



Universiteit  
Leiden  
The Netherlands

# The Datafication of the Contemporary Book Culture and the Two-Way Relationship Between Bookish Platforms and Users: A Goodreads Analysis

Lemonidou, Kyriaki

## Citation

Lemonidou, K. (2025). *The Datafication of the Contemporary Book Culture and the Two-Way Relationship Between Bookish Platforms and Users: A Goodreads Analysis*.

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master Thesis, 2023](#)

Downloaded from: <https://hdl.handle.net/1887/4180706>

**Note:** To cite this publication please use the final published version (if applicable).

The Datafication of the Contemporary Book Culture and the Two-Way Relationship  
Between Bookish Platforms and Users: A Goodreads Analysis

Kyriaki Lemonidou (s4027914)

MA Thesis

Supervisor: Dr. Peter Verhaar

Second Reader: Dr. Fleur Praal

9/8/2024

17.594 words

<b>Abstract .....</b>	<b>3</b>
<b>Acknowledgments.....</b>	<b>3</b>
<b>Introduction .....</b>	<b>3</b>
<b>Methodology .....</b>	<b>9</b>
<b>Limitations .....</b>	<b>10</b>
<i>Chapter I .....</i>	<i>10</i>
<b>The Datafication of the Reading Experience on Goodreads.....</b>	<b>10</b>
<b>A Close Analysis of the Goodreads Platform Functionalities.....</b>	<b>12</b>
<b>The Possible Implications of these Features for the Digital Reader Experience .....</b>	<b>17</b>
<i>Chapter II.....</i>	<i>20</i>
<b>Goodreads as an Undertaking of Amazon: Possible Consequences for the Reading Culture.....</b>	<b>20</b>
<b>Data Analysis Between Goodreads and Amazon Book Reviews: Identifying Platform-specific Behaviors .....</b>	<b>24</b>
<b>Data .....</b>	<b>25</b>
<b>Results .....</b>	<b>26</b>
i. Average Review Length.....	26
ii. Average Star Rating.....	26
iii. Average Sentiment Score .....	27
iv. Platform-Specific Linguistic Analysis of Goodreads and Amazon Reviews .....	28
<b>Key Findings .....</b>	<b>30</b>
<b>Discussion .....</b>	<b>31</b>
<i>Chapter III .....</i>	<i>33</i>
<b>The Shifting Notions of Identity, Community and Ownership Reflected Through the Datafication of the Contemporary Book Culture .....</b>	<b>33</b>
<b>Data-Generated Algorithms of Book-Centric Platforms: Is Privacy at Stake?.....</b>	<b>37</b>
<b>The Notion of Control in the Contemporary Digital Book Culture.....</b>	<b>41</b>
<b>Conclusion.....</b>	<b>44</b>
<i>Appendix.....</i>	<i>48</i>
<i>Bibliography.....</i>	<i>53</i>

### ***Abstract***

The present study employs a combination of qualitative and quantitative methods in order to establish the concept of datafication within the contemporary book culture, using the Goodreads platform as a primary case study. Through a close examination of the platform functionalities as well as of their implications for the overall reader experience today, the study attempts to analyze the changes that the current book culture has undergone in its hybrid form and to identify the dynamics between modern bookish platforms and their users. Upon conducting data analysis on Goodreads and Amazon reviews respectively, along with the observation of the shifting notions of Identity, Community, Ownership, Privacy and Control, it becomes evident that the modern reader is not merely a passive entity limited to the traditional 'Reader' archetype but rather emerges as a crucial agent of the contemporary book culture within the digital realm. Whether through their user-generated data or the composition of extensive reviews on popular titles, present-day readers possess the ability to determine emerging bookish trends and ultimately shape the book market. This pivotal role of the bookish-platform user culminates in a two-way relationship where users perpetually influence the platform they engage with, and vice versa.

### ***Acknowledgments***

First and foremost, I would like to express my deepest gratitude to my supervisor, professor Peter Verhaar for his encouragement of my research idea since our very first thesis meeting and for his ongoing support and guidance throughout this research journey. Your intellectual contributions as well as your practical assistance -such as with the data analysis- have been crucial to the completion of this thesis.

Secondly, I am eternally thankful for my family and friends for their (emotional) support and the patience that they showed throughout the process, even if they did not fully grasp the subject matter. It's safe to say that you kept me sane!

### ***Introduction***

The experience of reading a book, much like many others in the contemporary realm, has been digitized as well as digitalized. In terms of the former category, the book has been altered from a physical item to a numerical format, exemplified by the emergence of the e-book. Simultaneously, it has been digitalized as it has greatly shifted the societal perception of what a book as an item, the reader as an agent and

reading as an experience can entail within the context of contemporary digital media. In recent years digital book-centric platforms have managed to gain immense popularity among new age bibliophiles, with Goodreads, a social reading-focused website rendered as the most popular one, as well as various other alternatives such as the book-tracking LibraryThing or the –mostly data driven– Storygraph. Both as social networking platforms as well as book tracking ones, such websites have managed to play a pivotal role in this transformation of the Book into a cultural and increasingly social object in the twenty-first century. In these algorithmically driven platforms reader behavior is monitored through user generated data used to feed a recommendations algorithm aiming to provide personalized content. Reader behavior patterns are identified by these algorithms, contributing to the phenomenon named by journalist Chris Anderson as the ‘positive feedback loop’.<sup>1</sup> This loop entails users positively engaging with recommended works and subsequently providing positive feedback through reviews about these works. These platforms also allow users to ‘catalogue and annotate their book collections, connect with those of similar tastes, receive and make recommendations, and rate and discuss particular titles’,<sup>2</sup> while converting this user-generated content into metadata used to feed into the website algorithms and make the overall browsing experience more personalized to users. This way, acts such as reviewing on platforms like Goodreads or Amazon are entirely datafied; ‘criteria of similarity and difference—the forms of “value” identified and exchanged by reviewers—are coded and automated’.<sup>3</sup>

In short, the contemporary reading culture as a whole is intricately linked to data. Today, more than ever data is omnipresent in every aspect of the individual’s digital part of their reading activity to an extent that ‘every login, every page view, and every click leaves a digital trace’.<sup>4</sup> This data, resulting from one’s algorithmic

---

<sup>1</sup> C. Anderson, ‘The Long Tail’, in *The Social Media Reader* (New York University Press, 2020), ed. by Michael Mandiberg, p. 137.

<sup>2</sup> S. Murray, ‘Charting the Digital Literary Sphere’, *Contemporary Literature*, 56.2 (2015), p.324 <<https://doi.org/10.3368/cl.56.2.311>>.

<sup>3</sup> D. Wright, ‘Literary Taste and List Culture in a Time of “Endless Choice”’, in *From Codex to Hypertext: Reading at the Turn of the Twenty-first Century*, ed. by Anouk Lang (University of Massachusetts Press, 2012), p. 113.

<sup>4</sup> T. Gillespie, ‘The Relevance of Algorithms’, in *Media Technologies: Essays on Communication, Materiality, and Society*, ed. by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (The MIT Press, 2014), p.170.

identity<sup>5</sup> and the digital footprint left through their online activity, has given rise to a hybrid reading experience where the act of reading a physical copy of a book can be reviewed and discussed online. More specifically, while printed books do maintain their popularity and are widely used and even preferred by new age bibliophiles, these printed works are ‘created, promoted, sold, evaluated, consecrated, consumed and debated within a digital agential mesh’.<sup>6</sup> Of course, the online discussion of literary titles –or ‘book talk’<sup>7</sup>– is not solely generated on exclusively bookish platforms but can also be found in Social Networking Services (SNSs) such as Instagram, TikTok and YouTube. On the specific platforms, niche bookish communities can be found under the #bookstagram, #booktok or #booktube hashtags respectively, under which a community of aspiring micro-influencers gains a following by sharing their thoughts and reviews of recently read or currently reading titles. However, on this type of bookish communities, the book is oftentimes not approached as a cultural object but as a visual one, with the book cover playing a vital role in the reviewing process. In this paper I will be specifically focusing on Goodreads as the primary exclusively bookish platform at the time of writing and on the reader community that is cultivated on the platform. I will also critically examine the concept of the datafication of the contemporary reading culture through conducting data analysis on the Goodreads and Amazon reviews in order to establish any platform-specific behaviors. Furthermore, I will explore the social implications of this self-quantification process that the reader/user undergoes within the contemporary bookish culture, particularly focusing on the shifting notions and new affordances of identity, community, privacy, ownership, self-sovereignty and overall control.

The observation of user data and the algorithms generated through it on platforms such as Goodreads and Amazon appear to be of great importance in order to comprehend user and particularly reader behavior online as well as offline. Today, even if a reader uses the Goodreads platform in a passive manner, meaning without writing extensive reviews or even engaging with other users on the platform, they will inevitably become immersed in the Goodreads algorithm. After all, we are arguably living in ‘the “age of big data” [while] witnessing a quantitative “revolution” in

---

<sup>5</sup> S. Murray, ‘Secret Agents: Algorithmic Culture, Goodreads and Datafication of the Contemporary Book World’, *European Journal of Cultural Studies*, 24.4 (2021), p.978.  
<<https://doi.org/10.1177/1367549419886026>>.

<sup>6</sup> Murray, ‘Secret Agents’, p. 971.

<sup>7</sup> Murray, ‘Charting the Digital Literary Sphere’, p. 312.

human knowledge, driven by accompanying forms of data mining and analytics'.<sup>8</sup> During these times, according to Fuller and Sedo, ““reading with algorithms” is a necessary practice for readers who use social media to find out about books’.<sup>9</sup> This unavoidable algorithmic encounter arguably shifts and potentially shapes the reader’s perception on the act of reading, a fact that largely stems from the functionalities available on the Goodreads platform. More specifically, due to the fact that algorithms are essentially meant to create ‘normative models about readerly consumption’<sup>10</sup> through processing user-generated data, at the same time, these models also inevitably shape the reader’s perspective towards readerly consumption as well, thus creating a bidirectional relationship between users/readers and algorithms. In addition to that, the importance of data as an indicator of readerly consumption is also reinforced by certain policy changes from the part of the bookish platform, exemplified by the discontinuation of public API developer keys from December of 2020 onwards. This move from Goodreads signals a conscious choice to prevent developers from accessing data on the platform and/or manipulating it in order to create new reading lists, review aggregators, accessing user information etc. for personal gains, ultimately safeguarding their algorithmic data.

However, this user-generated data is not solely used to establish new and creative bookish possibilities for readers but could also become unreliable in specific contexts, this way potentially causing various conflicts on the platform. A recent instance of scandals spreading as a direct consequence of unreliable data on the Goodreads platform occurred in the recent December of 2023. Through the review functionality of the bookish platform the sci-fi novelist Cait Corrain ‘created fake accounts to leave positive ratings for her upcoming novel and sabotage other sci-fi novelists by leaving negative reviews on their work’.<sup>11</sup> The novelist then publically admitted her offence (which was referred to as ‘review-bombing’) on her personal social media accounts, a statement which ultimately did not stop from having her novel pulled from publication and her future book deal being cancelled. Following

---

<sup>8</sup> M. Andrejevic, Mark, A. Hearn, and H. Kennedy, ‘Cultural Studies of Data Mining: Introduction’, *European Journal of Cultural Studies*, 18.4–5 (2015), p. 379.

<sup>9</sup> D. Fuller and D. R. Sedo, ‘Young Adult Readers and the Genres of Online Book Reviewing’, <<https://post45.org/2023/12/it-was-plastered-all-over-my-instagram-last-year-at-least-for-my-algorithm-young-adult-readers-and-the-genres-of-online-book-reviewing/>>.

<sup>10</sup> Fuller and Sedo, ‘Young Adult Readers and the Genres of Online Book Reviewing’.

<sup>11</sup> A. Alter, ‘She “Review Bombed” Other Writers. Then Her Book Got Pulled’, *The New York Times*, 12 December 2023, <https://www.nytimes.com/2023/12/12/books/cait-corrain-goodreads-reviews.html?searchResultPosition=1> [accessed on 10 March 2024].

this incident, the Amazon-owned platform announced that they would modify their functionalities in order to ‘limit or suspend ratings and reviews “during times of unusual activity” [and ultimately] avoid review bombing’,<sup>12</sup> in an attempt to prevent such occurrences from happening in the future. The lack of reliability in the data that led to the scandal could also be considered a consequence of the Goodreads platform lacking adequate human moderation of its data, as it is still largely functioning under a voluntary librarianship program, namely ‘Goodreads librarians’. This program enlists volunteers to ‘help ensure the accuracy of information about books and authors in the Goodreads’ catalogue’,<sup>13</sup> as stated on the platform itself. This information suggests that the multimillion platform still operates with anachronistic features, a fact that proposes a stagnation even after its acquisition by Amazon. However, the possibility of hiring these people would arguably help ensure the platform’s smooth operation and human monitoring of the undergoing algorithmic processes. In essence, this example further demonstrates the various ways in which platform functionalities, like composing reviews, can be utilized for unfavorable purposes such as the ability to review titles on Goodreads prior to their official publication, or anonymously leave intentionally negative reviews on already published works. Consequently, such actions may have significant repercussions on the prevailing reading culture as well as on the individuals within this culture, specifically readers and authors.

Naturally, Amazon is also a pivotal agent in this contemporary book culture; not only because of the cheaper prices in books and an offering of the widest variety of titles online, but as it constitutes the owner of the Goodreads platform from 2013 to this day. The e-commerce giant is presently rendered as the most prominent book seller, with over 32.8 million published titles in total (as of 2023) as well as its rendering as the biggest distributor in the US, as it is responsible for 80% of book distribution in the States.<sup>14</sup> This focus on the book world was not new for the big-five corporation; the book zeitgeist actually constituted Amazon’s ‘first foray into online retailing’.<sup>15</sup> Books and the act of reading in particular remain a vital part of Amazon’s interest, with a prime example being the creation of the Kindle in November of 2007,

---

<sup>12</sup> Alter, ‘She “Review Bombed” Other Writers. Then Her Book Got Pulled’.

<sup>13</sup> Goodreads, ‘About Goodreads Librarians’, <<https://help.goodreads.com/s/article/About-Goodreads-Librarians>> [accessed 10 March 2024].

<sup>14</sup> D. McLoughlin, ‘Amazon Book Sales Statistics’, *WordsRated*, 9 November 2022, <<https://wordsrated.com/amazon-book-sales-statistics/>> [accessed on 10 March 2024].

<sup>15</sup> Murray, ‘Secret Agents’, p.972.



a handheld reading device and service that remains largely synonymous with the act of digital reading today. In addition to that, Amazon has recently launched its own book tracking feature called ‘Your Books’, a book discovery and recommendations feature similar to Goodreads, but which will be specifically linked to an Amazon account and therefore will focus on the book as a product available to encounter exclusively through the Amazon ecosystem. This feature will encompass physical copies, e-books found on a Kindle as well as audiobooks from the Amazon-owned Audible service.<sup>16</sup> Despite the fact that ‘Your Books’ is not available to the general public at the time of writing, Amazon’s interest in dominating the contemporary book world and ultimately altering the act of reading as well as the predominant book culture is undeniable.

