





# Measuring Literary Quality Proxies and Perspectives

Pascale Feldkamp<sup>1</sup>   
Yuri Bizzoni<sup>1</sup>   
Ida Marie S. Lassen<sup>1</sup>   
Mads Rosendahl Thomsen<sup>2</sup>   
Kristoffer L. Nielbo<sup>1</sup> 

1. Center for Humanities Computing, Aarhus University , Aarhus, Denmark.
2. Comparative Literature – School of Communication and Culture, Aarhus University , Aarhus, Denmark.

## Citation

Pascale Feldkamp, Mads Rosendahl Thomsen, Kristoffer L. Nielbo, and Yuri Bizzoni (2024). “Measuring Literary Quality. Proxies and Perspectives”. In: *CCLS2024 Conference Preprints* 3 (1). [10.26083/tuprints-00027391](https://doi.org/10.26083/tuprints-00027391)

**Date published** 2024-05-28

**Date accepted** 2024-04-04

**Date received** 2024-01-09

## Keywords

literary quality, literary success, canonicity, literary culture, computational literary studies, 19th-20th century literature

## License

CC BY 4.0 

## Note

This paper has been submitted to the conference track of JCLS. It has been peer reviewed and accepted for presentation and discussion at the 3rd Annual Conference of Computational Literary Studies at Vienna, Austria, in June 2024.

**Abstract.** Computational studies of literature have adopted approaches from statistics and social sciences to perform large scale studies of fiction, and recent work has sought to approximate the success of literary texts using some proxy for literary quality, such as collections of human judgments, sales-numbers or lists indicating canonicity. However, most quantitative studies of literary quality use one such measure as a golden standard of literary judgement without fully reflecting on what it represents. Conclusions drawn from these studies are nonetheless bound to mirror a particular conception of literary quality associated with the chosen metric. To address this issue, we provide a discussion of the interrelation of various “proxies of literary quality” within a corpus of novels published in the US in the late 19th and 20th century, performing correlations and comparisons across 14 different proxies. We start with a heuristic distinction between expert-based literary judgments, such as those represented by college syllabi and literary anthologies, and crowd-based judgments, such as GoodReads’ ratings, and explore the differences between these and other proxies that fall in-between, such as library holding numbers, prestigious literary prizes, and classics book series. Our findings suggest that works favored in expert-based judgments tend to score lower on GoodReads, while those long-listed for awards tend to score higher and enjoy greater circulation in libraries. Generally, two main kinds of “quality perception” emerge as we map the literary judgment landscape: one associated with canonical literature, and one with more popular literature, which may indicate that judgements of canonicity or literariness are not equal to popularity among readers. Additionally, our study suggests that prestige in genre-literature, as represented by main genre-fiction awards such as the Hugo or World Fantasy Award, constitute distinct proxies on their own, though more closely aligned to popular than canonical proxies.

## 1. Introduction

The concept of quality in literature is a fascinating riddle: it would seem that the idiosyncratic nature of reading precludes any objective standard for what constitutes a “good” book – and yet certain texts seem to have an enduring appeal: they interest

1  
2  
3  
4

readers across time and national borders and are consecrated in the institutional canons of different cultures. This paradox lies at the heart of discussions about what literary quality is, as well as of attempts to define, measure or predict it.<sup>1</sup>

The challenge of defining literary quality is complicated by the diversity of preferences of individual readers and reader-types (Riddell and Dalen-Oskam 2018), and even the tendency of readers to change their opinion about a text (Harrison and Nuttall 2018; Kuijpers and Hakemulder 2018). Moreover, the question of what constitutes literary quality and where it resides (in style, plot, emotional engagement, themes, etc.) quickly becomes a complicated matter of its own, one that schools of literary criticism have grappled with in many different ways (Bjerck Hagen et al. 2018).

While the evaluation of texts and the question of quality has naturally been prominent in literary criticism, its significance has often been eclipsed within scholarly discourse by various disciplinary shifts. Ethical and postcolonial shifts calling attention to canon representativity (Peer 2008), methodological transformations of the 20<sup>th</sup> century moving the focus from evaluation towards interpretation (Bjerck Hagen et al. 2018), and the expansion of the conceptual boundaries of literature to encompass texts ideologically opposed to aestheticism or “pleasing” the reader (Wellek 1972), are examples that have played a role in making terms like “literary quality”, or “classics” unpopular – said to belong to the “precritical era of criticism itself” (Guillory 1995). However, to attribute the longevity or popularity of certain books to purely contextual factors and reject the notion of literary quality altogether would seem to be at odds with both the resilience of canons and consensuses among readers at the large scale, which appear far from volatile (Archer and Jockers 2017; Bizzoni et al. 2021; Maharjan et al. 2017, 2018; Wang et al. 2019).<sup>2</sup> Moreover, literary cultures have consistently established and upheld proxies of literary excellence in practice, such as literary awards, classics book series, or prescriptions in creative writing courses. Thus, a disparity appears to have arisen between a scholarly “denial of quality” (Wellek 1972) and the multitude of evaluative criteria actualized within literary culture.

