Company Project Dataset: Lending Library Events. *https://doi.org/10.34770/39sq-bm51* (2021).
[38] Balázs Kovács and Amanda J Sharkey. 2014. The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative science quarterly* 59, 1 (2014), 1–33.
[39] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.
[40] Theodoros Lappas. 2012. Fake reviews: The malicious perspective. In *International Conference on Application of Natural Language to Information Systems*. Springer, 23–34.
[41] Theodoros Lappas, Gaurav Sabnis, and Georgios Valkanas. 2016. The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research* 27, 4 (2016), 940–961.
[42] Hongliu Li, Xingyuan Wang, Shuyang Wang, Wenkai Zhou, and Zhilin Yang. 2022. The power of numbers: an examination of the relationship between numerical cues in online review comments and perceived review helpfulness. *Journal of Research in Interactive Marketing* (2022).
[43] LibraryThing. 2021. About LibraryThing. https://www.librarything.com/about
[44] Zhiwei Liu and Sangwon Park. 2015. What makes a useful online review? Implication for travel product websites. *Tourism management* 47 (2015), 140–151.
[45] Hoyt Long. 2021. Culture at Global Scale. https://culturalanalytics.org/post/1160-culture-at-global-scale
[46] Ana Isabel Lopes, Nathalie Dens, Patrick De Pelsmacker, and Freya De Keyzer. 2020. Which cues influence the perceived usefulness and credibility of an online review? A conjoint analysis. *Online Information Review* (2020).
[47] Caimei Lu, Jung-ran Park, and Xiaohua Hu. 2010. User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of information science* 36, 6 (2010), 763–779.
[48] Michael Luca and Georgios Zervas. 2016. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science* 62, 12 (2016), 3412–3427.
[49] Suman Kalyan Maity, Abhishek Panigrahi, and Animesh Mukherjee. 2017. Book reading behavior on goodreads can predict the amazon best sellers. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. 451–454.
[50] Matthias Mauch, Robert M MacCallum, Mark Levy, and Armand M Leroi. 2015. The evolution of popular music: USA 1960–2010. *Royal Society open science* 2, 5 (2015), 150081.
[51] Megan McCluskey. 2021. How Extortion Scams and Review Bombing Trolls Turned Goodreads Into Many Authors' Worst Nightmare. https://time.com/6078993/goodreads-review-bombing/
[52] Ian Milligan. 2016. The problem of history in the age of abundance. (2016).
[53] Simone Murray. 2021. Secret agents: Algorithmic culture, Goodreads and datafication of the contemporary book world. *European Journal of Cultural Studies* 24, 4 (2021), 970–989.
[54] Lisa Nakamura. 2013. "Words with friends": socially networked reading on Goodreads. *PMLA/Publications of the Modern Language Association of America* 128, 1 (2013), 238–243.
[55] Edward Daniel Newell, Stefan Dimitrov, Andrew Piper, and Derek Ruths. 2016. To buy or to read: How a platform shapes reviewing behavior. In *Tenth International AAAI Conference on Web and Social Media*.
[56] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2 (2019), 13.
[57] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[58] Lambodara Ptuabhof and Phool Rani Da. 2018. Goodreads Ratings and Reviews Analysis of Booker Prize Titles. In *ICDT 2018:Publishing Technology and Future of Academia*. Segment Publication, 363–371.
[59] Monisha Rajesh. 2021. Pointing out racism in books is not an 'attack' – it's a call for industry reform. https://www.theguardian.com/books/2021/aug/13/pointing-out-racism-in-books-is-not-an-attack-kate-clanchy
[60] Juan Ramos et al. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242. Citeseer, 29–48.
[61] Rezvaneh Rezapour and Jana Diesner. 2017. Classification and detection of micro-level impact of issue-focused documentary films based on reviews. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1419–1431.
[62] Rui. 2013. Douban Top 250 Books Old Version 2013.06. https://www.douban.com/note/536479320/
[63] Nazanin Sabri and Ingmar Weber. 2021. Users Data. (2021).
[64] Ruchira Sharma. 2021. Black and LGBTQ+ authors say they're being harassed on Goodreads and trolled with one-star book reviews. https://inews.co.uk/culture/books/goodreads-book-reviews-black-lgbtq-authors-harrassed-trolled-949179
[65] Shuyang. 2016. douban.com top 250 movies and books. https://github.com/Shuyang/douban_top250/tree/master

[66] Eastern Express (Taiyuan). 2011. Douban Top 250 Books. https://www.douban.com/doulist/513669/

[67] Ted Underwood. 2019. *Distant horizons: digital evidence and literary change*. University of Chicago Press.

[68] Melanie Walsh and Maria Antoniak. 2021. The Goodreads 'Classics': A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism. *Journal of Cultural Analytics* 4 (2021), 243–287.

[69] Mengting Wan and Julian J. McAuley. 2018. Item recommendation on monotonic behavior chains. In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan (Eds.). ACM, 86–94. https://doi.org/10.1145/3240323.3240369

[70] Mengting Wan, Rishabh Misra, Ndapa Nakashole, and Julian J. McAuley. 2019. Fine-Grained Spoiler Detection from Large-Scale Review Corpora. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 2605–2610. https://doi.org/10.18653/v1/p19-1248

[71] Jing Wang, Anindya Ghose, and Panos Ipeirotis. 2012. Bonus, disclosure, and choice: what motivates the creation of high-quality paid reviews?. In *ICIS 2012 Proceedings*. Citeseer.

[72] Lotte M Willemsen, Peter C Neijens, Fred Bronner, and Jan A De Ridder. 2011. "Highly recommended!" The content characteristics and perceived usefulness of online consumer reviews. *Journal of Computer-Mediated Communication* 17, 1 (2011), 19–38.

[73] Adam Worrall. 2015. " Like a Real Friendship": Translation, Coherence, and Convergence of Information Values in LibraryThing and Goodreads. *iConference 2015 Proceedings* (2015).

[74] Yuanyuan Wu, Eric WT Ngai, Pengkun Wu, and Chong Wu. 2020. Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems* 132 (2020), 113280.

[75] Ruolin Xie. 2021. Investigaton into Douban water army: 15 RMB for a short review, votes and thumb-ups available as well[“”15 . http://www.xinhuanet.com/fortune/2021-02/25/c_1127136296.htm

[76] Gregory Yauney. 2021. shakespeare-and-company-online-readership. https://github.com/gyauney/shakespeare-and-company-social-readership

[77] Zebulon2020. 2020. Douban Read Top250 Crawler. https://github.com/zebulon2020/DoubanReadTop250Crawler

[78] Jialian Zhou. 2018. douban.com top 250 movies and books. https://doi.org/10.18170/DVN/X20PS1