The introduction of a new feature so close in similarity to Goodreads, coupled with the stagnation of the aforementioned platform following its purchase by Amazon in terms of updates in user functionalities, technological infrastructure and data security can be regarded as a prime example of a large corporation attempting to establish and uphold a monopoly of the book market and, hence of the book culture as a whole. This hegemony appears to be intended to be sustained by leveraging the user data generated on the Goodreads platform such as reviews, recommendations and genre preferences after the acquisition, this way further establishing the significance of this data in comprehending and shaping the reader experience. When asked about this approach in a Washington Post interview, former Goodreads employees stated that ‘Amazon seemed happy to mine Goodreads for its user-generated data and otherwise let it limp along with limited resources’.<sup>17</sup> Nevertheless, at the time of writing, Goodreads remains the most popular book tracking and exclusively book-oriented tool on the digital realm, with around 125 million members by late 2022.<sup>18</sup>

---

<sup>16</sup> S. Perez, ‘Amazon Competes with its Own Goodreads with Launch of Book Discovery Service, “Your Books”’, *TechCrunch*, 13 December 2023, <[https://techcrunch.com/2023/12/13/amazon-competes-with-its-own-goodreads-with-launch-of-book-discovery-service-your-books/?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\\_referrer\\_sig=AQAAAAYnBNMuGs8rnHg7DRByE1U8VFb8sXn9cE2fT\\_XnWMOOnoDnXLSe5DsuFDjeCmo-ij8l\\_TK1OSQLkzc471laRZcAqNttfHMYSqcvetv8X3TXzJduber3U4ISxzDcfmpML9McMVmtn416DHF5oBmpJoO7OVmAdQol4nu2wJlZQZjNw](https://techcrunch.com/2023/12/13/amazon-competes-with-its-own-goodreads-with-launch-of-book-discovery-service-your-books/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAAYnBNMuGs8rnHg7DRByE1U8VFb8sXn9cE2fT_XnWMOOnoDnXLSe5DsuFDjeCmo-ij8l_TK1OSQLkzc471laRZcAqNttfHMYSqcvetv8X3TXzJduber3U4ISxzDcfmpML9McMVmtn416DHF5oBmpJoO7OVmAdQol4nu2wJlZQZjNw)> [accessed on 10 March 2024].

<sup>17</sup> C. O’Donovan, ‘Goodreads Was the Future of Book Reviews. Then Amazon Bought It’, *The Washington Post*, 1 July 2023, <https://www.washingtonpost.com/technology/2023/07/01/amazon-goodreads-elizabeth-gilbert/> [accessed 10 March 2024].

<sup>18</sup> M. Kreizman, ‘Let’s Rescue Book Lovers From This Online Hellspace’, *The New York Times*, 24 December 2023, <https://www.nytimes.com/2023/12/24/opinion/goodreads-books-reviews.html?searchResultPosition=2> [accessed 10 March 2024].

As such, the platform constitutes an ideal case study for understanding the increasingly important role of data in the contemporary book culture.

### ***Methodology***

For the purpose of this research, a combination of qualitative and quantitative methodologies will be employed. Firstly, in regard to the qualitative segment, the Goodreads platform will be examined as a case study in terms of the functionalities it offers to its users as well as in order to observe how these features shape the contemporary readerly culture. More specifically, by drawing on these functionalities as well as on already existing literature, the research will focus on the potential implications of these features for the users and the reading culture as a whole as well as on their role in shaping the contemporary reading culture. In regard to the quantitative methods, the Goodreads API (Application Programming Interface) tool will be utilized as the source of the review data for the analysis of all English titles of The New York Times weekly bestseller lists from the year 2020, which will be used as samples. A text mining tool will then be utilized to conduct the analysis of the Goodreads and Amazon reviews for these titles respectively, as well as to visualize the outcomes of this comparative analysis. This part of the research will build upon the already existing study conducted by Dimitrov et al.<sup>19</sup> on the comparative analysis of Goodreads and Amazon reviews, published in 2021. The focus of this comparative study was on reviewer behavior and engagement in both the Goodreads and Amazon platforms respectively, specifically within the single genre of biography. The parameters examined included reviewer abundance, review abundance, review length, review star ratings, review sentiment and platform-specific review language. The primary findings of this research mostly centered on identifying similarities and differences in regard to the aforementioned parameters. Lastly, the qualitative component of this research will involve the examination of the findings of the aforementioned data inspection of the two platforms as well as an exploration of already existing literature. The aim of this last qualitative methodology will be to identify the sociocultural consequences that these platforms have brought about for the contemporary readers, particularly in terms of the evolving notions and new

---

<sup>19</sup> S. Dimitrov and others, 'Goodreads vs Amazon: The Effect of Decoupling Book Reviewing and Book Selling', *Proceedings of the International AAAI Conference on Web and Social Media*, 9.1 (2021), 602-605 <<https://doi.org/10.1609/icwsm.v9i1.14662>>.

affordances of identity, community, ownership, privacy, self-sovereignty and overall control within the digital book culture.

### ***Limitations***

In order to establish the quantitative segment of the research, which will be conducted in the second chapter of the paper, it is essential to acknowledge certain limitations in advance. As stated previously, the sampling process will originate from the analysis of all English titles of The New York Times weekly bestseller lists from the year 2020. Firstly, due to the fact that the research focuses exclusively on the year of 2020, the issue of temporal bias becomes inevitable, as focusing only on one year will ultimately lead to an inability to identify and observe any trends or patterns in the reviews of titles over a larger time period. In addition to that, temporal events such as the Covid-19 pandemic can potentially impact the selection process in terms of the titles published and selected within this particular year. In addition to that, a geographical bias also has to be recognized. The New York Times selection of titles will inherently cater to the US book market which, despite constituting the largest book market along with the UK, does inevitably restrict the options of titles available. Furthermore, due to the fact that the selection is based on the US market, there also arises a language limitation, as all titles listed as booksellers are in English, including the selection of Goodreads and Amazon reviews on these titles. Moreover, there is a level of representation bias, as the New York Times bestsellers list will ultimately overlook any niche publications or self-published work from that year, this way focusing more on widely known releases from larger publishers. Finally, titles marked as ‘bestsellers’ by a prominent international publication such as The New York Times may inherently attract more favorable reviews from readers, stemming from the perceived quality indicator of ‘New York Times bestsellers’.

## Chapter I

### ***The Datafication of the Reading Experience on Goodreads***

Goodreads, as the most prominent book-tracking and social reading platform constitutes the primary example of the datafication of the reading culture within the contemporary online realm. Otis and Elizabeth Chandler, a young couple with a passion for books, established the platform in 2006 and launched it in January 2007.

The current ‘about’ page of the website, written and signed by co-founder Otis, reads like a heartfelt letter to a friend. The co-founder explains that the idea for the platform essentially originated from him ‘scanning a friend’s bookshelf for ideas’.<sup>20</sup> This led to the realization that this actually seemed to be the most efficient method for obtaining reliable book recommendations; consulting the literary preferences of a like-minded individual. The initial setup of the website is also described in a warm and familiar tone by Otis; ‘we started in my living room, motivated by the belief that there was a better way to discover and discuss good books, and that we could build it’.<sup>21</sup> Keeping these words in mind, it is safe to assume that, when designing Goodreads in their living room, the two co-founders could have not possibly anticipated the immense popularity that their creation would have as well as its impact on the book culture as a whole.

However, the couple’s ultimate intention to provide a useful book-tracking tool for book enthusiasts in order to keep track of their reading as well as to create an online space for book lovers to interact, managed to ultimately constitute the epitome of the process of datafication of the book realm today. Interestingly enough, the acquisition of the platform by Amazon seems to have also played a crucial role in the process of datafication as well, as the Big Five corporation explicitly used Goodreads user-generated data for their own benefit, thus further validating the importance of this data as a reliable reflection of the state of the book market and an indicator of the purchasing preferences of its audience. Consequently, upon its acquisition, the Goodreads platform managed to shift its character from a cozy book-tracking and community-fostering platform created by book lovers for book lovers into a book-tracking, community-fostering tool with a focus on monitoring the trends of its literary community through user-generated data in order to effectively shape the trends in the book market. In this paper, I will also argue in favor of the idea that the current functionalities offered by the Goodreads platform shape the notion of the modern act of reading as well as the book culture as a whole, while at the same time, these notions are also constructed by the users themselves through the digital footprint left on the platform throughout their activity. More specifically, the architecture of the Goodreads website essentially provides the ‘guiding light for users’ identity

---

<sup>20</sup> Goodreads, ‘About Goodreads’, < <https://www.goodreads.com/about/us> > [accessed 15 March 2024].

<sup>21</sup> Goodreads, ‘About Goodreads’, < <https://www.goodreads.com/about/us> > [accessed 15 March 2024].

performances'<sup>22</sup> through the affordances available on the platform, setting specific parameters as to which acts are possible and acceptable within the digital space. Such acts are of course primarily influenced by political and economic dynamics as well as consumerist tendencies, as opposed to the users' best interest. This idea ultimately establishes a two-way relationship in terms of the impact of Goodreads users and the functionalities available on the bookish website. It is vital for users to recognize the affordances that shape this relationship, not solely for understanding how these functionalities affect one's 'interactions and broadcasting of the self, but as an indication of one's representation online.'<sup>23</sup> In order to establish this bidirectional relationship, the first step would involve the close examination of the functionalities available on the Goodreads platform at the time of writing.

### ***A Close Analysis of the Goodreads Platform Functionalities***

The Goodreads platform, according to the 'about Goodreads' page<sup>24</sup> essentially provides its users with two main types of functionalities. On the one hand, the platform was primarily established in order to track the books that a user has read, is currently reading or intends to read, therefore one essential type of functionalities would be the ones that have a direct book-tracking purpose. On the other hand, the page also highlights users' ability to 'see what their friends are reading',<sup>25</sup> a concept that alludes to the wide range of social and community-oriented type of functionalities present on the bookish platform and that have essentially contributed to establishing Goodreads as a social reading platform today. Indeed, it is evident that Goodreads currently features a digital book community that is 'global, multiple, and dynamic, and that [ultimately] brings into visibility an entirely new social dimension to reading'.<sup>26</sup> Nonetheless, besides these two feature types that are available for the Goodreads user, there exists another category of functionalities that is not explicitly mentioned in the 'about' webpage, as in the case of previous two categories. Amazon-related functionalities could serve as a third category of features presently available on the bookish website, with a couple of discreet features referencing either the

---

<sup>22</sup> A.M. Cirucci, 'Redefining Privacy and Anonymity through Social Networking Affordances', *First Monday*, 20.7 (2015).

<sup>23</sup> Cirucci, 'Redefining Privacy and Anonymity through Social Networking Affordances'.

<sup>24</sup> Goodreads, 'About Goodreads'.

<sup>25</sup> Goodreads, 'About Goodreads'.

<sup>26</sup> J. Pinder, 'ONLINE LITERARY COMMUNITIES: A Case Study of LibraryThing', in *From Codex to Hypertext: Reading at the Turn of the Twenty First Century*, ed. by Anouk Lang (University of Massachusetts Press, 2012), p. 68.

Amazon website or the Kindle, which is essentially an Amazon-affiliated product. I will first name and analyze the functionalities found within each of the categories and proceed to explore the potential implications of these features in the process of the datafication of the contemporary reading culture.

The essential and more straightforward functionalities of Goodreads are book-tracking ones. As mentioned before, the primary reason behind the creation of the platform essentially revolved around facilitating the process of a book lover monitoring the books they read in the past, present and future. This is mainly established through the ‘bookshelves’ feature found in every profile under the ‘my books’ tab. Three main classifications of bookshelves are readily available to users: ‘Read’, ‘Currently Reading’ and ‘Want to Read’. Additionally, there is also an ‘All’ segment, enabling one to browse the entirety of the titles included in their bookshelves. However, apart from these premade bookshelves, users also have the option to create new bookshelves for distinct purposes through the ‘Add a Shelf’ option. By clicking on this tab, users can give a name to their new shelf and commence the process of adding their preferred titles. In regard to the arrangement of these bookshelves, they are chronologically organized based on the time of their addition to the respective bookshelf. For instance, upon browsing the ‘all’ bookshelf (see Figure 1), which includes all the titles that a user has ever added to any of their bookshelves, categorized into ‘cover’, which refers to the cover of the saved book, ‘title’, ‘author’, ‘average rating’, referring to the average rating of the book by users who have already read and rated it until that point, ‘rating’, which refers to the rating given by the user, ‘shelves’, referring to the relevant bookshelf that the book has been added to, ‘review’, which refers to the review that the user has written on the respective title, ‘date read’, which refers to the date on which the user marked the book as ‘read; on the platform and ‘date added’, meaning the date on which the user added this book to any of their bookshelves. By default, this list is displayed in a linear way, although an alternative to be arranged by book cover is also available. Furthermore, users can also click on any book and find a ‘reading progress’ timeline, featuring its shelving date by the user, the date that the user started reading the book as well as any progress that has been tracked by the user until the date of its completion. Lastly, an ‘add comment’ function exists on the page, in order for the users to provide insights of their reading experience through comments shared with their followers.

Ratings and reviews are perhaps the most prominent functionalities of the Goodreads platform and they could be regarded as simultaneously occupying two different categories; the book-tracking and community-building ones. On the one hand, star ratings are the most direct method in order to evaluate a book on the platform, where users are prompted to assign a rating to any recently read book according to a five-star rating system, without the option of using half points. This feature partly plays into the Goodreads as a book-tracking function of the platform as the user is given the ability to monitor their preferences in specific titles at any given time. However, it also contributes to the community-building aspect of the website by displaying an average rating as well as the total number of ratings for each book. In addition to that, this central placement of the ratings on a book's profile page, after the title and author (see Figure 2), suggests that ratings are a vital factor in quickly assessing the quality of a book on the platform. Respectively, reviews, which are essentially longer textual expressions of users' opinion on a title, inherently offer a more holistic and comprehensive insight of the users' opinion that often complements and further expands upon their star rating. Reviews can be considered a book-tracking tool, as they can be used in order for the user to record their opinion on a title and possibly reflect on their opinion of the publication over time. However, these reviews can also serve as a community-building tool, akin to the comment sections found on typical Social Networking Services (SNSs). In the review section of a title, users do not solely write reviews for feedback sake, but also in order to interact and initiate discussions with other reviewers. This way, reviews, inherently assuming readership, essentially 'reach out to readers, authors, characters, and people in reviewers' lives, assembling complex social networks of reading'.<sup>27</sup> This engagement with reviews also helps to facilitate algorithms in tracking users' interests, linking them with other users with similar literary taste and finally deriving book recommendations from the shared taste patterns on the platform.

In addition to that, Goodreads reviews are an important feature of the platform as, through their community-building nature, implicitly but fundamentally contribute to the democratization of literary criticism. More specifically, the bookish platform allows for all users who have read a book to write their personal opinion on the

---

<sup>27</sup> B. Driscoll and D. R. Sedo, 'Faraway, So Close: Seeing the Intimacy in Goodreads Reviews', *Qualitative Inquiry*, 25.3 (2019), p. 254.

subject matter and publicize it for other bibliophiles to see. Through this process, the platform establishes the concept of literary criticism as not solely an opportunity for the selected few literary critics anymore; any well-thought out review can shape its reader's perception and lead them to make a final decision on whether to invest (both financially and intellectually) on the respective title or not. In this sense, on the Goodreads realm, the readers themselves constitute both the critics and the buying audience of a title. It can also be argued that this impact that Goodreads reviewers exert on the reputation and therefore (commercial) success of a work, a privilege which was previously exclusively held by professional literary critics, may be viewed as one of the motivating factors behind Amazon's acquisition of the bookish platform in the first place. Lastly, as posited by author Clay Shirky, the primary shift presented by this reviewing model is the transition from a 'filter then publish' toward a 'publish then filter' approach. Within this framework, anyone can publish a review on platforms such as Goodreads or Amazon, but the responsibility of filtering this review is placed upon the reader.<sup>28</sup>

Apart from book reviews, there are numerous other purely community-building functionalities on Goodreads. However, these social functionalities do not exist independently from the book-tracking ones, as the Goodreads algorithm essentially seems to use the data from the latter for the reinforcement of the community-building functionalities. These communal aspects bear a close resemblance to the design of social media platforms such as *Facebook*, *Instagram*, *X* and others. This similarity is reinforced by the fact that users of the bookish platform have the ability to find and connect with friends from other social media websites. In addition to that, users can also follow beloved authors and view their tracked activity, participate in user-generated discourse within discussion boards on bookish matters or privately discuss books with friends through a personal inbox feature. In addition to these functions, the 'Community' tab located at the website homepage provides even more community-oriented functionalities such as the 'groups' feature which showcases various groups formed by users who share interests in specific genres, as in the case of the 'Horror Aficionados' group which counts 25.500 members, celebrity book clubs such as 'Oprah's Book Club', or groups of fans of a particular author, for instance, the 'Stephen King Fans' group. Ultimately, the various

---

<sup>28</sup> Pinder, 'ONLINE LITERARY COMMUNITIES: A Case Study of LibraryThing', p. 75.



functionalities under the ‘community’ tab are fundamentally designed to ‘give users different mechanisms for participation and engagement, according to their different skills and interest’.<sup>29</sup> This, in turn, enables them to generate more user-generated data that can be incorporated into the Goodreads algorithm.