With recent computational inquiry into literary studies, and sizeable attempts at quantifying “quality”, this disparity is even more apparent. The stricter conditions of quantitative analysis – operationalizing traditional disciplinary concepts – bring the complexity of the idea of “quality” in literature to the fore. Computational studies of literary preferences have found that reader appreciation or success can to some extent be predicted by stylistic features (Cranenburgh and Bod 2017; Dalen-Oskam 2023; Maharjan et al. 2017), as well as narrative features such as plot (Jockers 2015), emotional valence and flow

1. In this article, we will use the term “literary quality” in a general sense – as “quality in literature” – independently from kinds of texts (e.g. high-brow/low-brow) and evaluative groups (e.g. universities, online communities). That is, we do not intend to imply perceived *literariness*, but rather we aim to denote some form of appreciation of a literary work. In other words, our focus is not on whether a text appears to be high-brow, have sophisticated references to other works of literature and so forth, but rather on whether a text is considered outstanding by different types of readership.

2. A very Marxist reader, Leon Trotsky, observed how the historical and aesthetic dimensions of art are utterly independent: “If I say that the importance of the Divine Comedy lies in the fact that it gives me an understanding of the state of mind of certain classes in a certain epoch, this means that I transform it into a mere historical document, for, as a work of art, the Divine Comedy must speak in some way to my feelings and moods... Dante was, of course, the product of a certain social milieu. But Dante was a genius. He raised the experience of his epoch to a tremendous artistic height. And if we, while today approaching other works of medieval literature merely as objects of study, approach the Divine Comedy as a source of artistic perception, this happens not because Dante was a Florentine petty bourgeois of the 13th century but, to a considerable extent, in spite of that circumstance” (Trotsky 1974)

(Maharjan et al. 2018; Reagan et al. 2016; Veleski 2020), or the predictability of novels' sentiment-arcs (Bizzoni et al. 2022a,b, 2021) – not to mention text-extrinsic features such as genre, promotion, author visibility and gender (C. W. Koolen 2018; Lassen et al. 2022; Wang et al. 2019). While such studies point to the existence of certain consensuses, it should be noted that these studies define the concept of success or quality very differently. The first and possibly most complex task of quantitative studies of literary quality is that of defining a “proxy” of quality itself: from where should we take the judgments we intend to explain?

In computational literary studies, a “proxy” serves as a formal method for approximating abstract constructs or concepts through operationalization. Proxies bridge qualitative interpretation with quantitative methodologies: they translate constructs or concepts, like “quality in literature”, into measurable variables. A “quality proxy” thus means a specific operationalization of appreciation among many. For example, we might differentiate between literary “fame” and “popularity”, since fame, such as the fame of James Joyce’s *Ulysses* does not necessarily mean that it is widely read. These different forms of quality may be measured in dissimilar ways – i.e., through different “proxies” – for example by looking at how often a book is subject of literary scholarship, vs. how many copies it sells, or how often it is rated on GoodReads.<sup>3</sup>

A large number of quantitative and computational works have used votes of popularity to approximate judgments of literary quality. GoodReads is a widely used resources (Jannatus Saba et al. 2021; Maharjan et al. 2017; Porter 2018), also since it provides a single scale of scores averaged on large numbers of individual readers. The “GoodReads approach” can be seen as an example of “counting votes”, where the majority decides: the number of votes or a higher average score defines quality. On the polar opposite, a number of studies have used individual canon-lists of works selected by individual or cohorts of established literary scholars to approximate what are “quality works” of literature (Mohseni et al. 2022). Canon-lists or anthologies represent the idiosyncratic perspective of the few. Naturally this approach has advantages and disadvantages: “canon-makers” with or without institutional backing presumably have a vast knowledge of literature, but the criteria of selection are not always explicit and may or may not represent a particular taste or kind of reader. These limitations are, however, are homologous to those of the “GoodReads approach” where criteria and type of reader is likewise unknown (is it a particular type of reader who rates books online?). Studies have also modelled literary quality by whether or not a book has won a literary award (Febres and Jaffe 2017), which is akin to the “canon perspective”, but may differ in terms of the institutional affiliation of actors. Another method is to seek judgements of quality in the reading population (C. Koolen et al. 2020). Yet efforts of gauging readers’ conceptions of quality with sophisticated questionnaires is naturally limited by the difficulty and costs of conducting extensive surveys. Either of these approaches nevertheless runs the risk of modelling but one kind of “literary quality”, prompting reflections on how they are related. While some studies have tried to map the relations and overlaps between kinds of quality proxies (Manshel et al. 2019; Porter 2018), usually experiments are conducted on a limited scale, either in terms of corpus, or in terms of

3. At present, *Ulysses* has 124,536 ratings on GoodReads and a relatively low average rating of 3.75, compared to works such as Suzanne Collins’ *The Hunger Games* and J.K. Rowling’s *Harry Potter and the Sorcerer’s Stone*, with above 8 million ratings and average ratings above 4.3.

