


A Global Book Reading Dataset

Nazanin Sabri ¹ and Ingmar Weber ^{2,*}

¹ School of Electrical and Computer Engineering, University of Tehran, Tehran 1439957131, Iran; nazanin.sabri@ut.ac.ir

² Qatar Computing Research Institute, Doha P.O. Box 34110, Qatar

* Correspondence: iweber@hbku.edu.qa

Abstract: The choice of what to read is both influenced by and indicative of such factors as a person's beliefs, culture, gender, and socioeconomic status. However, obtaining data including such personal attributes, as well as detailed reading habits and activities of individuals is difficult and would usually require either (i) data from e-readers, such as the Amazon Kindle, or from library checkouts, both of which are hard to obtain, or (ii) distributing questionnaires and conducting interviews, which can be expensive and suffers from recall bias. In this study, we present a dataset of over 40 million reading instances of 1,872,677 unique individuals collected from Goodreads. Goodreads is a book-cataloging social media platform with millions of users, where users share comments on the books they have read, while creating and maintaining social connections. We enrich the dataset with gender and location information. The dataset presented in this study can be used to perform cross-national and cross-gender analyses of reading behavior among book enthusiasts.

Keywords: reading; dataset; Goodreads



Citation: Sabri, N.; Weber, I. A Global Book Reading Dataset. *Data* **2021**, *6*, 83. <https://doi.org/10.3390/data6080083>

Academic Editor: Reinhold Decker

Received: 27 June 2021

Accepted: 31 July 2021

Published: 4 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Reading is a globally popular pastime that has been shown to be a beneficial non-medical strategy for improving mental health and well-being [1]. Reports from 2017 [2] and 2018 [3] have shown India to be the most active country with regards to reading in the world, with an average of more than ten hours a week spent on reading. According to both reports, 70% of Americans indicated having read at least one book during the past year, and the median number of books read in the U.S. per person per year is four, with an average of 12 [4]. The Pew Research Center explores the demographic traits that characterize the approximate quarter of the American population that does not read books in a given year, a percentage that has grown compared to a decade ago [5].

As reading is usually an activity performed alone and in the comfort of one's home, information on the reading habits and behaviors of nations and individuals is rarely recorded. While the growth of e-readers could result in the documentation of this data, the information would not be openly accessible to members of the research community. Additionally, print books are still more popular than digital books [6,7], meaning that such data would only account for a small proportion of reading instances.

Given the known, positive effects of reading on mental and psychological well being, there is continued interest in understanding factors that influence reading habits. Amid the COVID-19 pandemic, for instance, a web survey of the reading habits of Spanish and Italian readers was conducted [8], collecting a dataset of these habits during confinement. While the data can tell us a lot about these habits, the authors acknowledge the low response rates, showing that a large proportion of those who opened the questionnaire abandoned it before answering all the questions. Moreover, the data are limited to two countries, preventing a large scale cross-country comparison.

In this paper, we present data from Goodreads (<http://goodreads.com>, accessed on 20 July 2021) with the goal of enabling large-scale studies of reading behaviors. Goodreads

is a book-centered social media platform, launched in 2007. Based on their own claims (<https://www.goodreads.com/advertisers>, accessed on 20 July 2021), they have “45 million unique visitors a month”. Other statistics estimate that, as of 2019, the site had 90 million registered users [9]. Based on our explorations, the website has acquired well over 120 million users over its 13 years of activity, though many of these might no longer be active. We present a dataset of 41,253,535 book “reviews” (while Goodreads uses the terminology “review”, these reviews do not require a numerical rating or textual evaluation, and so the term “posting action related to a book” might be more appropriate), left by 1,872,677 Goodreads users with public profiles. Upon collection, the data are enriched with country information and inferred gender. The collection process and the details of the dataset are reported in the Section 2.