The last category of functionalities that can be detected on the Goodreads platform are Amazon related features. These particular features have been incorporated in the platform upon the acquisition of the website by the corporate entity and are meant to establish links between the two platforms in a subtle manner. More specifically, Goodreads refrains from explicitly stating its affiliation with Amazon on their website, opting instead to foster this relationship through providing discreet features that create connections between the two platforms. The most prominent feature that serves this connection is the ‘buy on Amazon’ option that accompanies each title that is catalogued on the Goodreads website, positioned under the ‘want to read’ option that relates to mere reader activity (see Figure 2). Of course, other platforms are also listed as viable options to purchase the respective title but –interestingly enough– the Amazon option is the default selection, while in order to access the list of other retailers a user should manually click on a small arrow beside the Amazon option. Additionally, Amazon had explicitly communicated their intention to connect the Kindle device with the Goodreads platform from the beginning of their ownership. Consequently, it is not surprising that a wide range of Kindle functionalities are available on the bookish website, including the ‘Kindle Notes & Highlights’ section within ‘my books’. This feature essentially enables Kindle users to automatically upload their annotations from their device to their Goodreads profile and share them with their friends and followers. In this sense, the Kindle, and by extension the Amazon account, is connected to the Goodreads account of the respective reader. Nevertheless, this correlation also extends beyond the Kindle device, as certain regions offer the option to immediately synchronize a user’s Amazon book purchases to their Goodreads profile by connecting their Amazon and Goodreads accounts.

---

<sup>29</sup> Pinder, ‘ONLINE LITERARY COMMUNITIES’, p. 80.

### *The Possible Implications of these Features for the Digital Reader Experience*

Arguably, the type of functionalities that are present on the Goodreads platform also serve to implicitly showcase the underlying motives driving the bookish website. For instance, the wide range of features that fall under the ‘community building’ category further reinforce the aim of the bookish platform to foster and cultivate a bookish community as well as to reinforce the practice of social reading. Overall, the platform seems to have been ‘clearly designed to enhance the social aspects of book reading and reviewing’.<sup>30</sup> A prime example of this transition towards a more social reading-oriented platform can be evident through the prompts that are featured primarily in the ‘currently reading’ bookshelf, where imperatives like ‘update your progress’, ‘rate this book’ and ‘share’ essentially encourage the user to engage and share their reading activity with their followers, therefore ultimately cultivating and subsequently establishing a ‘strong focus on the social nature of literary culture’.<sup>31</sup> Of course, this emphasis on the social aspect of reading cannot be rendered a new idea. On the contrary, the art of storytelling has existed for thousands of years while book clubs, which are also established on the basis of this collective reading concept, have been around since 1634, when Puritan religious advisor Anne Hutchinson began a scripture reading circle during her boat ride from England to the Massachusetts Bay Colony.<sup>32</sup>

However, this emphasis on the community-building functionalities along with the limited availability of book-tracking activity tools on the platform suggests a shift in the founders’ original vision of creating a (predominantly) book-tracking tool. In addition to that, this shift can also be regarded as part of a greater cultural trend according to which the Internet is increasingly used as a means of establishing and maintaining social –and specifically parasocial–<sup>33</sup> connections, with a prime example being the growing popularity of mainstream social media platforms within the last two decades. In this sense, Goodreads today can be regarded as a bookish social

---

<sup>30</sup> A.B. Albrechtslund, ‘Amazon, Kindle, and Goodreads: Implications for Literary Consumption in the Digital Age’, *Consumption Markets & Culture*, 23.6 (2020), p. 557.

<sup>31</sup> J. Vlieghe, M. Jaël and K. Rutten. ‘Everybody Reads: Reader Engagement with Literature in Social Media Environments’, *Poetics*, 54 (2016), p. 27.

<sup>32</sup> S. Polk, *The Long Legacy of Book Clubs* (2023) <<https://www.shondaland.com/inspire/books/a45563057/the-long-legacy-of-book-clubs/>> [accessed 10 April 2024].

<sup>33</sup> This form of digital relationships is forged between an individual and someone who they do not know personally. Particularly in the realm of social media, individuals tend to form parasocial bonds with influencers or other content creators whose content they interact with on a daily basis. Of course, such relationships lack mutuality and they are oftentimes susceptible to exploitation, more often than not from the part of the content creator.

media platform, as it ‘combines the friend relationship and communication elements of these generic social network sites’,<sup>34</sup> with book-tracking functionalities assuming a secondary role for contemporary users. Of course, these types of functionalities are not separate but interconnected, as the process of discovering new users who share similar literary interests on the bookish platform can influence the recommendations algorithm by leveraging user-generated data. In this sense, book-tracking functionalities currently found on the platform such as the ‘read’, ‘currently reading’ and ‘to be read’ bookshelves cannot be regarded solely as book-tracking tools in the datafied book culture. They are also used by the Goodreads and Amazon corporations as functionalities that facilitate ‘literary self-cataloguing’,<sup>35</sup> this way establishing the ‘allure of self-quantification via curating one’s reading life for a global audience of bibliophiles’,<sup>36</sup> a process that arguably could not have been predicted by Chandler in 2006.

Another observation that may hint at a shift from the original motivation behind the establishment of the Goodreads platform is the incorporation of Amazon-related functionalities, which essentially foster a commercial approach on the bookish website that was arguably not evident prior to the acquisition by the mentioned corporation. Consequently, a division has emerged between co-founder Otis Chandler’s initial intimate vision of creating a website to facilitate the process of exchange literary views among friends and Amazon’s more business-oriented and market-driven context. This shift from an intimate website to a commercially focused one can be detected merely through the act of writing a review on the platform. Upon crafting a review, an act which is automatically proposed when marking a book as ‘read’ on Goodreads, the user is required to address this review not solely to their Goodreads friends but to every user who may encounter the review section of the respective title. Therefore, the act of reviewing constitutes an inherently performative act on the platform, akin to ones detected on generic social media platforms. In addition to that, in the case that a user might read the review and attempt to purchase the respective book through the Goodreads platform, they will automatically be

---

<sup>34</sup> M. Thelwall, and K. Kayvan, ‘Goodreads: A Social Network Site for Book Readers’, *Journal of the Association for Information Science and Technology*, 68.4 (2017), p. 972.

<sup>35</sup> Murray, ‘Secret Agents’, p. 977.

<sup>36</sup> Murray, ‘Secret Agents’, p. 977.

redirected to the Amazon website, a fact that further reinforces the commercial essence that has been instilled within the platform by its current owner.

Through the utilization of the existing functionalities provided by the Goodreads platform, the users are prompted to track their reading habits, preferences and opinions about books, thus influencing the design of the Goodreads platform based on their personal activity through user-generated data. However, at the same time it is essential to acknowledge that the functionalities present on a platform inherently impose certain constraints and limitations in order to target a specific audience and monitor –or to an extent control– the activity and liberties of users on the platform. In essence, when we think about Goodreads as reflective of the contemporary book culture, it is also necessary to consider the range of functionalities offered by the platform to its users and assess the degree of limitations that they themselves may impose in their role of influencing the book culture. The default functionalities found on Goodreads essentially shape the prevalent norms around the hybrid book culture of today while also reflecting the reading habits of Goodreads users, consequently constituting both a trend-setter and a reflection of the trends set within the contemporary book culture. Therefore, if one wishes to ponder on the current state of the book culture, they can just take a look at the functionalities provided on the bookish website.

This reflection is not limited to trends regarding popular book titles, genres or authors through the star ratings and reviews as indicative of the current book market but can be extended to trends in reader behavior as well. For instance, the annual ‘Reading Challenge’ that the platform offers can be regarded as reflective of a particularly competitive approach to reading. More specifically, this reading challenge is a functionality according to which a Goodreads user can set a goal of the number of books they would like to read within the timeframe of a year. Users can set the number of books they would like to read at any point during the year and can change the amount at any moment. This functionality is seemingly meant to enhance the act of reading among users and help challenge their reading abilities, a motive evident from the motto of the challenge found on the website: ‘challenge yourself to read more this year!’. Interestingly enough, statistics from past years are also available. For the year 2023, almost 8 million users took part in the challenge and a total of 342 million books were registered as read with the average reading goal of a user being 43

books.<sup>37</sup> Through the challenge webpage a user can also see the results that their Goodreads friends scored for the past year under the ‘friends’ tab as well as the results of other users under the ‘community’ tab. Apart from the face-value motivation that the platform states as purely challenging and an endeavor to enhance user engagement in the act of reading, it could also be interpreted as cultivating a sense of competitiveness which ultimately leads in a rise in the volume of purchases. More specifically, this competitive undertone ultimately culminates in a process of commercial performative consumption on the platform, a model according to which establishing oneself as a ‘reader’ through registering multiple books on one’s Goodreads profile becomes more important than actually reading the books. This approach holds particular significance in an increasingly datafied book culture, where there is a prevalent commercialization of ‘clicks and views [being] monetized, [while] seeing printed books on screen [has become] a means of consuming them’.<sup>38</sup> In short, the more books a user adds to their lists, the more likely they would be to accomplish their yearly ‘Reading Challenge’ and the more profit would be generated from book-selling platforms, particularly Amazon, which is essentially the default book selling platform on the Goodreads website.

## Chapter II

### ***Goodreads as an Undertaking of Amazon: Possible Consequences for the Reading Culture***

As previously discussed, the acquisition of Goodreads by Amazon has arguably constituted a significant factor in shaping the current bookish platform. On a theoretical level, this shift in the Goodreads platform could have potentially played a role in the broader transformation of the contemporary reading culture. Arguably, this transformation can be regarded as a direct consequence of the Amazon platform constituting a very special agent in the book ecosystem. Robert Darnton’s infamous ‘Communication Circuit’<sup>39</sup> represents the closed system of the book market, through which books are conceived, designed, produced and consumed. Established in 1982,

---

<sup>37</sup> Goodreads, ‘2023 Reading Challenge’, < <https://www.goodreads.com/challenges/11633-2023-reading-challenge>> [accessed 12 April 2024].

<sup>38</sup> A. Lastoria, ‘Digital Masks of Printed Books: On-Screen Representations of the Materiality of the Codex’ in *Bookshelves in the Age of the COVID-19 Pandemic*, (Berlin: Palgrave Macmillan, 2022), p. 137.

<sup>39</sup> R. Darnton, ‘What Is the History of Books?’, *Daedalus*, 111.3 (1982), p. 68.

decades before the digital book era came into being, the Amazon paradigm would arguably assume a rather distinct position in this circuit. More specifically, theoretically speaking, it would certainly occupy the roles of the publisher, printer, shipper and bookseller, thereby controlling more than half of the circuit. Consequently, it would establish a sort of hegemony in the field and therefore would constitute a vital influence on the contemporary book market as a whole. In greater detail, in terms of its role as the publisher, Amazon Publishing is a publishing house established by the Big Five corporation in 2009 and manages 17 imprints under its name at the time of writing, covering various genres from adult to children's literature.<sup>40</sup> Apart from this venture into traditional publishing, the Kindle Self-Publishing program, introduced in 2007 concurrently with the homonymous e-reader, also offers the ability for individuals to self-publish their print or digital books through Amazon. Furthermore, through this self-publishing service Amazon also ventures into the printing industry, as they offer a Print-on-Demand service for their exclusive self-published content. Shipping services are also provided by the company through its warehouses, where they store their stock of books and other products. Interestingly enough, upon a BBC investigation into the working conditions at these warehouses conducted in 2013, the conditions were described as comparable to a 'slave camp', with workers being pressured to fulfill orders every 33 seconds<sup>41</sup> while their 'productivity was extensively tracked with a degree of exactitude previously unimaginable'<sup>42</sup> by the company. In addition to that, a significant amount of these orders, namely 10%,<sup>43</sup> pertain to Amazon's primary role, the one of the bookseller. Amazon has been the predominant bookseller since its establishment, with at least 300 million paperbacks being sold every year through the platform and 487 million e-books through the Kindle platform.<sup>44</sup> These figures rightfully render the platform as the leading online bookseller today, with an estimated 70% takeover of the US book

---

<sup>40</sup> D. Curcic, *Amazon Publishing Statistics* (2023), *WordsRated*, 12 January 2023, <<https://wordsrated.com/amazon-publishing-statistics/>> [accessed 12 April 2024].

<sup>41</sup> A. Bennett, *Amazon Warehouse Staff in 'Slave Camp' Conditions, Workers Say* (2013), <[https://www.huffingtonpost.co.uk/2013/11/25/amazon-staff-investigation\\_n\\_4335894.html](https://www.huffingtonpost.co.uk/2013/11/25/amazon-staff-investigation_n_4335894.html)> [accessed 12 April 2024].

<sup>42</sup> L. Nakamura, "'Words with Friends': Socially Networked Reading on *Goodreads*", *PMLA/Publications of the Modern Language Association of America*, 128.1 (2013), p. 243.

<sup>43</sup> WordsRated, 'Amazon Book Sales Statistics' <<https://wordsrated.com/amazon-publishing-statistics/>> [accessed 15 April 2024].

<sup>44</sup> WordsRated, 'Amazon Book Sales Statistics'.

market in 2025 and at least 50% of the UK market,<sup>45</sup> known to be the two largest book markets globally. These numbers have also had a lasting negative impact on independent book retailers, with the numbers of such bookshops dropping by 63.2% between the years from 1995, when the Amazon platform was established, to 2022 according to statistics found on WordsRated.<sup>46</sup>

Keeping these statistics in mind, one could argue that Amazon maintains a monopoly in the contemporary book market, a fact that consequently leads to a monopoly of the broader prevailing book culture as well. Furthermore, the fact that Amazon ‘has attempted to concentrate all varieties of online bibliophilia under its own umbrella’<sup>47</sup> can also be traced through the Amazon website; Amazon, akin to Goodreads, functions as a ‘recommendation-based retailer’,<sup>48</sup> where the user-generated recommendations and reviews, mostly from avid users of the platform, essentially work as advertising tools for the respective titles. This behavior is not exclusively associated with Amazon or Goodreads users but appears to be a common and rather fundamental attribute of all platforms where online interactions take place among users and is known as ‘e-word-of-mouth’ (eWOM). This behavior seems to be intricately linked to new age marketing strategies that have emerged alongside the proliferation of the Internet and social media, leveraging the affordances of fostering relationships and networking, which are vital particularly in this case for social reading platforms. More specifically, the concept of eWOM can be defined as ‘any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet’.<sup>49</sup> In the context of bookish websites, of course, such as Amazon and Goodreads, the products under discussion are exclusively books, while the reader serves as the customer. This idea of eWOM appears to be of great importance for the platforms where it transpires, as it has the ability to exert substantial impact on customer views about products or services, this way ‘leading to changes in judgements, value ratings and likelihood of purchase’.<sup>50</sup> This significant influence of the particular strategy primarily stems from the fact that upon recommending a

---

<sup>45</sup> WordsRated, ‘Amazon Book Sales Statistics’.

<sup>46</sup> WordsRated, ‘Amazon Publishing Statistics’.

<sup>47</sup> Murray, ‘Charting the Digital Literary Sphere’, p. 325.