the number and types of quality proxies considered.	83
The question remains of how different proxies relate to an overall concept of literary quality: do different proxies offer windows or perspectives into a more or less universal perception of quality, or do such proxies represent vastly different forms of appreciation?	84
Do, for instance, GoodReads scores mirror, on a larger scale, the selection of experts, such as for literary anthologies, or do they diverge to such an extent that we may assume that what is judged to be “quality” in each proxy is based on different criteria?	85
	86
	87
	88
	89
To address the question of differences between quality proxies, we collected 14 different possible proxies for literary quality, ranging from popular online platforms to university syllabi and prestigious awards, and used them to annotate a corpus of over 9,000 novels (note that we do not analyze the texts themselves in this article). <sup>4</sup> Our central question was whether and to what extent these metrics measure the same thing: if the “quality” measured by GoodReads data differs from that represented by the number of library holdings, the two metrics will have nothing in common; if instead there is a significant overlap - that is, books popular on GoodReads are also acquired by many libraries - they will correlate. To the best of our knowledge, this is the first study that tries to compare several judgements of literary quality on a large collection of modern titles, trying to understand, by a rigorous approach, the relation between them.	90
	91
	92
	93
	94
	95
	96
	97
	98
	99
	100

## 2. Related Works 101

Studies have found that there seems to be a consensus among readers about what works are “classics”. Walsh and Antoniak (2021) tested the relation between GoodReads’ Classics, a user-compiled list, and titles included in college English syllabi (as collected by the OpenSyllabus project), showing that there is a significant overlap between what is perceived as classics on GoodReads and what appears on college syllabi (Walsh and Antoniak 2021). Thus, users seem to be replicating a particular perception of the “canonicity” of titles. 102  
103  
104  
105  
106  
107  
108

Similarly, Koolen et al. (2020) surveyed a large number of Dutch readers, asking for both judgments of how “enjoyable” and how “literary” a novel is, and have shown that there is a more substantial consensus among readers about “literariness” than “enjoyability”-ratings, which appear less predictable than those of literariness (C. Koolen et al. 2020). 109  
110  
111  
112

Another study by Porter et al. (2018) sought to model differences in popularity and prestige in their corpus, using, on the one hand, GoodReads’ average ratings and, on the other hand, the Modern Language Association’s database of literary scholarship, counting the number of mentions of an author as the primary subject of a scholarly work. They show that there is a clear difference in the equilibrium between popularity and prestige across genres. Books from genres like sci-fi are rated very often on GoodReads but are sparsely represented in scholarly work, while poetry exhibits an opposite tendency. Based on Pierre Bourdieu’s conceptualization of the literary field, they define two axes of literary “success”, prestige and popularity as online popularity (on GoodReads) and prestige among literary scholars (represented in the MLA database), so that their 113  
114  
115  
116  
117  
118  
119  
120  
121  
122

4. See section 4 for a discussion of this corpus, which, it should be noted, is heavily skewed toward American and Anglophone authors.

“map” risks to look overly neat. Literary scholars, for example, may not be the primary nor most important actors in processes of literary prestige, and Manshel et al. (2019) have shown how literary prizes – appointed by committees who may be either authors themselves, scholars, or lay-readers – appear to have an important role in positively influencing both prestige and popularity.<sup>5</sup>

While only a few studies have tried to measure differences and convergences of literary quality judgments quantitatively, the question of how literary cultures evaluate texts has been central to sociological approaches to literature. Especially the attempts of Pierre Bourdieu to “map” the literary field is central in this context and has given rise to a string of seminal works on power dynamics in literary cultures (Bennett 1990; Casanova 2007; Guillory 1995; Moretti 2007). Bourdieu’s map of the French “literary field” (1) focuses on literary genres and their interrelation in terms of prestige (and not actors in literary quality judgments *per se*). However, Bourdieu makes an important distinction between types of audiences and considers “consecration by artists, by institutions of the dominant classes, and by popular success” as distinct axes, that are more or less mutually exclusive.<sup>6</sup>