Over the past couple of years, data from the Goodreads platform have been used for several academic studies. Thelwall and Kousha [10] explore the user base of the website, comparing the behavior and activity on the platform with respect to gender by, for instance, showing that females register more books and rate them less positively. As the study is conducted through the analysis of 50,000 random users, the dataset we share as part of this study could also be used to answer similar questions. More broadly speaking, we see this dataset as useful for supporting user-centric studies on reading behavior. Other researchers have focused solely on the reviews and ratings left on Goodreads. For example, [11] studies the sentiment, emotion and language expressed in reviews. Others have looked at what aspects of a book, e.g., characters or storyline, are being discussed [12]. Kousha et al. [13] investigate the feasibility of using Goodreads’ book metrics and reviews as a means to assess the impact of books. Alghamdi and Ihshaish [14] address related questions of potential influence, looking only at Arabic book reviews. Maity et al. [15] study how much user behavior on Goodreads could be indicative of sales on other book retail platforms, such as Amazon. As we have chosen not to share the review text, or the book identities due to potential risk of abuse, our dataset does not directly support these review-centric studies.

Rather, we hope that these data will allow studies of reading habits and behaviors, and how these habits are impacted by various events and social movements. In particular, we believe that certain cross-cultural, cross-gender, and cross-country studies of reading are enabled through the use of these data.

2. Data Collection and Exploration

To collect the data in this study, we used the official Goodreads API [16], using a Python program to connect to and collect data from the API. As of 8 December 2020, the website has declared that it no longer provides new API keys (<https://help.goodreads.com/s/article/Does-Goodreads-support-the-use-of-APIs>, accessed on 20 July 2021), and it has since started to retire previously issued API keys (<https://help.goodreads.com/s/article/Why-did-my-API-key-stop-working>, accessed on 20 July 2021).

For our data collection, we chose a user-centric approach, i.e., collecting a “complete” snapshot of data for a sample of users, rather than, for example, collecting all readers of a sample of books. To select the sample of users, we proceeded as follows. First, we observed that the internal Goodreads user ID seems to be consecutively assigned, with user ID 1 belonging to the Goodreads Founder Otis Chandler. The largest user ID, as of September 2020, was 121,761,242. We then proceeded to sample from the user ID space as follows.

We initially began by querying the user space by selecting a few user IDs at random and then continuing the collection by adding a constant number to these values and collecting those accounts. The precise details of this changed during the collection process as we gradually refined our data collection objectives. For example, initially, the clustering of many near-adjacent IDs, corresponding to users who registered around the same time, was not a concern. In fact, we were interested in investigating accounts that had joined during the COVID-19 pandemic, and a more rigorous collection of IDs (with additions of smaller numbers) was conducted for IDs in that space (causing the peak that is visible in

Figure 1). However, later, we changed to uniformly sampling the entire ID space to avoid oversampling users who registered on particular dates.

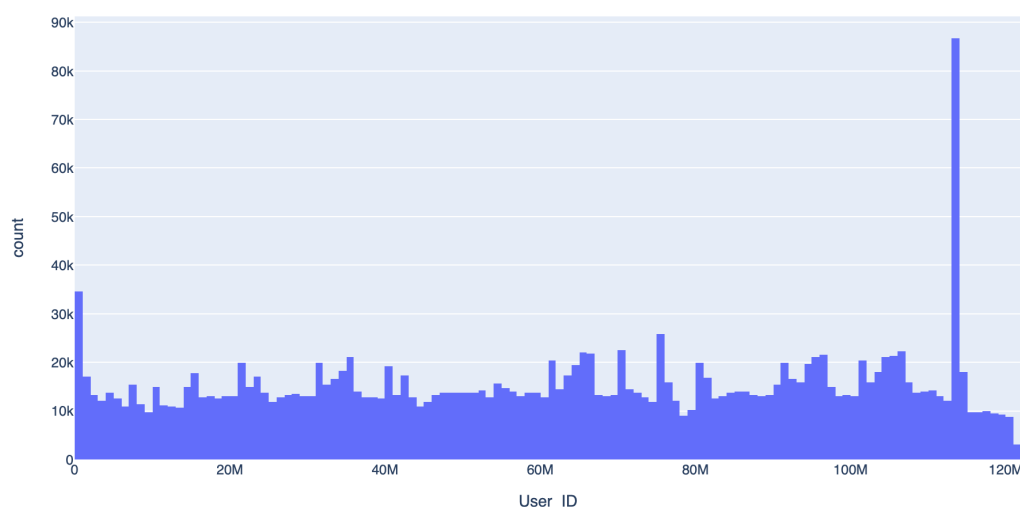


Figure 1. Distribution of user-id numbers.