<sup>48</sup> Nakamura, “‘Words with Friends’: Socially Networked Reading on *Goodreads*”, p. 239.

<sup>49</sup> P.S. Kapoor, K.R. Jayasimha and A. Sadh, ‘Brand-Related, Consumer to Consumer, Communication via Social Media’, *IIM Kozhikode Society & Management Review* 2.1 (2013), p. 48.

<sup>50</sup> Kapoor, Jayasimha and Sadh, p. 51.

specific title as a bibliophile on the Amazon or Goodreads platform, users often associate such reviews with a notion of authenticity, akin to receiving a reliable book recommendation from a trusted acquaintance, ultimately leading to a ‘conflation of the commercial and the personal’<sup>51</sup> By expanding on this notion of eWOM, the Amazon algorithm seems to utilize ‘social listening and monitoring tools’<sup>52</sup> in order to collect this user-generated data and delineate the current trends of the book market, which are the reinforced and reproduced by more user-generated data. Consequently, this concept serves to reinforce the notion that all activities conducted in digital spaces, both through engagement with other users or with the interfaces themselves, are ultimately recorded and archived<sup>53</sup> for purposes that greatly remain undisclosed to users of social media platforms.

A prime example of the occurrence of eWOM is showcased in a survey conducted by Maity et al.<sup>54</sup> on the basis of whether Goodreads user reviews could predict bestseller titles on Amazon. According to the findings of this research, Goodreads users tend to post more status updates on titles that are listed as ‘Amazon bestsellers’, a fact that further reinforces a connection between the two book-oriented platforms as well. Particularly notable is also the fact that the Amazon corporation possesses Goodreads, therefore also exerting control over the user data generated on the bookish platform. It is only natural that the bookish website would be of interest for book industry professionals as it holds commercial value for them; ‘because the site contains millions of user reviews of books, it may also inform librarians and other professionals for purchasing strategies and other services’.<sup>55</sup> Consequently, this intertwining of factors elucidates the apparent Amazon hegemony within the book market.

However, these statistics also may imply that the influence of Amazon on the book culture is not genuine, as in the case of Otis Chandler wanting to create a bookish environment for avid readers, but that the approach of Amazon is consumer-based and focused on profit-generating. In this sense, Amazon can be regarded as manipulating user-generated data and contributing to the datafication of the

---

<sup>51</sup> Pinder, ‘ONLINE LITERARY COMMUNITIES: A Case Study of LibraryThing’, p. 71.

<sup>52</sup> Kapoor, Jayasimha and Sadh, p. 54.

<sup>53</sup> Cirucci, ‘Redefining Privacy and Anonymity through Social Networking Affordances’.

<sup>54</sup> S.K. Maity, A. Panigrahi and A. Mukherjee, ‘Book Reading Behavior on Goodreads Can Predict the Amazon Best Sellers’, *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (2017).

<sup>55</sup> Thelwall, and Kayvan, ‘Goodreads: A Social Network Site for Book Readers’, p. 972.



contemporary book culture in order to boost sales and enhance a performative consumption element among readers. This process ultimately feeds into an ‘algorithmic culture’ within which ‘computational processes of sorting, classifying and hierarchizing of people, places, objects and ideas on sites like Amazon, have profoundly altered the way “culture”, as a category of experience, is now practiced, experienced and understood’,<sup>56</sup> in this case referring to book culture specifically. In short, Amazon and, to a certain extent, Goodreads users, seem to find themselves in the vicious circle of advertising works through their reviews and user-generated digital traces, a process which essentially leads them to consume more books and therefore, for the corporate giant to ultimately make a larger profit. The difference, however, between the approach of the two platforms seems to rest in the fact that Goodreads does not exclusively lead reviewers to a buying decision but does give them the option to do so (i.e. through the ‘buy on Amazon’ tab) while ultimately merely fostering discourse about books. Of course, it is safe to assume that there is a ‘metaphorical commodity value’<sup>57</sup> to these reviews as they also function as ‘reputational markers’<sup>58</sup> for a book and therefore as ‘purchaser advice’.<sup>59</sup> On the other hand, Amazon, as predominantly an online retailing platform, does exclusively foster consumer behavior among users through their reviews. This approach, established here on a more theoretical level, will be supported on the following segment by a data analysis conducted on the basis of a comparison between Amazon and Goodreads reviews in order to identify any platform-specific behaviors by the respective users.

### ***Data Analysis Between Goodreads and Amazon Book Reviews: Identifying Platform-specific Behaviors***

In order to pinpoint the aforementioned bidirectional relationship between the bookish platform and its users, as well as any platform-specific behaviors exhibited by users of both Goodreads and Amazon, a data analysis will be carried out using the reviews on all English titles of The New York Times weekly bestseller lists from 2020 as samples. This qualitative segment of the present paper seeks to further elaborate on the concept of datafication within the contemporary book culture and demonstrate how readers interact with a book within the context of a social reading platform. Prior

---

<sup>56</sup> Andrejevic, Hearn, and Kennedy, ‘Cultural Studies of Data Mining: Introduction’, p. 385.

<sup>57</sup> Murray, ‘Secret Agents’, p. 980.

<sup>58</sup> Murray, ‘Secret Agents’, p. 980.

<sup>59</sup> Murray, ‘Secret Agents’, p. 980.

to presenting the results of the data analysis, it is crucial to acknowledge that while Goodreads reviews were easily accessible via the use of the API tool, Amazon reviews proved to be harder to export due to restrictive download limits per day imposed by the company. Hence, a publicly accessible set of Amazon reviews from the HuggingFace repository was utilized. Despite the fact that this data set is relatively incomplete, with a limited number of reviews per book as well as per unique users, it does arguably provide the opportunity to examine distinct characteristics and user behaviors exhibited in Goodreads and Amazon reviews, respectively.

### ***Data***

With the valuable help of the supervisor of the present research, Dr. Peter Verhaar, the Goodreads and Amazon platforms were crawled for reviews on titles featured in the weekly New York Times bestseller lists of 2020. The selection of this specific sample was deliberate in order to avoid any genre-specific limitations when evaluating the reviews on both platforms. Initially, data was retrieved and downloaded for the weekly New York Times bestseller lists, resulting in a total of 51 lists, corresponding to approximately one list per week, and encompassing a total of 182 unique titles. The first step involved using the Goodreads API tool in order to retrieve the reviews from the bookish platform. Secondly, in order to retrieve the reviews for the respective titles on Amazon, certain challenges arose, necessitating the exploration of various strategies in order to circumvent the anti-webscraping measures taken by the Amazon website. Due to the fact that none of these strategies proved to be efficient in this case, coupled with the restrictive number of downloads on their website per day, a publically accessible set of Amazon reviews on the repository HuggingFace<sup>60</sup> was used. This data set was provided by the McAuley Lab and despite the fact that it is rather incomplete, having a limited number of reviews per book as well as a limited number of reviews per user, the data still does facilitate the examination of the notable differences between the data available on the Amazon and Goodreads platforms. Various analyses were carried out in order to highlight these distinctions; the parameters we focused on were average review length

---

<sup>60</sup> McAuley Lab, 'Amazon Reviews 2023', 2024, HuggingFace, <<https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023/tree/main>> [accessed 22 June 2024].

(measured in words, sentence length and number of sentences), average rating, average positive sentiment score and average negative sentiment score. On the basis of lexical analysis, the most frequent, less frequent and distinct words in reviews of each platform were examined while sentiment analysis involved identifying the amount of positive and negative vocabulary in the reviews of each platform. Finally, the calculation of the type-token ratio was also observed.

## ***Results***

### *i. Average Review Length*

In regards to average review length on each platform, upon combining all of the ISBNs included in the weekly New York Times bestseller lists from both the Goodreads and Amazon datasets, a roughly similar length of reviews was established in the two platforms. As seen in the scatter plot visualization (figure 3), there is a cluster showcasing that most reviews in both platforms revolve around shorter reviews while there appear to be some outliers, particularly in the case of Goodreads, indicating the existence of some longer reviews as well. In addition to that, another parameter that was studied was the number of words included in the reviews of the two platforms. Interestingly enough, there appeared to be a large difference between the words included in Goodreads and Amazon reviews. More specifically, the average number of words in a Goodreads review was 257.36 words while the respective number of words in Amazon book reviews was 85.71 (see figure 4). Moreover, the sentence length of the reviews found on each platform was also analyzed. The result of this analysis showed that there was a difference in the sentence length with Goodreads having a sentence length of 20.69 words while the one for Amazon was of 16.36 words (figure 5). Lastly, there was also a difference in the parameter of the average number of sentences found in reviews: on Goodreads, the average number of sentences was 12.28 while on Amazon it was fairly lower, merely 4.87 (figure 6).

### *ii. Average Star Rating*

Another vital parameter studied during the data analysis was the star ratings. Both the Goodreads and the Amazon platform operate under a five-star rating system without the option of half stars. Upon conducting an analysis on both platforms, the results showed that the average star rating for the Goodreads platform was 4.02, while for the Amazon platform it was 4.13. Despite the fact that the two averages are quite similar,

the most striking differences were observed in the graphs that displayed the quantity of reviews for each rating. Interestingly enough, the Goodreads ratings graph (figure 7) indicated that the ratings were distributed roughly from 2.7 to 4.6 stars, without any occurrences of one or five-stars ratings. On the other hand, in the graph depicting the Amazon reviews (figure 8), users seemed to have given both one and five-stars ratings. The Amazon chart suggested that users were more inclined to give higher ratings to a title, as the most star ratings are distributed between the four and five-star ratings. Conversely, the Goodreads chart indicated a tendency towards ratings around four stars. Therefore, despite the close average star ratings for both platforms, the ratings on the Amazon platform seemed to be more evenly distributed across the scale compared to those on Goodreads. This discrepancy could indicate that Amazon users are more prone to expressing both their dissatisfaction with a book, as in the case of the one-star ratings, as well as their utter satisfaction, as in the case of the five-star ratings. Conversely, Goodreads users tend to express their opinions in a more moderate manner, by selecting ratings between three and four stars without opting for the extremes.

### *iii. Average Sentiment Score*

Both positive and negative sentiment scores were studied in the reviews gathered from the Goodreads and Amazon platforms with the aim of determining whether the aforementioned distinctions between the two platforms corresponded to respective differences in the emotional language used by reviewers. Firstly, the positive sentiment score was analyzed in the reviews by conducting an analysis on the total sum of the positive words found in reviews of Goodreads and Amazon respectively. It is worth noting that the comparison of the positive sentiment on the two platforms did not reveal any significant differences, as the average positive sentiment score on Goodreads was 0.05, with a very close 0.07 for Amazon. In general, even though both platforms showcased similar results, Amazon did exhibit a wider range of positive sentiment scores, spanning from 0.03 to 0.10, whereas Goodreads reviews mostly ranged from 0.04 to 0.07. This subtle disparity between the reviews suggests that Amazon reviewers are more inclined to use positively-charged vocabulary when discussing a title in comparison to Goodreads reviewers, a fact that aligns with the previous observation that Amazon users tend to assign higher star ratings and

indicates a generally more resilient and positive Amazon audience. Nonetheless, this variance is minor and should not be broadly generalized given that the average positive sentiment scores of the reviews across both platforms remain similar.

Negative sentiment scores between Goodreads and Amazon were also examined in the analysis. The average negative sentiment score on Goodreads was found to be 0.0293, with Amazon at a very close 0.0244. Consequently, the findings for the two platforms are nearly identical. Nevertheless, some disparities in the score range for each platform were observed in this case as well. More specifically, in this instance Amazon again showcased a broader spectrum of scores, ranging from 0.005 to 0.045. In contrast to that, Goodreads reviews maintained a consistent range between 0.015 and 0.045. These results further reinforce the notion that Amazon reviewers tend to use more emotionally-charged language, whether conveying positive or negative sentiments, while Goodreads reviewers adopt a more moderate approach towards the expression of emotions.

#### *iv. Platform-Specific Linguistic Analysis of Goodreads and Amazon Reviews*

The distinctive language used in the reviews on both platforms was also analyzed through a textual analysis of the reviews in an attempt to gain insight into the typical language patterns employed on each platform and potentially identify trends in user language for the two platforms. More specifically, the Dunning-Log likelihood statistical approach was employed in order to pinpoint the distinctive terms for each platform. This method is used to assess the distinctiveness of words in a particular set of texts in relation to the texts in a reference corpus, by calculating probabilities based on word frequencies. This way, using the frequencies already calculated for the reviews on Goodreads and Amazon, distinctive words for each platform can be distinguished. Figures 9 and 10 illustrate the distinctive words found in the Goodreads and Amazon reviews, respectively. In terms of the Goodreads-specific lexicon, frequently mentioned terms like 'spoiler', 'lucas', 'decker', 'davenport', 'ryan' and 'mary' highlight that the Goodreads reviews strongly emphasize the narrative, characters and the overall reading experience of the book that is being reviewed. This suggests a highly detailed and analytical approach from the part of the Goodreads users, focusing not only on the commercial value of the title, but also in its thematic content, thus delving into critical analysis. In addition to that, the term 'netgalley',

alluding to the homonymous online service that offers pre-released titles for book enthusiasts to read and review, indicates an advanced level of awareness on the book market among Goodreads users. Lastly, terms like ‘war’, ‘mother’, ‘women’ and ‘death’ indicate a tendency of the reviewers on the platform to engage in highly intellectual discussions on themes and sociopolitical topics, rather using their reviews as a mere indicator of the quality of a title.

Respectively, the distinct terms observed on Amazon seemed to carry different connotation for the users. Within the Amazon-specific review vocabulary, positive terms such as ‘great’, ‘good’, ‘excellent’ and ‘interesting’ were rather popular, suggesting that the Amazon reviewers were more prone to having a direct approach in providing feedback on the corresponding title. More specifically, Amazon reviewers opted for common descriptive terms to express their opinions of the respective work rather than delving in a detailed analysis as in the case of Goodreads reviewers. These terms could also indicate a general tendency towards offering a more surface-level evaluation of the respective book, possibly hinting at a decision-making approach centered around making purchasing decisions. Moreover, the presence of words such as ‘gift’, ‘bought’ and ‘condition’ points towards a purchasing context that is absent in Goodreads reviews and which further reinforces the aforementioned argument of a review process driven by decision-making implications.

In an endeavor to examine the vocabulary used in the two platforms and detect any platform-specific linguistic features, data analysis was conducted in order to ascertain the type-token ratio in the reviews. More specifically, upon tokenizing the text from the reviews on both platforms, the visualization of this data (see figure 11) revealed a slight disparity indicating that Amazon reviews exhibited a higher type-token ratio, this way showcasing a more diverse vocabulary compared to the reviews found on Goodreads. This was a particularly intriguing discovery given that the aforementioned data suggested a preference of Goodreads users for more in-depth analyses and a more varied vocabulary would be expected because of this tendency. It is important to note, however, that the difference between the two ratios was minimal; slightly above 0.3 for Goodreads and slightly above 0.5 for Amazon reviews. These results highlight that Amazon reviews demonstrate a slightly higher lexical diversity than those on Goodreads, with users displaying a propensity for employing a broader range of vocabulary when composing their reviews, a fact that could potentially signify the existence of more intricate and nuanced reviews on the platform. This

finding could also be attributed to the distinct orientations of the two platforms; the purchasing-focused nature of Amazon is a platform-specific circumstance which is absent on Goodreads. In addition to that, the fact that Amazon reviews also offer content-based commentary of the titles could contribute to the platform possessing a broader vocabulary that encompasses both the commercial and the content-focused contexts. On the other hand, a lower type-token ratio on Goodreads could imply a less lexically diverse audience, resulting in more repetitive and straightforward reviews on the specific platform as a whole. Nevertheless, given that the difference between the two ratios is only marginal (0.2 points), caution should be taken in overgeneralizing and stressing these observations, which should remain within the context of the specific sample sets used in this research.