conference version

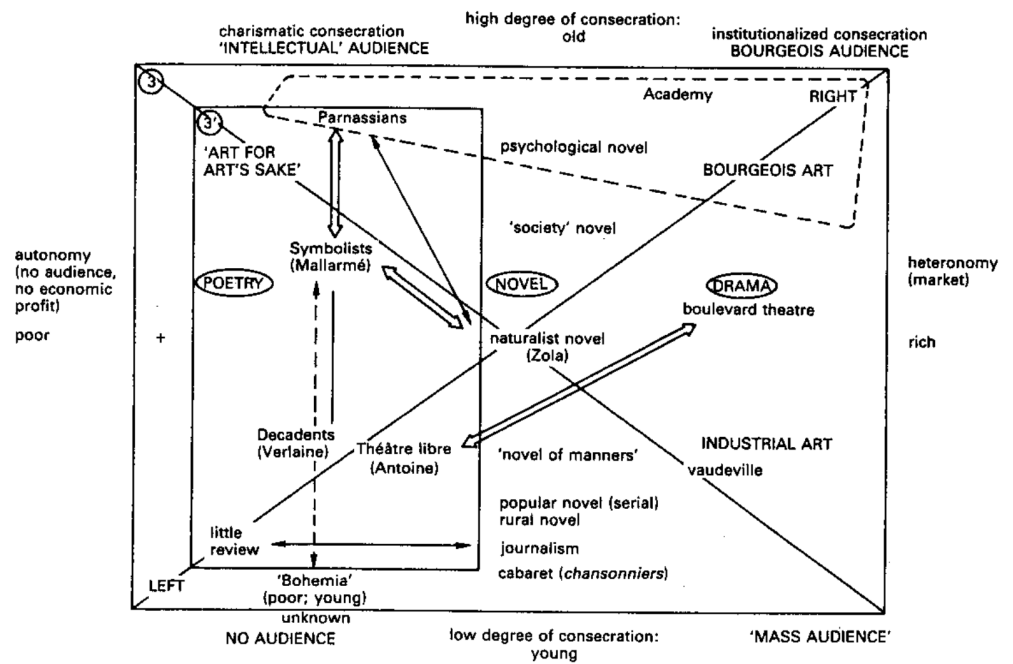


Figure 1: Bourdieu’s French literary field of the late 19<sup>th</sup> century, with audience or popularity on the x-axis and consecration or prestige on the y-axis.

While the relation between these actors is only sketched out (and it is the present study’s aim to inspect these more closely), Bourdieu’s map can serve as a heuristic conceptualization of types of actors in literary quality judgments. Here, the idea of expert-based and crowd-based literary judgments is apparent at either pole, represented

5. Using the same definitions of popularity and prestige as Porter et al. (2018), it seems that whether or not books had received a prize significantly raised the probability of both being popular and prestigious (Manshel et al. 2019).  
 6. Bourdieu writes: “there are few fields [beyond the literary] in which the antagonism between the occupants of the polar positions is more total” (Bourdieu 1993, p. 46).

on one side by intellectual and bourgeois audiences, recognized intellectuals such as “Parnassians” and institutions such as *l’Académie Française*; and on the other hand by amateur and mass audience such as the artistic underdogs “bohemia” and popular media. As Porter et al. (2018) have shown, “on a broad level, real-world data about popularity and prestige appear to confirm Bourdieu’s intuitions” (Porter 2018). In their visualization the genres “Mystery & Thriller” and “Science Fiction & Fantasy” appear where Bourdieu places the “Popular novel” (at low consecration and high economic profit), while poetry is in the upper left area of the map, representing high prestige and low popularity. However, the focus of Porter et al. is on the right-hand part of Bourdieu’s map, with prestige defined as institutional or academic consecration: the place for literary works in academia. For a more comprehensive “map” based on real world data, various actors, including literary prizes and publishers, should be considered. It is to this end that the present paper uses a sizeable corpus to examine the interrelation judgments of a type of “success” in the literary field, including various actors under the general categories of expert-based and crowd-based literary success based off Bourdieu’s “map”. We discuss the selection of various proxies and what they represent, before moving on to looking at their distribution and interrelation in the Chicago corpus.

### 3. Selecting Types of Literary Judgments

By considering various proxies of literary quality, our aim was to examine the interrelation of conceptually different types. We considered three distinct approaches to literary quality:

1. Approaches that seek to approximate literary canonicity or quality in an institutional sense, looking at which works or authors are included in school or university syllabi, literary anthologies, or that win literary awards.
2. Approaches that seek to approximate reader-popularity, basing proxies of literary quality on larger populations, where the selection process appears more “democratic”, seeking the quality perception of “layman readers”, by collecting user-generated data such as ratings from sites like GoodReads, Amazon, or Audible.
3. In-between approaches that seek to measure the market success or market resilience of works, looking at, for example, sales figures.

#### 3.1 Expert-based Quality Proxies

Expert-based proxies of literary quality may to an extent be synonymous with canonicity, that is, consecration and institutionalization. Often, quantitative studies of reader appreciation define canonicity or prestige through canon lists compiled by, i.a., individual magazines (Vulture 2018, as in Porter 2018), editors (Karlyn and Keymer n.d., as in Algee-Hewitt et al. 2018), or literary scholars (Bloom 1995, as in Mohseni et al. 2022). However, such lists resemble personal canons that may not have a wide reach, e.g., it is unclear how widely accepted Harold Bloom’s chosen canon is among scholars. In this study, we have preferred canonicity proxies that do not depend on the selection of

very few. To examine expert-based proxies of literary quality and estimate the amount of “canonic” literature in our dataset, we marked all titles by authors that appear in selected institutional or user-compiled proxies that indicate literary prestige: a literary anthology, the most assigned titles in English Literature course syllabi, literary awards, and a publisher’s classics series.