To obtain information about a given user through the API, we used the `user.show` method (<https://www.goodreads.com/api/index#user.show>, accessed on 20 July 2021). Note that only information for users who set their profile to be viewable by “anyone (including search engines)” was used, and the API does not support the collection of data for private accounts. Next, after making sure that the account was public, we collected all the books that the user had added through the `reviews.list` method (<https://www.goodreads.com/api/index#reviews.list>, accessed on 20 July 2021), paginating through long result lists where necessary. While we follow the Goodreads terminology and use the term “review”, these reviews are not required to have any textual review, nor any type of rating. Instead they can merely indicate that a user posted a book to one of their shelves, for example, the “read” shelf.

The data were collected in 2020, encompassing any books that users had read since first joining the website until August 2020. Tables 1 and 2 display the fields available for each individual and review respectively (this dataset is publicly available at https://figshare.com/projects/A_Global_Book_Reading_Dataset/118854 (accessed on 20 July 2021)).

Table 1. Explanation of fields available for each user.

Field	Description	Included in Public Dataset
User ID	A unique, numerical identifier for the user on the website.	A hashed version of the ID is made available.
Name	Name of this user. In contrast to many other pseudonymous social networks, Goodreads users tend to use real names and even full names, as the input form has separate first, middle, and last name fields.	No
Username	The username that the user has selected. This field is optional; name is the field each user must fill in to create an account.	No
Profile Image	URL of the user’s profile picture.	No
Friend Count	The number of friends that the user has. Being friends on Goodreads is a bidirectional property, independent of uni-directional following; 62% of users do not have any friends.	Yes
Review Count	The total number of books added to any of the user’s shelves, in other words, the total number of books in the user’s automatically generated “all” shelf. Only 4.5% of users have more than 100 books in their shelves.	Yes

Table 1. Cont.

Field	Description	Included in Public Dataset
Groups Count	The number of groups the user is part of. Some groups can be freely joined, for others the user needs to be admitted.	Yes
Location	An optional self-reported location of the user. By default, Goodreads seems to infer a user's country, presumably based on IP address. This selection can be changed later and a drop-down list of countries is available. Only 3.7% of users have left the field empty.	Self-reported locations are not reported but detected country-level values are.
Age	Self-reported age of the user; 97% of users have not completed this value.	No
Gender	Self-reported gender of the user. Only 7735 have filled in this value. Options, from a drop-down, include male, female and custom, which supports free-text.	Inferred gender values are included, but not the self-reported ones.
About	An optional self-description of the user.	The numerical length of this section is included, but not the textual content.
Favorite Authors	Favorite authors of the user.	Yes, but author IDs are replaced by hashed values.
Website	An optional field, allowing users to share their website or any other link.	No
Joined	The month and year in which the user joined the platform.	Yes
Last Active	The month and year that the user was last active on this website (since our collection was conducted in 2020, dates within this year do not necessarily indicate that the user has abandoned the website).	Yes

Table 2. Explanation of fields available for each review.

Field	Description	Included in Public Dataset
Book ID	A unique identifier for the book this review is about.	A hashed version of the ID is provided.
Rating	A numerical rating, taking integer values from 1 to 5. Ratings are optional and can be left empty. Only 47.8% of book additions include ratings.	Yes
Shelve Names	Users are able to make different shelves. While there are no restrictions on the shelves you are allowed to create, it is at times viewed as genres or tags used for recommendations.	Yes, but shelf names used by fewer than 200 distinct users are replaced by <i>small-count</i> to prevent the tracking of users with a certain taste.
Spoiler Flag	A Boolean flag indicating if the review contains spoilers (the flag is set by the user)	Yes
Review Body	The text of the review.	No, but the character length of the text is provided.
Likes	Number of likes for the review.	Yes
Date Added	A systematically generated date of when the user first added this book to one of their shelves.	Yes
Date Updated	A systematically generated date of the last time the user updated this book.	Yes
Started At	An optional user-inputted date indicating when the user started reading the book.	Yes
Read At	An optional user-inputted date indicating when the user finished reading the book.	Yes
Owned	Whether the user owns the book.	Yes
Read Count	Number of times this book was read by this user (re-reads are possible on the platform).	Yes