### ***Key Findings***

*i. Goodreads users exhibit higher levels of engagement and produce more intellectually sophisticated reviews compared to Amazon users.* According to the findings, both in terms of the length of the reviews as well as the words within them, Goodreads reviews do seem to be more advanced and aligned with the current trends in the contemporary book realm than Amazon ones, as exemplified by the extensive reference to the Netgalley service.

*ii. Goodreads users refrain from expressing extreme opinions within the context of their reviews.* Users of the platform demonstrate an inclination towards adopting a more moderate position, both when assigning star ratings and when composing their reviews. They also tend to refrain from expressing extreme emotions, either positive or negative, thus avoiding the use of emotionally charged language in their reviews.

*iii. Amazon reviewer base appears to exhibit a greater inclination towards articulating extreme emotions in their reviews.* The findings indicate a discernible trend of greater emotional expression among user when conveying both positive and negative sentiments, thereby characterizing Amazon reviewer base as more emotionally expressive and direct.

*iv. Goodreads users tend to compose more content-focused reviews.* More specifically, users of the platform, being highly engaged in the task of review writing,

have a propensity to provide more extensive and in-depth critical analyses of the respective titles, with these reviews predominantly emphasizing the narrative and the characters in the books.

*v. Amazon users tend to compose reviews that are geared towards purchase-making.*

Due to the fact that the reviews on the platform are generally more straightforward and concise, they seem to suggest an underlying decision-making approach for the readers. While they do involve some plot analysis on the respective titles, their short and direct nature as well as the frequent use of terms related to purchasing collectively serve as a form of quality assessment for the titles. The higher lexical diversity that was observed in these reviews could potentially be attributed to the dual focus on both the content and the commercial aspect of a title, a characteristic which is specific to Amazon.

### ***Discussion***

The prevalent observation that arises from the findings of this present study indicates that the Goodreads and Amazon platforms have different foci, a fact that can also be evident through the analysis on the users' reviews. On the one hand, Goodreads appears to be an inherently bookish platform, where users craft detailed and thorough reviews in a journalistic manner, refraining from expressing either extremely positive or negative sentiments. The emphasis appears to be on generating discourse around the themes and the characters of a book in an attempt to engage with fellow readers, therefore establishing the 'social reading' nature of the website. On the other hand, the Amazon platform appears to focus more on assisting readers in making an informed decision into purchasing a book based on the respective reviews on the platform. While Amazon reviews do entail some level of critical analysis, it is kept concise, as the primary focus of the website seems to rest in helping readers make purchase-related decisions. Consequently, the reviews seem to mirror the overall commerce-oriented nature of the Amazon platform, therefore reinforcing the concept of a bidirectional relationship between bookish platforms and their users.

The present study was partially based on the research conducted by Dimitrov et al. in 2021.<sup>61</sup> Concerning the findings, both similarities as well as differences have

---

<sup>61</sup> Dimitrov and others, 'Goodreads vs Amazon: The Effect of Decoupling Book Reviewing and Book Selling'.



been identified between the two studies. In regard to the similarities, the key findings about Goodreads having a ‘more engaged reviewer base’,<sup>62</sup> Amazon users exhibiting ‘more purchase-oriented reviews’<sup>63</sup> and the Amazon platform producing ‘more extreme valued reviews’<sup>64</sup> are similar in both studies. The primary distinction lies in the final key finding of the Dimitrov et al. research, which indicates that ‘sentiment is stable across platforms’.<sup>65</sup> In the current study, the sentiment values were also examined, with the results showing that despite similar average sentiment scores (specifically, in terms of positive sentiment, 0.05 for Goodreads and 0.07 for Amazon and in terms of negative sentiment, 0.0293 for Goodreads and 0.0244 for Amazon), the broader range of sentiment scores for Amazon suggests a tendency for expressing both extremely positive as well as negative sentiments in reviews on the platform. On the other hand, the narrower scope of the Goodreads sentiment scores indicates that users generally opt for more moderate and less emotionally-charged language in their reviews. This curious<sup>66</sup> finding of the Dimitrov et al. study about sentiment stability between the two platforms could potentially be attributed to the focus of the study on reviews of books categorized under the ‘biography’ genre on both platforms, a genre that may not necessitate the use of emotional language by reviewers. However, the present study, which did not focus on a specific genre but on weekly New York Times bestseller lists of the year 2020, demonstrated that despite the generally similar sentiment averages, sentiment across platform was in fact not stable.

While the present research does offer a broad overview of some general trends detected in Goodreads and Amazon reviews, opportunities for future research do exist in order to help better establish and contextualize the aforementioned findings. To begin with, conducting a similar study using different datasets, such as more recent best seller lists or even samples across multiple years, could enhance the contextual understanding of the findings of this study and potentially identify specific patterns in review writing as well. The peculiarity of Amazon reviews utilizing a more extensive lexicon compared to Goodreads reviews should also be further studied. Finally, research on sentiment expressions that transcend the binary distinction between positive and negative emotions in reviews on both platforms could yield valuable

---

<sup>62</sup> Dimitrov and others, p. 3.

<sup>63</sup> Dimitrov and others, p. 3.

<sup>64</sup> Dimitrov and others, p. 4.

<sup>65</sup> Dimitrov and others, p. 4.

<sup>66</sup> Dimitrov and others, p. 4.

insight into the nuanced feelings that users experience while composing their reviews and the potential variation of these feelings across platforms.

### Chapter III

#### ***The Shifting Notions of Identity, Community and Ownership Reflected Through the Datafication of the Contemporary Book Culture***

To begin with, through both the qualitative and quantitative segments of the present research a great shift in the approach that bookish platforms such as Goodreads and Amazon have brought towards the contemporary book culture has been established. Correspondingly, a similar shift can also be observed among the audience that engages with such platforms. Today, the concepts of Identity, Community and Ownership within the bookish realm have shifted *for* as well as *by* the audience of book culture itself. Firstly, in terms of the notion of Identity, this concept has become increasingly datafied within the contemporary book culture, as modern social networks seek to essentially endorse digital embodiments of the users' three-dimensional off-line selves. The data involved in algorithmic operations related to the identification of users online does not interpret 'symbols to be understood and differentiated [by entities such as Goodreads or Amazon] but inputs to be sorted [and categorized],<sup>67</sup> as in the case of the aforementioned data analysis. In the case of the notion of Identity on Goodreads, categorization of data is vital; 'what the categories are, what belongs in a category, and who decides how to implement these categories in practice, are all powerful assertions about how things are and are supposed to be'<sup>68</sup> in the bookish digital realm. In the digital sphere, it is also common for individuals to forsake the idea of the private self in favor of social collectivity. More specifically, users on Goodreads do not have to use their real names on the platform; most of the time they opt for abbreviated versions of it or nicknames instead. This fact, coupled with their augmented digital traceability on these platforms, essentially leads users to associate themselves, and therefore their individuality, with their virtual library and their social performances on the website as a representation of their interests, experience and ultimately their identity.

---

<sup>67</sup> Andrejevic, Hearn, and Kennedy, p. 384.

<sup>68</sup> Gillespie, 'The Relevance of Algorithms', p. 171.

This finite notion of identity on the Goodreads platform, based on the idea of approaching one's tastes and preferences, particularly in the case of books, as a representation of '[their] innermost selves and most personal experiences'<sup>69</sup> can also be regarded as lost when considering the startling accuracy of the Goodreads algorithm in connecting users with similar interests and generating personalized book recommendations for them. In this sense, the uniqueness and individuality of curating one's own personal library is somewhat lost on bookish platforms such as Goodreads, where users can access a plethora of book suggestions based on a single title that they found interesting with the click of a button and within which 'digital affordances can only be programmed to represent a user with a finite number of "adjectives" and "verbs"'.<sup>70</sup> In essence, within such bookish algorithmically-driven websites a paradox occurs according to which 'the self takes the form of a decentered (or centerless) network of codes that, on another level, also serves as a node within another centerless network'.<sup>71</sup> In this context, the user's digital identity resembles an avatar or a digital persona, and can be defined as a process of 'costumization'<sup>72</sup> of a profile that is intended to be broadcasted online while confined within the boundaries of the 'Reader' image. This limitation of the individual to their literary preferences on online bookish websites appears to be a phenomenon that is exclusive to the online domain despite the omnipresent 'rhetoric of selfhood'<sup>73</sup> that these website claim to foster. More specifically, in real-world scenarios, one is not solely defined or perceived based on their interest in one particular thing, in this case books and reading. Rather, such interests are generally regarded as complementary to one's holistic personality or identity. This seems to be the sense of the individual that Goodreads co-founder Otis Chandler sought to approach in his motivation to establish a virtual space for individuals to connect and share book recommendations. In addition to that, the performative nature of engaging with social media, or in this case with a social reading website, coupled with the limited amount of information on the individual that this entails, may result in certain users fabricating an untrue identity online. This has been an issue on Goodreads, exemplified in the aforementioned incident of author Cait Corrain creating multiple fake Goodreads profiles to sabotage her colleagues'

---

<sup>69</sup> Pinder, 'ONLINE LITERARY COMMUNITIES: A Case Study of LibraryThing', p. 83.

<sup>70</sup> Cirucci, 'Redefining Privacy and Anonymity through Social Networking Affordances'.

<sup>71</sup> Pinder, 'ONLINE LITERARY COMMUNITIES: A Case Study of LibraryThing', p. 84.

<sup>72</sup> Cirucci, 'Redefining Privacy and Anonymity through Social Networking Affordances'.

<sup>73</sup> Murray, 'Charting the Digital Literary Sphere', p. 325.

publications. Consequently, one could argue that the shift that the datafication of the book culture has brought forth could be regarded as a shift away from Identity as a multifaceted yet fixed concept in real life towards a more fluid, partial, unstable and potentially deceptive notion within the machine-readable online realm of Goodreads.

Along with the notion of Identity, the concept of Community has also undergone a significant shift within the digital book culture realm. To begin with, the fact that Goodreads constitutes a social reading website that places a strong emphasis on cultivating a communal feeling among its users already became evident from the analysis of the very functionalities found on the platform in the first chapter of the research. Such functionalities are essentially designed to establish and nurture relationships among users and to foster ‘properties that define the nature of social relationships, the communal characteristics and the strength of these relational properties’,<sup>74</sup> this way offering the possibility for generating an abundance of ‘book-talk’ online. In today’s digital age, individuals who are interested in reading have a better chance to connect with like-minded readers now more than ever. Through platforms like Goodreads, users can easily initiate conversations on their favorite titles, authors or genres with the click of the button and build relationships while transcending previously vital boundaries such as time and space. The online realm also allows for immediate interaction through multimedia content such as hyperlinking, features that were missing prior to the emergence of social websites, thus enhancing the overall user experience. Furthermore, these connections do not have to be limited solely to one’s already existing circle in real life but can further expand to ‘a very large audience through the friends of friends’ network’,<sup>75</sup> essentially including circles of connection that extend beyond the users’ time or space. In a sense, one could also argue that social reading websites such as Goodreads even encourage engagement based on shared interests over already established social connections outside of the platform.

This preference of engaging with people who particularly hold similar interests constitutes a key characteristic of the online approach to building relationships and is reflected in the concept of ‘affinity spaces’, described by Vlieghe, Muls and Rutten as ‘physical, visual or blended spaces that are often spread across

---

<sup>74</sup> Kapoor, Jayasimha and Sath, p. 50.

<sup>75</sup> Kapoor, Jayasimha and Sath, p. 48.

many sites, such as face-to-face meetings, message boards, blogs and web-pages and offer multiple interest-driven trajectories, opportunities to learn with others and paths towards becoming an authentic participant'.<sup>76</sup> The feeling of belonging to a group of like-minded individuals, in this case readers, is also of fundamental importance on the general Internet ecosystem and fosters the continuation of social media platforms with many users not focusing as much on the reading experience or the cultivation of their literary taste, but on engaging in the online book-related discourse. However, this grouping of individuals based on their interests arguably constitutes a deliberate strategy employed by social websites such as Goodreads in order to collect metadata. This data is crucial for enhancing the functionality and precision of algorithmic processes within the bookish website. More specifically, on Goodreads the users' preferences seem to be mined not solely to understand and interpret information about the individuals themselves, but to arrange and sort them as well as their interactions according to the priorities set by the bookish platform's algorithmic priorities.<sup>77</sup> This approach further serves to reinforce the evolution of the dynamics surrounding the notion of community as well as the ways in which it is practiced, experienced and perceived within the contemporary digital realm of books.

Lastly, in conjunction with the notions of Identity and Community, another essential component of book culture, Ownership, has also undergone a significant transformation during this era of increasing datafication. To begin with, because of the intangible and fluid nature of the digital, coupled with the fact that the modern book culture predominantly occurs online, books are also mostly interacted with online rather than in their physical, tangible and stable form in real life. Therefore, according to Albrechtslund, despite the fact that

books are commonly associated with some form of private possession [as well as with] the ability to shelve or pass [them] on to others, [this notion of ownership and possession has been] obviously tested by the increased entanglement of reading with networked technologies and devices.<sup>78</sup>

---

<sup>76</sup> Vlieghe, Jaël and Rutten. 'Everybody Reads: Reader Engagement with Literature in Social Media Environments', p. 27.

<sup>77</sup> Andrejevic, Hearn, and Kennedy, 'Cultural Studies of Data Mining: Introduction', p. 381.

<sup>78</sup> Albrechtslund, 'Amazon, Kindle, and Goodreads: Implications for Literary Consumption in the Digital Age', p. 554.

Of course, some readers do still prefer the physical copy of a book as their primary mode of reading, but alternative book formats such as e-books and audio books have become increasingly popular in recent years, this way rendering the act of owning or collecting books a rather secondary motive for modern readers. This can be regarded as a great shift from the traditional archetype of the Reader, who would be conventionally associated with a book collector. The emergence of book subscription services such as Kindle and Audible (interestingly enough, both owned by Amazon) further complicates the concept of book ownership in the modern book culture. Hence, one could posit that the notion of ownership is no longer inherently linked to being a reader in the digital realm. Particularly on platforms like Goodreads, users establish their Reader identity simply by engaging in discourse around books and exchanging their ideas with fellow users, with the concept of owning or collecting books regarded as a secondary practice. At the same time, this great shift in the concept of ownership from a vital characteristic of book enthusiasts to a secondary trait in the online book culture could also be regarded as a part of the greater shift from the traditional perception of the act of reading as an individualistic act towards a more communal and social process established online. Today, reading a book, and by extension owning it, is directly associated with the reader sharing their opinion with other users through star ratings or detailed reviews. Therefore, generating discourse around a title on Goodreads appears to hold more importance than accumulating an extensive personal library. This approach can be explained by the inherently community-inclined nature of the Goodreads platform. It is important to note, however, that on other platforms, particularly social networking sites like Instagram and TikTok, users displaying their extensive personal libraries or their new book purchases serves as a token for establishing their Reader status. Nevertheless, even in this case, the ultimate goal of this act does not rest in individual enjoyment but rather in the fostering of community connections with other book enthusiasts.