### 3.1.1 Anthologies

Students of English or of Literature will often be acquainted with anthologies that are compiled in part for educational use, facilitating easy access to some key works. In this context, the Norton Anthology in particular is a leading literary anthology (Pope 2019), with diachronic series of English and American literature that are widely used in education (Shesgreen 2009). For the present study, we marked all titles in our corpus written by authors mentioned in these two series, where the anthology of English Literature is the most widespread (Ragen 1992).

### 3.1.2 Syllabi

While titles assigned on Literature or English syllabi surely vary across colleges and regions, it is possible to find trends and most assigned titles via large collections of data, such as by the OpenSyllabus project, which has collected 18.7 million college syllabi in an attempt to map the college curriculum.<sup>7</sup> From this data, we took all titles in our corpus by authors who appear as authors of one of the top 1000 titles assigned in English Literature college syllabi.

### 3.1.3 Awards and Long Lists

We collected long-listed titles (winners and finalists) for both prestigious general literature awards: The Nobel Prize in Literature, the Pulitzer Prize, the National Book Award (NBA); as well as various genre-based awards (for the full list, see Table 1). The choice of long-lists allowed us to have a more titles annotated, but also an annotation possibly less susceptible to the extrinsic factors that can influence the choice of a winner among a small selection of candidates in the moment (politics, topic, prominence of the author, and so forth).

Manshel et al. (2019) have shown that winning an award does contribute to long-term prestige – but also popularity – of titles in academia as well as on GoodReads. Interestingly, Kovács and Sharkey (2019), found that while awards may initially make a title more popular and gather more ratings on GoodReads, this may also affect a drop in average rating as the reception of a book becomes polarized. As such, the choices of award-committées do seem to be in touch with the general public, but also diverge from consensuses among readers at the very large scale Kovács and Sharkey 2014. We keep genre-awards and more general literary awards separate in our analysis, as we expect titles to be received differently across genres. As our corpus catalogues mainly American and British authors, the focus of our selection was the topmost known committee-based awards in anglophone literary culture.

7. See: <https://www.opensyllabus.org>.

**3.1.4 Classics Series** 223

Various large publishing houses, like Vintage or Penguin<sup>8</sup>, maintain a classics series. As Penguin is arguably one of the biggest publishers of anglophone literature (Alter et al. 2022), we marked all titles or authors in our corpus that appear in their classics series. We looked at both the specific titles (title-based) with matches in our data, and at all titles by authors featured in the series (author-based), keeping these separate in our analysis.

**3.2 Crowd-based Quality Proxies** 230

Where proxies of quality are clearly vote-based and the result of equal weight for each individual in a large population, we call them “crowd-based”, remembering, however, that these votes are cast within a system and social structures (e.g., on the social platform GoodReads), which are not non-hierarchical as the term “crowd-based” generally implies, nor isolated from tendencies of expert-based proxies. For example, the canonicity perception of GoodReads’ users may have more to do with expert-based proxies of literary quality than we think (Walsh and Antoniak 2021). Among crowd-based measures, we have opted for GoodReads and Audible average rating (number of “stars” given to a title) and rating count (number of votes). We also used two GoodReads user-compiled lists: the “GoodReads classics” and the “Best books of the 20<sup>th</sup> century” which may represent canonic literature but at a larger scale than expert-based canonicity lists.

**3.2.1 GoodReads** 243

GoodReads is a social network or “social catalogue site” with links to other social networks (Facebook, Twitter, Instagram, and LinkedIn), designed for readers to discover, review, and share their thoughts. Otis Chandler, GoodReads’ co-founder, states on the homepage that the idea was to make a social forum akin to looking at the bookshelf at a friend’s house: “When I want to know what books to read, I’d rather turn to a friend than any random person or bestseller list.” With its 90 million users, GoodReads arguably offers an insight into reading culture “in the wild” (Nakamura 2013), as it catalogues books from a wide spectrum of genres and derives book-ratings from a heterogeneous pool of readers in terms of background, gender, age, native language and reading preferences (Kousha et al. 2017). GoodReads’ average ratings represent the average user rating of titles. Rating ranges from 0 stars (indicating low appreciation) to 5 stars (indicating high appreciation). The average score provides a general indication of the book’s reception, but is problematic as it conflates types of literary appreciation, i.e., satisfaction, enjoyment, and evaluation, to one scale. While it is important to note that these GoodReads’ ratings and number of raters (rating count) do not present an absolute measure of literary quality or even popularity (GoodReads did start with predominantly American users), they do offer a valuable perspective on a work’s overall popularity among a diverse population of readers. Beyond ratings, GoodReads also compiles vote-based lists and “shelves”, arranged according to the titles most often either assigned to a particular list or tagged to a particular shelf. These are, for example, GoodReads’ Classics, Best Books of the 20th Century, The Worst Books of All Time, etc.