The data obtained through the Goodreads API were then enriched, in particular by attempting to infer a user's gender. To infer gender from a user's self-declared name and/or username, we used the Name2GAN tool [17] to detect the most probable female or male gender of the name. Unlike name dictionaries from the U.S.A. Social Security Administration (<https://www.ssa.gov/oact/babynames/limits.html>, accessed on 20 July 2021), Name2GAN is trained on multi-lingual Wikipedia and social media data and recognizes names from many cultures. While the tool only supports a binary male-or-female classification, as well as an "unknown" option for unrecognized names, we are not implying that gender is binary, and we acknowledge that many people self-identify as non-binary. Goodreads also supports free-text, non-binary gender in the user profile. However, for the small set of users where a self-reported gender was available, including self-reported non-binary gender, we decided not to include this information in the shared dataset so as not to provide an easy way to identify users based on their gender identity. We believe that the inferred binary gender still provides a meaningful signal for studying gender differences between women and men in reading behavior, without exposing vulnerable minorities to the risk of identification.

Using this approach, we inferred a gender of either male or female for 87% (1,634,103 out of 1,872,677) of users. To estimate the accuracy of the gender inference, we compared the detected gender against the self-declared gender for the set of users who added their genders manually. We found that, among those with self-reported binary gender values, 86.4% of the instances are labeled correctly. Upon inspection of the not-correctly-classified values by hand, we found that these names are often either abbreviated versions of the person's name (e.g., E. M.), truly ambiguous names (e.g., Mallia Chris), or not people's names at all (e.g., DR, International School). The distribution of gender values is displayed in Figure 2. We can observe that there are disproportionately more female users in our dataset than male users. This is, however, on-par with other statistics on the users of the website showing that the user base of the website is predominantly female [18].

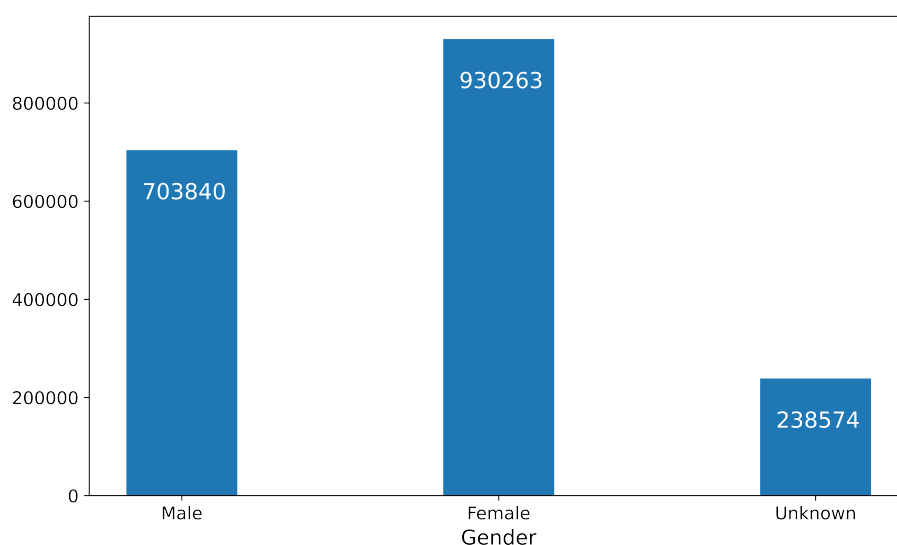


Figure 2. Distribution of detected gender values.