### ***Data-Generated Algorithms of Book-Centric Platforms: Is Privacy at Stake?***

The datafication of the contemporary book culture has also resulted in changes to the notion of privacy experienced by readers in the context of bookish platforms. The sense of privacy can be defined as the right to control personal information as well as

the agency to choose when this right to privacy should be enacted,<sup>79</sup> rights that can generally be regarded as particularly difficult to acquire in online spaces. The current datafied state of social media has effectively managed to ‘blur and complicate the levels of privacy and sociality’<sup>80</sup> within all practices that occur within them. A prime example of this shift would be the complete digital traceability of user generated data on these platforms. More specifically, the digital footprint of each user/reader can be traced by the Goodreads and Amazon websites through the utilization of cookies, which are oftentimes also exploited by these corporations for their benefits. This data is then used to feed into the complex algorithms of the platforms, the operations of which remain a black box for the users as well as for the general public, as their ways of functioning remain untraceable. User data as well as their personal information registered online can also be shared by for-profit bookish websites with third parties for advertising purposes, as in the case of publishing houses or its owner Amazon in an attempt to monitor the predominant trends and preferences in the contemporary book culture. The correlation between the two corporations is explicitly stated in the Goodreads ‘Privacy Policy’ webpage, where they state that ‘if you have an account on [www.amazon.com](http://www.amazon.com), information gathered by Goodreads may be correlated with any personal information that [Amazon] has and used by Goodreads and Amazon to improve the services we offer’.<sup>81</sup>

Ultimately, all user-generated content on the platform culminates in a process in which users are ‘both collecting and being collected under a new regime of controlled consumerism’,<sup>82</sup> a fact that leads to a blurring of the boundaries between publicness and privacy. In this case, the concept of eWOM (e-word-of-mouth) is a prime example of how the aforementioned user data could potentially serve as indicative of the current trends and preferences within online book communities, arguably vital information for bookish third parties. Apart from the exploitation of user data, privacy issues may also arise from the fact that the act of reading has acquired an increasingly social nature in the public sphere, with platforms such as Goodreads categorized as ‘social reading websites’. Consequently, reading has become a highly commodified and performative act on such platforms, departing

---

<sup>79</sup> Cirucci, ‘Redefining Privacy and Anonymity through Social Networking Affordances’.

<sup>80</sup> Driscoll and Sedo, ‘Faraway, So Close: Seeing the Intimacy in Goodreads Reviews’, p. 248.

<sup>81</sup> Goodreads, ‘Privacy Policy’, <<https://www.goodreads.com/about/privacy>> [accessed 28 May 2024].

<sup>82</sup> Nakamura, “‘Words with Friends’: Socially Networked Reading on *Goodreads*”, p. 241.

from its past perception as an individualist and solitary activity to a more communal and social experience, no longer based on solitude or regarded as an activity confined to the private time of the reader. Finally, the practice of a for-profit openly accessible website like Goodreads sharing user generated data with external parties for commercial advertising purposes establishes reading on such websites as an increasingly ‘algocratic practice’,<sup>83</sup> existing within the complex blurred ground between public and private data. These blurred boundaries between the two notions prompt further questions about the validity of the notion of privacy in the social media realm in the first place, especially upon considering that users who perform in the digital realm of social media have essentially ‘already chosen to disclose a multitude of content to *someone*’.<sup>84</sup>

Coupled with this evolving notion of privacy on platforms like Goodreads and Amazon, self-sovereignty represents another shifting concept within the social media realm. The concept of self-sovereignty can be defined as the ability of the user to ‘retain control over their personal data and, to a certain extent, over the representation of their identity (or persona) within a particular identity management system’,<sup>85</sup> thereby asserting control over the knowledge that they wish to acquire, their personal data and their digital identity overall and ‘preserving the right to selective disclosure of different aspects and components of one’s identity in different contexts’.<sup>86</sup> Consequently, an additional crucial aspect of this notion could arguably be user autonomy. As has already been established, social media websites, as well as social reading ones like Goodreads in this case, do not afford users complete autonomy over their data or activity on the platform, a fact that prompts inquiries about the extent to which users have autonomy in freely shaping their online identities rather than conforming to the algorithmic processes of the platform. A prime example of the insufficient level of autonomy and control that Goodreads users possess over their personal data is the fact that, according to the Privacy policy webpage, ‘user content’, which is described as all activity posted by a user, cannot be entirely removed from the platform after posting; ‘copies may remain viewable in cached and archived pages

---

<sup>83</sup> Nakamura, p. 242.

<sup>84</sup> Cirucci, ‘Redefining Privacy and Anonymity through Social Networking Affordances’.

<sup>85</sup> A. Giannopoulou, ‘Digital Identity Infrastructures: A Critical Approach of Self-Sovereign Identity’, *Digital Society*, 2.18 (2023), p. 18.

<sup>86</sup> Giannopoulou, ‘Digital Identity Infrastructures: A Critical Approach of Self-Sovereign Identity’, p. 18.



or if other users have copied and stored your User Content'.<sup>87</sup> This fact alone deprives users of autonomy and renders control over their already-online data impossible.

Various measures should be deliberated for the reinforcement of the idea of self-sovereignty in bookish social media in order to cultivate a more democratic environment that prioritizes the user as the central agent and safeguards their data and personal information within the prevailing data-centric book culture. Firstly, it is imperative to exercise control over the users' personal data registered on platforms such as Goodreads. This could be achieved through more robust privacy regulations as well as the explicit disclosure of the platform in terms of their intentions with the user data prior to registration, including explicitly naming external parties that may potentially have access to this data, such as Amazon in the context of Goodreads. By doing so, users would have full awareness of the ways in which their personal data is utilized by the platform as well as the option to express the requisite consent for such usage. Introducing the concept of data portability could also significantly enhance self-sovereignty on Goodreads. This concept essentially means that users would have the ability to extract their personal information registered on the platform. While Goodreads does currently offer a variation of the data portability concept by enabling users to import or export their bookshelves to and from Goodreads, this limited data portability does not effectively reinforce self-sovereignty but rather focuses on facilitating portability exclusively in terms of the books registered on the platform. Another potentially beneficial approach towards enhancing self-sovereignty could involve the users' ability to browse the platform anonymously. In this case, by operating in a decentralized technological infrastructure, users would only have access to book-tracking related functionalities, a fact that could potentially affect the primarily social and community-building objectives of the platform. Nevertheless, users should undoubtedly have some degree of authority over the information they wish to share on social media platforms as well as over the parties that could potentially have access to this information. This ability, contextualized in a sustainable technological context, would cultivate a stronger sense of privacy and self-sovereignty for users, this way ultimately enhancing a smooth user experience for the increasingly datafied reader on Goodreads.

---

<sup>87</sup> Goodreads, 'Privacy Policy', <<https://www.goodreads.com/about/privacy>> [accessed 30 May 2024].

### *The Notion of Control in the Contemporary Digital Book Culture*

While examining the aforementioned concepts of Identity, Community, Ownership and Privacy, it becomes evident that throughout all of them and their shift in the bookish digital realm, there appears to be an underlying need for renegotiation of the notion of Control as well. More specifically, the whole infrastructure of the contemporary digital book culture, with Goodreads as the central focus of this research, is essentially centered around the idea of reevaluating and potentially redistributing control among the various agents present within this culture, including users/readers, algorithms and corporations. Goodreads users, on their part, are currently faced with the never-ending challenge of first understanding what the highly technical and non-transparent data operations actually entail and subsequently asserting a sense of control over the data that they generate through the bookish platform. Therefore, the new age bibliophile has to take control over their personal data in the online sphere as well as comprehend and gain control over their representation on the platform. In short, the Goodreads user today has to establish this sense of control in order to remain ‘the master of [their] profile instead of a third-party service provider’<sup>88</sup>, and rather than allow for the ones who privately own and control<sup>89</sup> the algorithmic operations of the platform to control the identities embodied within this platform as well.

Apart from this exertion of control, the modern-day reader has to also understand their control in influencing and therefore shaping the bookish realm. As already established through the quantitative segment of this research, the user generated data essentially serves as a primary resource for the identification of emerging trends formed within the book market through reviews, ratings and general bookish discourse, this way serving as a form of quality control for the respective titles. The significance of this data for the broader book realm is further reinforced by the interest of external entities such as Amazon to also gain access to this data, as evidenced by the acquisition of the Goodreads platform in 2013. This notion of control that users attempt to obtain is further complicated by algorithmic processes, which notoriously remain opaque for the users, this way limiting their understanding of the ways in which their data is utilized on platforms such as Goodreads, or which

---

<sup>88</sup> Giannopoulou, ‘Digital Identity Infrastructures: A Critical Approach of Self-Sovereign Identity’, p. 18.

<sup>89</sup> Andrejevic, Hearn, and Kennedy, ‘Cultural Studies of Data Mining: Introduction’, p. 380.

specific external parties have access to this data. According to Microsoft researcher Tarleton Gillespie,

the criteria public information algorithms take into account are myriad; each is fitted with a threshold for what will push something up in the results, position one result above another, and so on. So evaluations performed by algorithms always depend on inscribed assumptions about what matters, and how what matters can be identified.<sup>90</sup>

Consequently, algorithmic operations also prompt questions in regard to who has control over the formatting the Goodreads recommendations and the level of control that users actually have over the choices they make within the bookish platform. On the other hand, one could argue that Goodreads does attempt to provide a sense of control to its users; the existence of recommendations in itself offers a sense of control to users, as the abundance of books available on the platform is filtered and narrowed down to exclusively titles of personal interest specifically catered to each user. Of course, the community and the cultivation and fostering of it, all core aspects of the Goodreads platform, also offers a heightened sense of control to users, allowing them to initiate interactions exclusively based on shared literary interests, an ability that is provided through the recommendation of user profiles as well as through functionalities such as ‘groups’ and ‘discussions’.

Moreover, interestingly enough, the use of cookies and other monitoring mechanisms is generally portrayed by the platform as an inherently beneficial process for users. Surprisingly, within the Privacy Policy section of the website, cookies are described as a means to ‘honor [the user’s] preferences’, ‘provide [them] with personalized content’, ‘recognize [them] as a member’, ‘provide customized features and services’ and overall ‘improve security’<sup>91</sup> on the platform. Despite the fact that there is an option to block cookies on Goodreads, the ‘Cookies Notice’ webpage suggests that by doing so, the user will potentially cease to have access to ‘some features and services’<sup>92</sup> and will also have to manually readjust the cookies

---

<sup>90</sup> Gillespie, ‘The Relevance of Algorithms’, p. 177.

<sup>91</sup> Goodreads, ‘Privacy Policy’, <<https://www.goodreads.com/about/privacy>> [accessed 30 May 2024].

<sup>92</sup> Goodreads, ‘Cookies Notice’, <[https://www.goodreads.com/about/cookies\\_notice](https://www.goodreads.com/about/cookies_notice)> [accessed 30 May 2024].

preferences every time they visit the website, overall rendering the process of browsing the platform less efficient for the users who do not accept the default cookies. Furthermore, as already stated, the relationship between the users and the platform's algorithm can be viewed as bidirectional, with users shaping the platform through their data while the platform itself, through the functionalities available, also shapes user behavior and preferences. This two-way relationship can also become evident through the data analysis findings that have been established in chapter II, which essentially suggest that different types of bookish platforms may attract different approaches from users, leading to different types of discourse within each platform, even when discussing identical subject matters. In this sense, there is also a shifting notion of control in the extent to which the users can influence the platform, as well as in the extent to which the functionalities of Goodreads (or Amazon) can shape the activities and preferences of users.

Algorithms may not be autonomous agents in the digital bookish culture, yet corporations like Goodreads and Amazon arguably use them as tools to exert a sense of control over the data generated on bookish websites. The algorithmic operations present on these platforms appear to wield influence over the content produced on them; fueled by user-generated data, various algorithms are employed in order to shape the social reading platforms as we know them today. Tailored algorithms such as collaborative filtering ones are designed specifically for the purpose of creating book recommendations, user experience-enhancing browser 'cookies', machine learning models are used for identifying behavioral patterns among readers etc. Despite the fact that algorithms are primarily employed to establish a digital environment that is personalized to each individual user, thus granting a sense of control to the platform's users, they also have a significant impact on formulating the content as well as the user profiles with which users interact on the platform, and therefore hold significant control over how the platform is shaped. In addition to that, even though algorithms do not possess agency themselves, they are regulated and supervised by major corporations such as Amazon. This oversight renders these corporations as primary authoritative agents in the digital bookish realm, as they essentially hold considerable control over both the algorithmically-driven platforms as well as over the users who engage with these platforms. The key findings of the second chapter of this research can be regarded as a prime example of the bidirectional relationship between platforms such as Goodreads and Amazon and their

users in shaping the predominant ‘book talk’ and therefore the modern book culture as a whole. It is evident that the objectives of a platform, and by extension, of a corporation, are reflected through the activity and the ways in which the users engage within the particular platform. Fundamentally, all sociocultural discourse generated around the phenomenon of the datafication of the contemporary book culture appears to predominantly center around the concept of control and the power dynamics that are (still being) established among the agents of this digital era.

### ***Conclusion***

In the present study, both qualitative as well as quantitative methods were employed, revealing the process of datafication within the contemporary book culture. It can be argued that the act of reading today has evolved into a hybrid practice, enhanced by bookish digital platforms. Goodreads stands out as the predominant social reading website to this day, while its parent company, Amazon, reigns as the most popular online book retailing platform. However, the relationship between readers and the bookish websites is by no means unidirectional. Through close examination of the Goodreads functionalities and their impact on users as well as data analysis conducted on the Goodreads and Amazon platforms, it was established that both websites fostered a reciprocal relationship with their audience. Within this connection, user-generated data is processed by the platforms’ algorithms in order to identify trends and patterns in user behavior that could potentially influence the book market. Consequently, the datafication of the contemporary book culture fundamentally creates a dynamic relationship between bookish platforms and their users, according to which one perpetually influences the other.

In order to comprehend this bidirectional relationship, the architecture of the Goodreads website was closely observed, focusing on the available functionalities and their implications as reflective of the users’ behaviors and overall activity. This detailed analysis led to the categorization of the Goodreads features based on their purpose on the platform. More specifically, the predominant types of features on the bookish platform were identified as book-tracking, community-building and Amazon-related features. Each category was closely observed in terms of the number of features included, their frequency and positioning on the website, the vocabulary used to describe them, and their overall integration within the structure of the website from the perspective of a Goodreads user. Some noteworthy findings emanating from this

observation include the prevalence of community-building functionalities, with the platform encouraging users to rate, review, and engage in discourse through group discussions in order to reach a wider audience. This ultimately highlights the platforms' emphasis on 'social reading' while also reflecting the broader cultural trend of cultivating and nurturing social connections within social media networks. In addition to that, an intriguing aspect is the limited presence of book-tracking functionalities, despite their foundational role in the platform's inception in 2007. The Amazon-related features were also quite interesting to observe, as they remain present but rather subtle for the users and they overall add a commercial dimension to the Goodreads website, not present before its acquisition by the Big Five corporation. Through this addition of Amazon-related functionalities, acts such as composing reviews are rendered as performative and can potentially influence readers' purchasing decisions.

The implications of these functionalities were subsequently discussed, revealing the ways in which they serve to showcase the underlying motives of the Goodreads platform itself. One such motive appears to be the enhancement of the social and communal feeling among users, ultimately positioning the Goodreads platform akin to a contemporary social media website. In addition to that, another motive seems to be the focus on commercial aspects of reading, established particularly through the Amazon-affiliated features. In essence, the default functionalities on Goodreads indicate a dual emphasis on fostering a literary community as well as promoting a sense of performative consumption among its users. Taking into consideration that Goodreads constitutes the predominant and most widely used bookish media today, these motives, evident through the default functionalities of the platform, reflect the prevailing trends and incentives within the contemporary book culture. The influence that Amazon exerted in shaping this reading culture, particularly in its capacity as the owner of Goodreads, was subsequently analyzed. Upon looking into Amazon as a corporation, a multifaceted role in the book market can be identified. More specifically, Amazon appears to function as a publisher, printer, distributor and bookseller within the contemporary book realm, a fact that effectively establishes a monopoly within the contemporary book market and consequently within the broader literary culture. Despite the fact that both Amazon and Goodreads platforms operate as recommendation-based interfaces, the main difference lies in the fact that the former operates as a recommendation-

based retailer with reviews serving as advertising tools, whereas the later does not overtly connect recommendations or reviews with purchasing decisions. However, it is particularly intriguing to note that Amazon possesses authority over the Goodreads user-generated data, thereby enabling the exploitation of this data for its own benefit. This process of classifying and hierarchizing users based on their activity on bookish platforms ultimately shifts the interpretation of ‘book culture’ in the ways in which its members perceive and engage with it, thus creating a data-driven book culture. Nevertheless, albeit the similarities found between the Goodreads and Amazon websites that underpin this datafied culture, the data analysis conducted on the reviews found on the two platforms revealed significant distinctions in user engagement. Drawing on the study conducted by Dimitrov et al.,<sup>93</sup> the key findings of the data analysis showcased variances in the ways in which users interact with the platform through their reviews. Interestingly enough, these ways appear to align with the respective focus of each platform, thus further substantiating the initial proposition of a reciprocal relationship between bookish platforms and their users.