8. See: <https://www.penguin.com/penguin-classics-overview/>.



For the present study, we used the top 100 of a popular list, the Best Books of the 20th Century<sup>9</sup>, and a shelf, the GoodReads' Classics<sup>10</sup>, where titles were *listed* by users 600 to 10,000 times, and *shelved* 15,588 to 64,903 times, respectively.

### 3.2.2 Audible

We use the average rating and number of ratings of title on Audible, the Amazon audiobook service. Like GoodReads, the site uses a five-star scale for user ratings, however, the amount of users and the rating counts are significantly lower for Audible compared to GoodReads: while Dan Brown's *The Da Vinci Code* has 2,259,837 ratings on GoodReads, it has 3,225 ratings on Audible at the moment of writing, and the average Audible rating is inflated in comparison to the GoodReads' average rating for our corpus, which may be an effect of a smaller number of users.

### 3.3 In-between Quality Proxies

The number of copies sold is often adopted as a reliable standard to estimate the success novels, for example to gauge a set of signals that land a book on the bestseller list Archer and Jockers 2017. It is interesting because a proxy like sales figures seems to stand in-between the crowd- and expert-based proxies, including a degree of resilience or canonicity of titles (as classics will continue to sell) as well as popular demand. The NPD BookScan<sup>11</sup>, for example, is a popular resource in this regard (as used in Wang et al. 2019), which provides data for the publishing industry both regarding genre, prices, and weekly sales figures for all books published in the US since 2003. It is clear that such data is market- and location-specific, and is only an option for studies of more contemporary works. As with any other approximation of literary quality, but perhaps especially pertaining to sales figures, the issue is both that data pertains to more recent publications, is not readily available, and that contextual factors may influence the data. For book-sales, Wang et al. (2019) have shown that marketing, the particular publishing house, and visibility of the author plays a central role for sales numbers. Instead of sales-figures, we may use proxies that also include an aspect of resilience and popular success. Thus, we have used the number of libraries holding a given title on Worldcat and the number of translations of a work into other languages, as well as the author's presence on Wikipedia and a bestseller list. The number of library holdings as a proxy is conceptually intermediate between a completely free, crowd-based vote count and an expert-driven single choice, as the list of books held by libraries depends on both popular demand (of library-card holders) and expert choices (librarians). Similarly, the translational success of a work shows a degree of market success (if translation is seen as a token of publishers seeking to expand sales of bestselling books outside the national market) and canonicity or resilience (if translation is seen as a token of a work's cultural longevity or durable popularity). Similarly, Wikipedia rank and bestseller lists appear conceptually to include a degree of resilience and popular success.

9. See: [https://www.GoodReads.com/list/show/6.Best\\_Books\\_of\\_the\\_20th\\_Century](https://www.GoodReads.com/list/show/6.Best_Books_of_the_20th_Century).

10. See: <https://www.GoodReads.com/shelf/show/classics>.

11. See: <https://www.npd.com/industry-expertise/books/>.

**3.3.1 Library Holdings** 303

For each title, the Chicago Corpus provides the number of US libraries holding a copy of it. The idea is that libraries' choices could help indicate a canon that is not arbitrary (as libraries supposedly respond to institutional demands like school reading requirements) but also remains essentially crowd-based (as libraries also respond to other demands, including from leisure-readers). Libraries are institutions managed by experts, but adding together the choices of thousands of different libraries allows the selection to partly overcome the risks involved in electing one single, if well-informed, authority.

**3.3.2 Translations** 311

The *Index Translationum* database<sup>12</sup> collects all translations published in ca. 150 UNESCO member states, compiled from their local bibliographical institutions or national libraries. It catalogues more than 2 million works across disciplines. Note that the database was created in 1979 and stopped compiling in 2009. Thus, we are not looking at the most translated works through time, where the "classics" may be more frequent, but at a particular period, and the results should be interpreted with that in mind.

**3.3.3 Wikipedia Author-page-rank** 318

Using wikipedia page-views, that is, the number of times visits to an author's page on Wikipedia is also sometimes used as a proxy for popularity or resilience. Hube et al. (2017) have used Wikipedia metrics to measure in the centrality of authors in digital space (Hube et al. 2017), with a variation of page-rank, the original google algorithm. It is an efficient way to navigate graphs: hubs or author-pages on Wikipedia that have the highest number of other pages referencing them have a higher rank, which means a higher rank for more referenced authors. The Wikipedia page rank thus measures a type of "canonicity" of authors, but also their presence in the popular and cultural sphere, if we consider that Wikipedia-pages are created both by experts and lay-readers. For the present study, we used Wikipedia author-page (WAP) rank, where it should be noted that ranks refer to authors, so that books by the same author will have the same rank, independent from differences between individual titles.

**3.3.4 Bestseller Lists** 331

To gauge the commercial success of titles, we also marked titles in our corpus that were also extant in the Publisher's Weekly American 20<sup>th</sup> century bestseller list.<sup>13</sup> Publishers Weekly is a trade news magazine which is published once a week (from 1872) and targeted at agents within the field: publishers, literary agents, booksellers, and librarians. While sales numbers are considered, the full set of selection criteria for the list are unknown.