Next, we analyzed the location values of users, aiming to detect country of origin based on the unstructured texts users have shared on the platform. By default, Goodreads appears to automatically infer a user's country, most likely based on the user's IP address. (This is based on the authors' own observation when creating a test account.) After sign-up, users can then choose to edit this location information, which includes selecting a country from a drop-down menu, including an "-" (empty) option. They can also provide free-text city and state information. Given the enforced country-level scheme, almost all users have a clearly identifiable country. To extract country information, both in the majority of easy

cases, as well as in a smaller number of harder cases, we used a combination of rule-based approach, followed by the use of GeoPy (<https://github.com/geopy/geopy>, accessed on 20 July 2021). GeoPy is a Python client for several popular geocoding web services. More specifically, the system makes use of the *Google Maps Platform*, *OpenStreetMap Nominatim*, and *Bing Maps*, among others, to work. We performed the following steps, one after the other, stopping if we found a country:

- Comma separate the string, checking only the last part of the string against a list of countries and state names, labeling the country if the value is on that list. This is because most people use the convention of mentioning their country as the last part of their address. A total of 96% of locations are detected in this manner.
- Comma separate the string, checking only the first part of the string against a list of country names, labeling the country if the value is on that list. Similar to the intuition of the last part, this time, consider those who start their address by writing their country name. A total of 0.07% of locations are detected in this manner.
- Input the entire string to GeoPy. A total of 0.06% of locations are detected in this manner.

Eventually, a total of 96.2% of user locations were detected. As 3.7% of users had an empty location field, this means that only a tiny fraction of users did not have a usable location that could be mapped to a country. Figure 3 shows the distribution of these locations across the world. We can see that U.S.-based (711,889 users, making up 38% of the users in our dataset) and India-based (163,521 users, making up 8.7% of the users in our dataset) users make up a large proportion of our dataset.

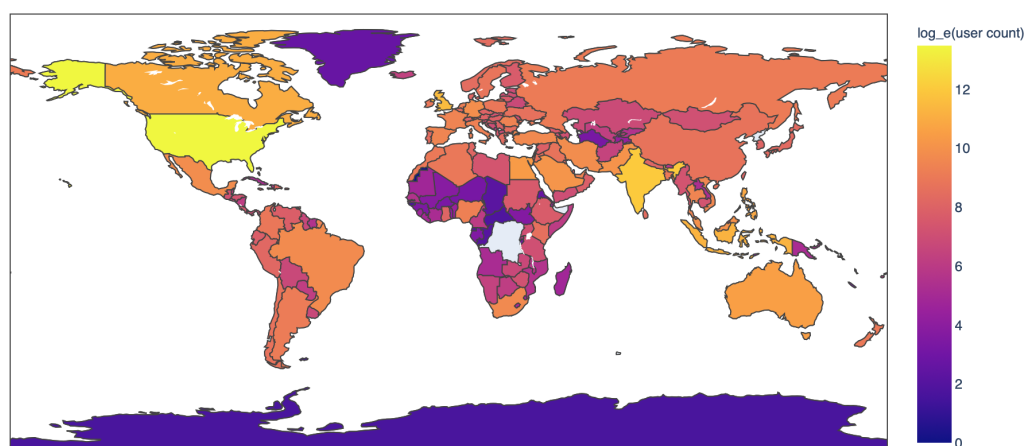


Figure 3. Distribution of detected location values. The values displayed in the figure are natural logarithms (base e).

As shown in Figure 4, most users who join Goodreads are not active after the first month they join. However, there are users who have been active on the website for 13 years.

Table 3 displays the statistics of our dataset. Among the book additions in our dataset, *Harry Potter and the Sorcerer's Stone* by J.K. Rowling is the most added book in our dataset, followed by *The Hunger Games* by Suzanne Collins and *To Kill a Mockingbird* by Harper Lee. (While we have decided not to include any book titles in the public dataset, we include these three titles here to provide a sense of what is popular on Goodreads).