The last chapter of the present research aimed at examining the evolving notions of identity, community, ownership, privacy and overall control in the equally transitional digital bookish realm. These notions have arguably shifted *for* as well as *by* the bookish culture. In terms of Identity, a shift has been established in the sense that social media platforms aim to promote digital representations of the three-dimensional individuals behind user profiles. Consequently, on such platforms, the self takes the form of a combination of codes and algorithms suitable for the machine-readable realm of Goodreads or Amazon and transcends from its multifaceted yet fixed nature in real life towards a more fluid and uncertain concept on book-centric websites. In addition to that, a shift in the notion of Community can also be observed in terms of the ways it is practiced, experienced and perceived on bookish platforms. Similarly, the concept of Ownership has arguably undergone a shift within the digital bookish realm as it is no longer inherently linked to the modern concept of the Reader. Owning books is regarded as a secondary trait within the online book culture, while the primary emphasis seems to rest in generating discourse around titles and cultivating a communal feeling through this discourse with like-minded readers.

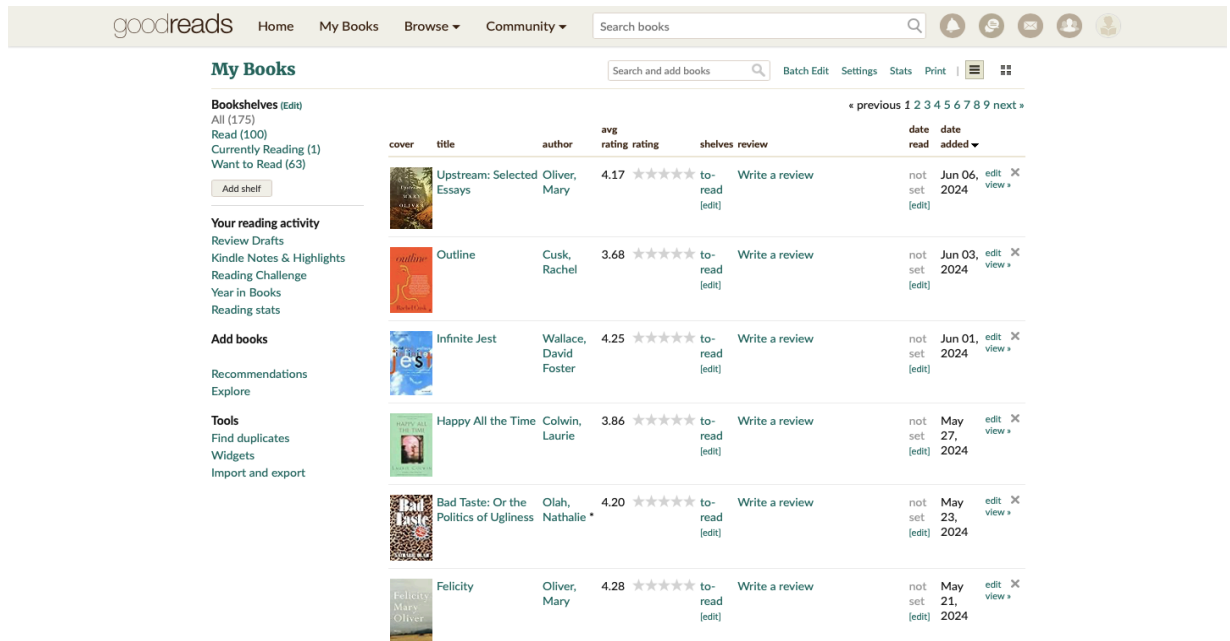
---

<sup>93</sup> Dimitrov and others, ‘Goodreads vs Amazon: The Effect of Decoupling Book Reviewing and Book Selling’.

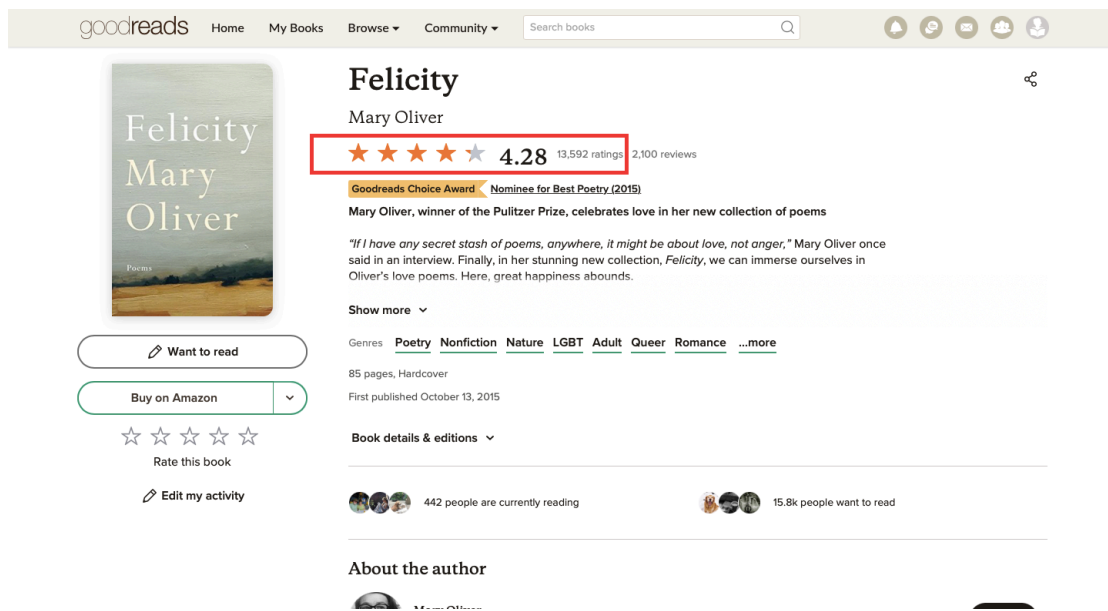
The concept of Privacy has also changed significantly within the online domain, as the act of reading has become a highly performative act on bookish platforms, to an extent that the validity of the notion of privacy needs to be debated and potentially redefined in order to align with the social media paradigm. Alongside privacy, self-sovereignty, denoting user autonomy and authority over their personal data, were also discussed. The examination of privacy-related practices on Goodreads highlighted the need for policy adjustments to reinforce self-sovereignty through actions like privacy regulations, explicit disclosure of terms and conditions, promotion of data portability and anonymous browsing on bookish platforms. Finally, the last notion discussed in this paper was that of Control. Interestingly, all of the aforementioned notions seemed to reveal a shared need for the reassessment and redistribution of control among the agents of the contemporary bookish realm. On the one hand, modern-day bibliophiles have to assert control over their personal data and subsequently of their online representation, while also asserting influence over the bookish landscape. On the other hand, book-centric platforms such as Goodreads and Amazon also hold significant control over user data and the algorithms processing it. Ultimately, this research, encompassing both qualitative and quantitative segments, highlighted a shifting notion of control in the extent to which users can influence a bookish platform, as well as the extent to which the features and overall architecture of a bookish platform can shape the behaviors and preferences of its users. Taking all of these factors into consideration, the datafication of the contemporary bookish culture has arguably led to significant changes impacting all aspects of the literary realm. From shifting the way in which readers engage with books to great changes in the application of key concepts such as identity, community, ownership, privacy and control within it. In addition to that, a prime shift occurs in the relationship between users and book-centric platforms like Goodreads and Amazon. This relationship does not render the user a passive entity or confine them to the conventional 'Reader' stereotype but rather culminates in a two-way relationship where users continually influence the bookish platforms they engage with and vice versa.



## Appendix



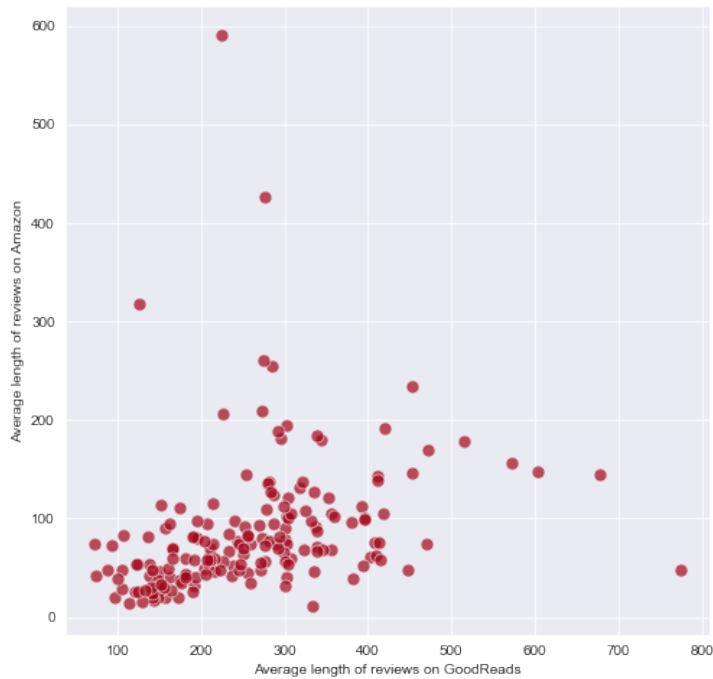
**Figure 1.** An example of the details included in the ‘all’ bookshelf of a Goodreads user profile. Image accessed via the Goodreads website.<sup>94</sup>



**Figure 2.** An instance of the profile of a book on Goodreads, where the star rating is presented as a central part of the description. Image accessed via the Goodreads website.<sup>95</sup>

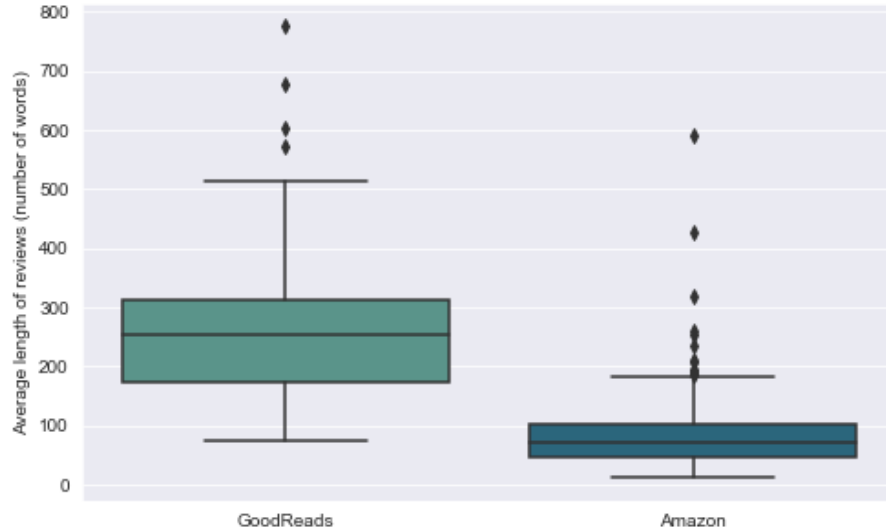
<sup>94</sup> *My Books*, Goodreads, <<https://www.goodreads.com/review/list/143713973?shelf=%23ALL%23>> [accessed June 6 2024].

<sup>95</sup> *Felicity*, Goodreads, <<https://www.goodreads.com/review/list/143713973?shelf=%23ALL%23>> [accessed June 6 2024].



**Figure 3.** Scatter plot visualization showcasing the trends in average review length for the Goodreads and Amazon platforms respectively.

Image accessed via the GitHub repository.<sup>96</sup>

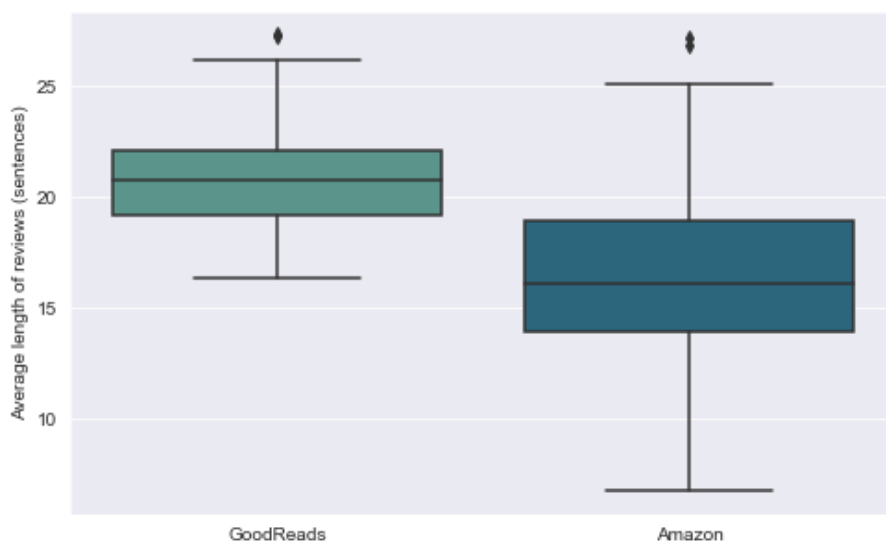


**Figure 4.** Candlestick chart depicting the average words in reviews found on Goodreads and Amazon respectively.

Image accessed via the GitHub repository.<sup>97</sup>

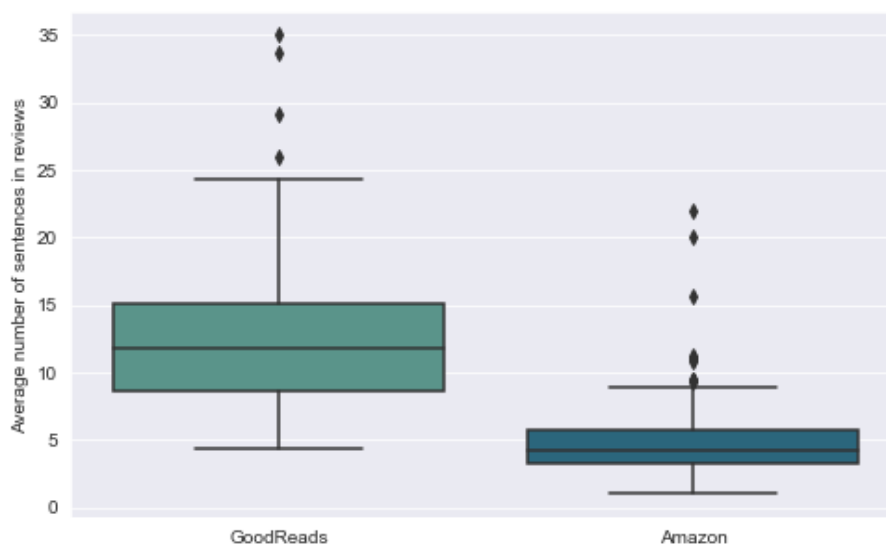
<sup>96</sup> *Analysis\_Reviews.ipynb*, Datafication\_bookculture, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 25 June 2024].

<sup>97</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 26 June 2024].