12. See: <https://www.unesco.org/xtrans/bsform.aspx>.

13. Extracted from the database by John Unsworth at the University of Illinois: <https://web.archive.org/web/20111014055658/http://www3.isrl.illinois.edu/~unsworth/courses/bestsellers/picked.books.cgi>.

## 4. Dataset: the Chicago Corpus 338

In order to quantify the possible convergence of these proxies, we need a dataset of chosen titles. A large dataset of titles would allow us to see whether different ways of scoring or judging literary works tend to have something in common (e.g. valuing similar texts) or not. Ideally, for a first experiment, we would also require a selection of texts that are not too widespread in time, written/read in the same language, and in the same narrative form (e.g. all prose novels).

We base our study on the *Chicago corpus*,<sup>14</sup> a corpus of over 9,000 manually compiled novels that were either written or translated into English and published in the US between 1880 to 2000. The corpus was compiled based on the number of libraries holding a copy of the novel, with a preference for novels with more holdings. Beyond responding to the constraints detailed above, the Chicago corpus allows us to access the number of libraries holding each title in the US. Moreover, the Chicago corpus has been curated and used by teams of literary scholars, and offers access to the full text of all its titles, which makes a study of correlations between quality judgments and textual features possible in the future.

Because of its unique method of compilation, the Chicago corpus is a rare dataset in terms of its diversity: it spans works from genre-fiction and popular fiction (i.a., Isaac Asimov, Agatha Christie, George R. R. Martin), to seminal works from the entire period, central modernist and postmodernist texts (e.g. James Joyce's *Ulysses* and Don DeLillo's *White Noise*), as well as winners of the Nobel Prize (i.a., Ernest Hemingway, William Faulkner, Toni Morrison), and other prestigious literary awards (i.a., Cormac McCarthy). As such, it represents a sizeable subsection of both prestigious or "canonic" works, as well as popular and genre-fiction classics.

It should be noted that the Chicago corpus contains only works either written in or translated into English, and therefore exhibits an over-representation of Anglophone authors.

We previously discussed the essential characteristics of these proxies of literary quality, as well as the kind of outlook on literary judgments that they seem to model or approximate. Some are on the free and vote-counting end of the spectrum, putting equal weight to the rating of each user. Resources like the Norton collection, as well as prestigious literary awards, arguably fall on the expert-based side of the spectrum, as they are managed by small groups of authoritative readers, usually professional literary critics.

By collecting and annotating proxies of quality for titles in the Chicago corpus, we collected a wide variety of "quality judgments" for each title, some continuous (as GoodReads' average ratings) or progressive (as the number of library holdings), some discrete, as any list that either includes or excludes titles. This, as we will see, constitutes a fundamental divide between our measures, and in some sense mirrors two different ways of assessing literary quality. The resources that in one way or another score each book – number of ratings, number of library acquisition, average rating – represent quality on a continuum, while the resources that select books – anthologies, syllabi and

14. For more on the corpus, see the resource at: [https://github.com/centre-for-humanities-computing/chicago\\_corpus](https://github.com/centre-for-humanities-computing/chicago_corpus).

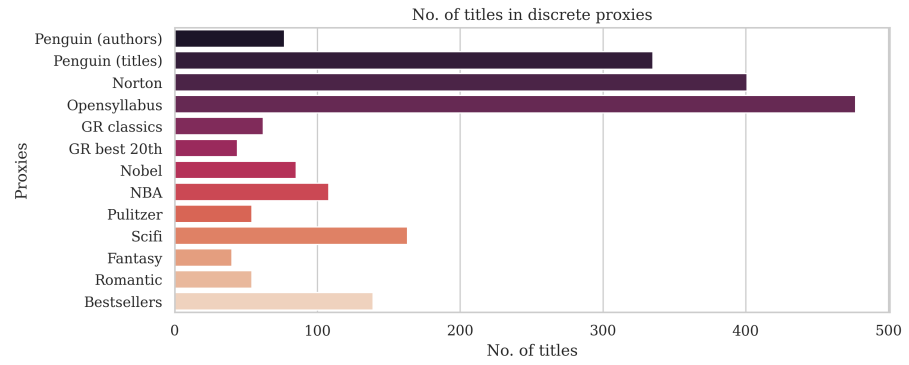


Figure 2: Sizes of discrete proxies in our corpus.