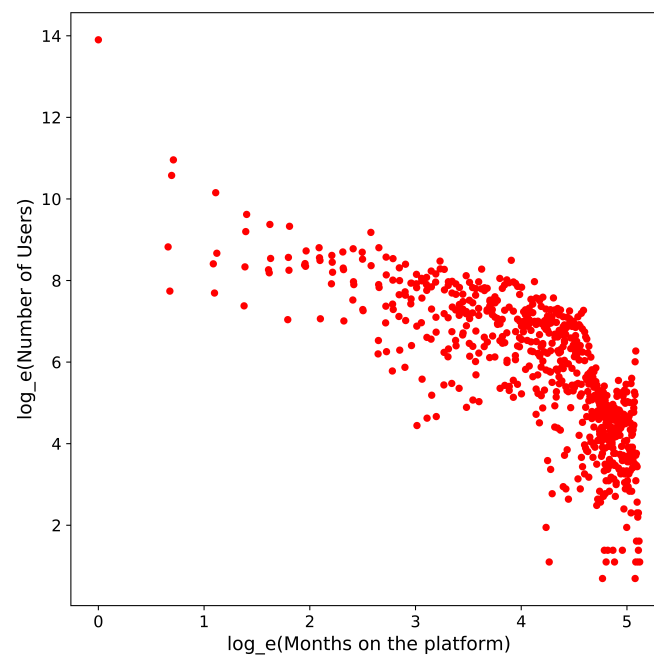


Figure 4. Distribution of the number of months the user was active on the website. The values displayed on both axes are on the natural logarithm scale.

Table 3. Statistics of our dataset.

Instance	Count
Users	1,872,677
Books	3,594,304
Book Additions (Referred to as Reviews)	41,253,535
Book Additions (Reviews) with Rating	19,852,290

The most used tags for female and male users of the platform are shown in Table 4.

Table 4. Top 10 tags used by men and women.

Female	Male
read	read
to-read	to-read
currently-reading	currently-reading
favorites	fiction
fiction	fantasy
fantasy	favorites
romance	owned
own	history
non-fiction	own
young-adult	science-fiction

Anonymity

As previously mentioned, all data were collected through the Goodreads's official API, only including information about public accounts. Despite the public nature of the

data, we believe that data about individuals must be published only in an anonymized form to minimize the risk of harm to users whose information is included in the data release. Correspondingly, the released data do not include a user's name, their username, their precise location, or, in fact, any text input at all, as any free text field might leak personally identifiable information. The user ID in the data release is a hash of the original ID, where the hash function includes a random "salt" to guard against lookup attacks. A particular type of risk that we have tried to mitigate relates to identifying incidents of users reading "forbidden books". This in particular relates to books that are banned for their political, religious, or sexual content. Even though no personally identifiable information was included in the data release, we have chosen not to include the identity of any book or author in the released dataset to minimize the particular risk of identifying users reading "forbidden books". For the same reason, we have decided to remove information about shelf names used by fewer than 200 distinct users, which included such shelf names as "LGBTQ" or "Erotica" (shelf names used by more than 200 distinct users, as well as the number of unique users who have used them, are shown in Table 5). However, we acknowledge the risk that globally popular books might be reidentifiable through their popularity level and their global distribution pattern.

Table 5. The most used shelves by the users in the dataset.

Shelf	Number of Users	Shelf	Number of Users	Shelf	Number of Users
to-read	1,087,410	manga	550	cookbooks	313
read	758,974	humor	546	didn-t-finish	304
currently-reading	309,831	my-books	542	school	301
favorites	5959	2015	531	dystopia	301
fantasy	2420	favourites	525	childrens	300
fiction	2037	psychology	523	plays	299
nan	1968	business	521	library	286
non-fiction	1831	books	511	economics	272
classics	1561	memoir	508	chick-lit	267
history	1322	comics	506	want-to-read	266
poetry	1309	owned	486	suspense	265
romance	1214	self-help	449	children	263
mystery	1197	wishlist	429	sports	262
historical-fiction	1184	to-buy	422	series	261
2020	1101	2014	418	drama	258
dnf	1078	travel	407	novels	251
2019	1035	contemporary	406	urban-fantasy	247
young-adult	959	audiobook	404	parenting	245
biography	898	politics	402	mythology	240
2018	879	historical	400	2012	239
sci-fi	869	dystopian	397	kids	236
science-fiction	848	crime	394	literature	233
horror	846	religion	392	couldn-t-finish	232
abandoned	823	re-read	382	favorite	225
nonfiction	746	kindle	378	maybe	224
philosophy	740	graphic-novel	373	feminism	222

Table 5. Cont.