**Figure 5.** Candlestick chart depicting the average length of sentences in the reviews found on Goodreads and Amazon.

Image accessed via the GitHub repository.<sup>98</sup>

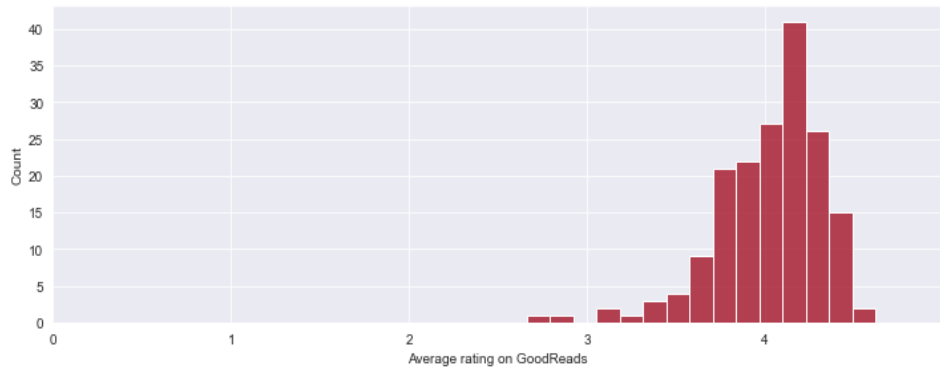


**Figure 6.** Candlestick chart depicting the average number of sentences in the reviews found on Goodreads and Amazon.

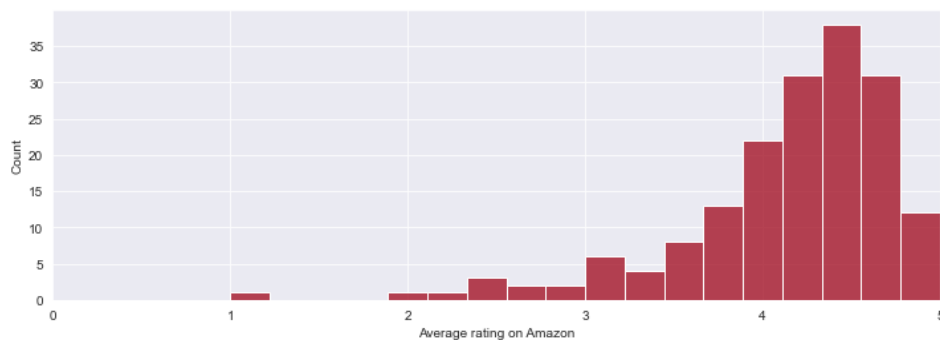
Image accessed via the GitHub repository.<sup>99</sup>

<sup>98</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 26 June 2024].

<sup>99</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 26 June 2024].



**Figure 7.** Bar chart showcasing the average star rating on the Goodreads platform.  
Image accessed via the GitHub repository.<sup>100</sup>



**Figure 8.** Bar chart showcasing the average star rating on the Amazon platform.  
Image accessed via the GitHub repository.<sup>101</sup>

```

br 618.3910112105156
spoiler 150.8533544846391
- 140.37315453049231
hide 105.06773623159091
-- 93.08256246683644
netgalley 90.52936161990037
steel 76.84648240572514
lucas 75.13700184360934
favourite 69.11556198500472
war 67.67992612963971
decker 63.730700279719926
stone 61.96323671461357
mother 58.4405450367571
davenport 55.85860535857253
also 55.811371789458406
women 52.88897468007005
death 51.471126656201605
blog 51.13421779773901
ryan 48.80598884557139
side 46.77369042018657
review 45.64709334985898
mary 43.60647222146341
investigation 43.51023096309083
exchange 41.73861175166631
man 41.66213167357702

```

**Figure 9.** List of distinctive words found in the Goodreads reviews.

<sup>100</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 29 June 2024].

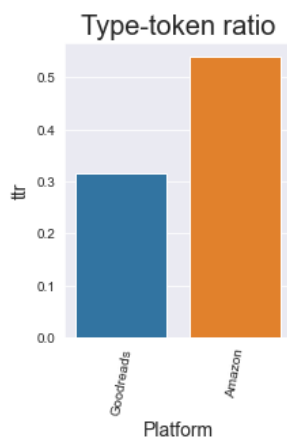
<sup>101</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 29 June 2024].

Image accessed via the GitHub repository.<sup>102</sup>

```
great -2853.6257755751285
book -2235.0765730823387
read -2137.6373704265025
good -1097.0872430601335
put -692.7893737747808
story -628.5111995789043
excellent -527.9213182580925
written -446.9808814532157
gift -384.15989118143136
recommend -375.1009199182497
loved -362.79423550756974
kya -334.4357579895992
marsh -315.0023726903097
author -285.8461491181372
best -276.666393926806
turner -263.42217035004194
grisham -261.3483573049625
wonderful -253.32586963749606
bought -240.6471510977662
enjoyed -238.34452003025058
page -231.1371140025053
interesting -211.90282204864542
reading -198.61348045598277
condition -194.2009601432839
disappointed -193.79956921430073
```

**Figure 10.** List of distinctive words found in the Amazon reviews.

Image accessed via the GitHub repository.<sup>103</sup>



**Figure 11.** Bar chart showcasing the type-token ratio in Goodreads and Amazon reviews respectively.

Image accessed via the GitHub repository.<sup>104</sup>

<sup>102</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 30 June 2024].

<sup>103</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 30 June 2024].

<sup>104</sup> *Analysis\_Reviews.ipynb*, GitHub Repository, Peter Verhaar, <[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 5 July 2024].

## Bibliography

### 1. Primary Sources:

Amazon <<https://www.amazon.co.uk>> [accessed 20 July 2024]

Goodreads <<https://www.goodreads.com>> [accessed 26 July 2024]

Verhaar, Peter, *Analysis\_Reviews.ipynb*, *GitHub Repository*

<[https://github.com/peterverhaar/datafication\\_bookculture/blob/main/Analysis\\_Reviews.ipynb](https://github.com/peterverhaar/datafication_bookculture/blob/main/Analysis_Reviews.ipynb)> [accessed 5 July 2024].

### 2. Published Secondary Material:

Albrechtslund, Anne-Mette Bech, 'Amazon, Kindle, and Goodreads: Implications for Literary Consumption in the Digital Age', *Consumption Markets & Culture*, 23.6 (2020), 553–68 <<https://doi.org/10.1080/10253866.2019.1640216>>

Alter, Alexandra, 'She "Review Bombed" Other Writers. Then Her Book Got Pulled.', *The New York Times*, 12 December 2023  
<<https://www.nytimes.com/2023/12/12/books/cait-corrain-goodreads-reviews.html>>  
[accessed 10 March 2024].

Anderson, Chris, 'The Long Tail', in *The Social Media Reader* (New York: New York University Press, 2012), pp. 137–52 <<https://www-fulcrum-org.ezproxy.leidenuniv.nl/epubs/fj236520q?locale=en#page=147>>

Andrejevic, Mark, Alison Hearn, and Helen Kennedy, 'Cultural Studies of Data Mining: Introduction', *European Journal of Cultural Studies*, 18.4–5 (2015), 379–94  
<<https://doi.org/10.1177/1367549415577395>>

Bennett, Asa, 'Amazon Warehouse Staff In "Slave Camp" Conditions, Workers Say', *The Huffington Post*, 25 November 2013  
<[https://www.huffingtonpost.co.uk/2013/11/25/amazon-staff-investigation\\_n\\_4335894.html](https://www.huffingtonpost.co.uk/2013/11/25/amazon-staff-investigation_n_4335894.html)> [accessed 12 April 2024].

Cirucci, Angela M., 'Redefining Privacy and Anonymity through Social Networking Affordances', *First Monday*, 20.7 (2015), 1–9  
<<https://doi.org/10.5210/fm.v20i7.5465>>

- Curcic, Dimitrije, 'Amazon Publishing Statistics', *WordsRated*, 12 January 2023  
<<https://wordsrated.com/amazon-publishing-statistics/>> [accessed 12 April 2024].
- Darnton, Robert, 'What Is the History of Books?', *Daedalus*, 111.3 (1982), 65–83.
- Dimitrov, Stefan, Faiyaz Zamal, Andrew Piper, and Derek Ruths, 'Goodreads vs Amazon: The Effect of Decoupling Book Reviewing and Book Selling', *Proceedings of the International AAAI Conference on Web and Social Media*, 9.1 (2021), 602-605  
<<https://doi.org/10.1609/icwsm.v9i1.14662>>
- Driscoll, Beth, and DeNel Rehberg Sedo, 'Faraway, So Close: Seeing the Intimacy in Goodreads Reviews', *Qualitative Inquiry*, 25.3 (2019), 248–59  
<<https://doi.org/10.1177/1077800418801375>>
- Fuller, Danielle, and DeNel Rehberg Sedo, 'Young Adult Readers and the Genres of Online Book Reviewing' <<https://post45.org/2023/12/it-was-plastered-all-over-my-instagram-last-year-at-least-for-my-algorithm-young-adult-readers-and-the-genres-of-online-book-reviewing/>> [accessed 10 February 2024]
- Giannopoulou, Alexandra, 'Digital Identity Infrastructures: A Critical Approach of Self-Sovereign Identity', *Digital Society*, 2.18 (2023), 1–19  
<<https://doi.org/10.1007/s44206-023-00049-z>>
- Gillespie, Tarleton, Pablo J. Boczkowski, and Kirsten A. Foot, *Media Technologies: Essays on Communication, Materiality, and Society* (The MIT Press, 2014)  
<<https://doi.org/10.7551/mitpress/9780262525374.001.0001>>
- Kapoor, Payal S., K.R. Jayasimha, and Ashish Sadh, 'Brand-Related, Consumer to Consumer, Communication via Social Media', *IIM Kozhikode Society & Management Review*, 2.1 (2013), 43–59 <<https://doi.org/10.1177/2277975213496514>>
- Kreizman, Maris, 'Let's Rescue Book Lovers from This Online Hellscape', *The New York Times*, 24 December 2023  
<<https://www.nytimes.com/2023/12/24/opinion/goodreads-books-reviews.html>>  
[accessed 10 March 2024].
- Lastoria, Amanda, 'Digital Masks of Printed Books: On-Screen Representations of the Materiality of the Codex' in *Bookshelves in the Age of the COVID-19 Pandemic*, (Berlin: Palgrave Macmillan, 2022).
- Maity, Suman Kalyan, Abhishek Panigrahi, and Animesh Mukherjee, '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*, (Sydney Australia: ACM, 2017), 451–54  
 <<https://doi.org/10.1145/3110025.3110138>>
- Matthews, Jolie C, ‘Professionals and Nonprofessionals on Goodreads: Behavior Standards for Authors, Reviewers, and Readers’, *New Media & Society*, 18.10 (2016), 2305–22  
 <<https://doi.org/10.1177/1461444815582141>>
- McLoughlin, Danny, ‘Amazon Book Sales Statistics’, *WordsRated*, 9 November 2022  
 <<https://wordsrated.com/amazon-book-sales-statistics/>> [accessed 10 March 2024].
- Murray, Simone, ‘Charting the Digital Literary Sphere’, *Contemporary Literature*, 56.2 (2015), 311–39 <<https://doi.org/10.3368/cl.56.2.311>>
- Murray, Simone, ‘Secret Agents: Algorithmic Culture, Goodreads and Datafication of the Contemporary Book World’, *European Journal of Cultural Studies*, 24.4 (2021), 970–89 <<https://doi.org/10.1177/1367549419886026>>
- Nakamura, Lisa, “‘Words with Friends’”: Socially Networked Reading on *Goodreads*’, *PMLA/Publications of the Modern Language Association of America*, 128.1 (2013), 238–43 <<https://doi.org/10.1632/pmla.2013.128.1.238>>
- O’Donovan, Caroline, ‘Goodreads Was the Future of Book Reviews. Then Amazon Bought It.’, *The Washington Post*, 1 July 2023  
 <<https://www.washingtonpost.com/technology/2023/07/01/amazon-goodreads-elizabeth-gilbert/>> [accessed 10 March 2024].
- Perez, Sarah. ‘Amazon Competes with Its Own Goodreads with Launch of Book Discovery Service, “Your Books”’, *TechCrunch*, 13 December 2023  
 <[https://techcrunch.com/2023/12/13/amazon-competes-with-its-own-goodreads-with-launch-of-book-discovery-service-your-books/?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\\_referrer\\_sig=AQAAAHcMgq-MAQNaPq\\_DLnSH9P4VdamhGTxwgbhICMHq9IQSYsTBVKMP2zLZ9iINvOhZkk7wc29wKkuuDHflfEkbtI5CL8Q\\_l9ZdynSiOS3sgiUCeTJ8mjfv5V4LnpKpyppfzbiVkmMVyY0NWZltFG0lboHydCwxm6ILVgZlgWb0HNa](https://techcrunch.com/2023/12/13/amazon-competes-with-its-own-goodreads-with-launch-of-book-discovery-service-your-books/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAHcMgq-MAQNaPq_DLnSH9P4VdamhGTxwgbhICMHq9IQSYsTBVKMP2zLZ9iINvOhZkk7wc29wKkuuDHflfEkbtI5CL8Q_l9ZdynSiOS3sgiUCeTJ8mjfv5V4LnpKpyppfzbiVkmMVyY0NWZltFG0lboHydCwxm6ILVgZlgWb0HNa)> [accessed 10 March 2024].
- Pinder, Julian, ‘ONLINE LITERARY COMMUNITIES: A Case Study of LibraryThing’. *From Codex to Hypertext: Reading at the Turn of the Twenty First Century*, (2012), 68–87. <http://www.jstor.org/stable/j.ctt5vk982.7>.



- Polk, Shelbi, ‘The Long Legacy of Book Clubs’, *Shondaland*, 23 October 2023  
<<https://www.shondaland.com/inspire/books/a45563057/the-long-legacy-of-book-clubs/>> [accessed 10 April 2024].
- Thelwall, Mike, and Kayvan Kousha, ‘Goodreads: A Social Network Site for Book Readers’, *Journal of the Association for Information Science and Technology*, 68.4 (2017), 972–83 <<https://doi.org/10.1002/asi.23733>>
- Vlieghe, Joachim, Jaël Muls, and Kris Rutten, ‘Everybody Reads: Reader Engagement with Literature in Social Media Environments’, *Poetics*, 54 (2016), 25–37  
<<https://doi.org/10.1016/j.poetic.2015.09.001>>
- Wright, David, ‘Literary Taste and List Culture in a Time of “Endless Choice”’, in *From Codex to Hypertext*, (University of Massachusetts Press, 2012), pp. 108-123  
<[https://web-p-ebSCOhost-com.ezproxy.leidenuniv.nl/ehost/ebookviewer/ebook/ZTAwMHh3d19fMTI0NTQ1OF9fQU41?sid=d3f247e2-2736-4336-997d-b2c0a89fa7e7@redis&vid=0&format=EB&lpid=lp\\_108&rid=0](https://web-p-ebSCOhost-com.ezproxy.leidenuniv.nl/ehost/ebookviewer/ebook/ZTAwMHh3d19fMTI0NTQ1OF9fQU41?sid=d3f247e2-2736-4336-997d-b2c0a89fa7e7@redis&vid=0&format=EB&lpid=lp_108&rid=0)>

### 3. Websites:

- Goodreads, ‘2023 Reading Challenge’ <<https://www.goodreads.com/challenges/11633-2023-reading-challenge>> [accessed 12 April 2024].
- Goodreads, ‘About Goodreads’ <<https://www.goodreads.com/about/us>> [accessed 15 March 2024].
- Goodreads, ‘About Goodreads Librarians’, <<https://help.goodreads.com/s/article/About-Goodreads-Librarians>> [accessed 10 March 2024].
- Goodreads, ‘Cookies Notice’, <[https://www.goodreads.com/about/cookies\\_notice](https://www.goodreads.com/about/cookies_notice)> [accessed 30 May 2024].
- Goodreads, ‘Privacy Policy’, <<https://www.goodreads.com/about/privacy>> [accessed 28 May 2024].
- McAuley Lab, ‘Amazon Reviews 2023’ (2024), *HuggingFace*,  
<<https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023/tree/main>> [accessed 22 June 2024].