Shelf	Number of Users	Shelf	Number of Users	Shelf	Number of Users
2017	730	on-hold	371	education	219
science	711	books-i-own	360	ebook	219
1	683	paranormal	359	writing	218
did-not-finish	667	adventure	355	vampires	217
book-club	661	audiobooks	355	food	215
thriller	658	unfinished	350	first-reads	214
2016	615	art	344	comedy	214
own	590	music	332	picture-books	212
short-stories	590	classic	323	children-s-books	209
graphic-novels	562	reference	317	health	204
ya	553	2013	314	true-crime	201

Furthermore, we accept the reidentification possibility with another Goodreads data collection. In other words, if an attacker was to recollect a dataset similar to the one that we are sharing, then they would likely be able to link users on things such as their activity patterns. However, in that scenario, the attacker would not gain any additional benefit from having access to our particular data.

3. Potential Use Cases

In its current form, we see the biggest values in user-centric studies that make use of the international nature, as well as of the inferred gender. While fine-grained book information was withheld, knowing when, where, and what type of popular genre (i.e. shelf name) is being read and posted about could still serve as the basis of important studies on the interplay between country, gender, book genre, and activity patterns over time. For example, this dataset enables knowledge of how temporal patterns affect the reading behavior of book enthusiasts. In fact, the initial motivation for the creation of this dataset was to observe the gender-specific impact of the reading of women during the COVID-19 pandemic. Surprisingly, we did not find a clear pattern here, possibly due to the dominance of book enthusiasts, who might continue reading, even when faced with calamity. Future work could examine the differences between the most popular book genres (see, for example, Table 4), analyzing temporal changes to what each gender uses most, and investigating if and how they are affected by real-world events.

We acknowledge that many interesting use cases would require knowing additional information, such as a book's identity, or the text of a review, both of which were withheld from this data release as explained in the previous section. For well-specified use cases with a mission of social good, and where external, ethical review can be demonstrated, we invite researchers to contact the authors to discuss additional data access options.

Data Limitations

The collected dataset has certain limitations, including the following. Firstly, at the beginning of the collection, the API was queried, using constantly-spaced user identifiers rather than randomly sampled numbers. This method of collection was then later changed to sampling IDs uniformly from the ID space. However, due to the initial approach, some ID ranges, and hence, some sign-up periods, were over sampled, compared to others (please see Figure 1).

A technical limitation is that the API does not allow us to capture the dates of re-reads of the same book. In other words, if a user reads a book more than once, that "review" instance is updated to now reflect the new dates and status of the review. No new instance regarding that book is created; instead, each user only has one review instance for each book,

which is updated whenever a change to the status of the book is made. Consequently, while we can detect the number of times they have read the book (using the “read-count” field in the dataset), we are not able to find when each round of reading took place, and only have access to the dates for the last time the book was marked as read. Information regarding the date at which the book was re-read is available on each review page on the website. However, to the best of our knowledge, this information cannot be collected through the API.

Finally, it is important to remember that the data represent the reading habits of avid readers, as joining a social network for books is not something done by the majority of readers. Any findings derived from these data will correspondingly need to be interpreted with the underlying user selection bias in mind.

Author Contributions: Conceptualization, N.S. and I.W.; methodology, N.S. and I.W.; software, N.S.; formal analysis, N.S. and I.W.; data curation, N.S.; writing—original draft preparation, N.S.; writing—review and editing, N.S. and I.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset is publicly available at https://figshare.com/projects/A_Global_Book_Reading_Dataset/118854 (accessed on 20 July 2021).

Acknowledgments: We thank the anonymous reviewers for their constructive feedback on the submission.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Billington, J.; Dowrick, C.; Hamer, A.; Robinson, J.; Williams, C. An Investigation into the Therapeutic Benefits of Reading in Relation to Depression and Well-Being. Liverpool: The Reader Organization, Liverpool Health Inequalities Research Centre. 2010. Available online: https://www.academia.edu/download/32364850/An_investigation_into_the_therapeutic_benefits_of_reading_in_relation_to_depression_and_well-being.pdf (accessed on 20 July 2021)
2. Brown, B. The Ultimate Guide to Global Reading Habits (Infographic). 2017. Available online: <https://geediting.com/world-reading-habits/> (accessed on 19 December 2020).
3. Brown, B. World Reading Habits in 2018 (Infographic). 2018. Available online: <https://geediting.com/world-reading-habits-2018/> (accessed on 19 December 2020).
4. Perrin, A. Book Reading 2016. 2016. Available online: <https://www.pewresearch.org/internet/2016/09/01/book-reading-2016/> (accessed on 20 December 2020).
5. Perrin, A. Who Doesn't Read Books in America? 2019. Available online: <https://www.pewresearch.org/fact-tank/2019/09/26/who-doesnt-read-books-in-america/> (accessed on 20 December 2020).
6. Perrin, A. One-in-Five Americans Now Listen to Audiobooks. 2019. Available online: <https://www.pewresearch.org/fact-tank/2019/09/25/one-in-five-americans-now-listen-to-audiobooks/> (accessed on 20 December 2020).
7. CNBC. Physical Books Still Outsell e-Books—And Here's Why. 2019. Available online: <https://www.cnbc.com/2019/09/19/physical-books-still-outsell-e-books-and-heres-why.html> (accessed on 20 December 2020).
8. Salmerón, L.; Arfé, B.; Avila, V.; Cerdán, R.; De Sixte, R.; Delgado, P.; Fajardo, I.; Ferrer, A.; García, M.; Gil, L.; et al. READ-COGvid: A Database From Reading and Media Habits During COVID-19 Confinement in Spain and Italy. *Front. Psychol.* **2020**, *11*, 2639. [CrossRef] [PubMed]
9. Clement, J. Goodreads: Number of Registered Members 2011–2019. 2020. Available online: <https://www.statista.com/statistics/252986/number-of-registered-members-on-goodreadscom/> (accessed on 19 December 2020).
10. Thelwall, M.; Kousha, K. Goodreads: A social network site for book readers. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 972–983. [CrossRef]
11. Driscoll, B.; Rehberg Sedo, D. Faraway, so close: Seeing the intimacy in Goodreads reviews. *Qual. Inq.* **2019**, *25*, 248–259. [CrossRef]
12. Hajibayova, L. Investigation of Goodreads' reviews: Kakutanied, deceived or simply honest? *J. Doc.* **2019**, *75*, 612–626. [CrossRef]
13. Kousha, K.; Thelwall, M.; Abdoli, M. Goodreads reviews to assess the wider impacts of books. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 2004–2016. [CrossRef]
14. Alghamdi, A.; Ihsaish, H. The use and impact of Goodreads rating and reviews, for readers of Arabic Books. *Int. J. Bus. Inf. Syst.* **2020**. Available online: <https://uwe-repository.worktribe.com/OutputFile/4448322> (accessed on 20 July 2021).
15. Maity, S.K.; Panigrahi, A.; Mukherjee, A. Analyzing Social Book Reading Behavior on Goodreads and How It Predicts Amazon Best Sellers. In *Influence and Behavior Analysis in Social Networks and Social Media. ASONAM 2018. Lecture Notes in Social Networks*; Springer: Cham, Switzerland, 2018. Available online: [doi:10.1007/978-3-030-02592-2_11](https://doi.org/10.1007/978-3-030-02592-2_11) (accessed on 20 July 2021) [CrossRef]

-
16. Goodreads. Goodreads API. 2020. Available online: <https://www.goodreads.com/api> (accessed on 19 December 2020).
 17. Jung, S.; Salminen, J.; Jansen, B.J. Name2GAN (Version 1.1) [Computer Software]. Qatar Computing Research Institute. 2020. Available online: <https://quecst.qcri.org/tool/Name2GAN> (accessed on 20 July 2021).
 18. Johnson, J. Distribution of the Online Audience of Goodreads.com in Great Britain (GB) in 2018, by Age Group and Gender. 2020. Available online: <https://www.statista.com/statistics/490362/gb-online-audience-of-goodreads-com-by-age-group-and-gender/> (accessed on 21 December 2020).