	Titles
National book award	108
Pulitzer prize	53
Nobel prize*	85
Scifi awards	163
Hugo award	
Nebula award	
Philip K. Dick award	
J.W. Campbell award	
Prometheus award	
Locus sci-fi award	
Fantasy awards	40
World fantasy award	
Locus fantasy award	
British fantasy award	
Mythopoeic award	
Romantic awards*	54
Rita awards*	
RNA awards*	
<hr style="border-top: 1px dashed black;"/>	
Norton anthology*	401
OpenSyllabus*	477
Penguin classics series (titles)	77
Penguin classics series*	335
GoodReads' classics*	62
GoodReads' best books of the 20th century*	44
<hr style="border-top: 1px dashed black;"/>	
20th century bestsellers (Publisher's Weekly)	139
Wikipedia AP rank*	3558
Translations	5082
GR avg. rating	8989
GR rating count	8989

Table 1: Number of titles in the corpus per quality proxy. Proxies followed by \* are author-based: For these, we included all titles extant in the corpus by the author mentioned, either due to the scarcity of awards in the genre or the nature of the award/list, e.g., the Nobel prize given to authors rather than to individual titles. All other proxies are title-based.

awards – are discrete, representing quality as a threshold. 379

In the following sections, we examine the relation between these proxies, assessing the 380  
 correlation between them, how they are situated in a network, and their intersections. 381

## 5. Results

382

### 5.1 Correlation

383

Having annotated the titles in our corpus for these proxies, we looked at the correlations 384  
 between them to see how and whether they interplay. As some values are discrete and 385  
 others are not, the correlation matrix is often a measure of overlap: if the correlation 386  
 coefficient at the intersection of *Penguin classics* and *Norton* is a high number, the two 387  
 proxies have large overlaps. Computing a Spearman or Pearson correlation between two 388  
 discrete lists means checking whether and to what extent the two lists include the same 389  
 items. Finally, correlations between discrete and continuous values tell us whether there 390  
 is a sizable change in values when switching from one category to another – for example, 391  
 whether there is a sizable change in scores between books that were long-listed for a 392  
 given award and books that were not.<sup>15</sup> 393

conference version

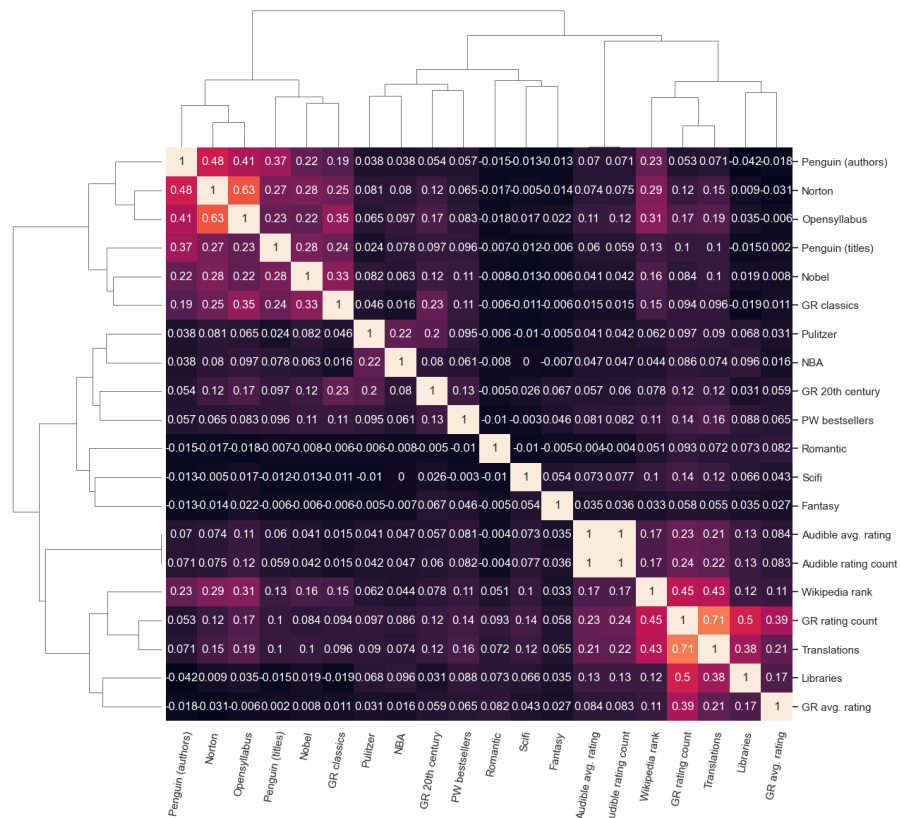


Figure 3: Correlations between discrete and continuous measures of literary quality (Spearman correlation). The matrix shows hierarchical clustering by Ward’s method.

Looking at the correlation matrix resulting from our dataset we find intriguing correlations 394  
 between proxies of appreciation. Firstly, we find that there seem to be two “islands” 395  
 with stronger internal correlations: one spans, roughly, GoodReads and Audible number 396  
 of ratings and average ratings along with the Library holdings; the other is more or less 397  
 connecting what we could call “canon lists” – GoodReads’ best books of the 20th century, 398  
 GoodReads’ Classics, the Nobel, Opensyllabus, the Norton anthology, and the Penguin 399

15. It is crucial to remember that a correlation between a discrete and a continuous variable is not equivalent to a t-test of significance, as we will discuss later; that is, random samples from the same population could show a valid correlation, and vice versa: samples from two populations could show no correlation at all.

Classics Series, and (somewhat surprisingly) the bestsellers. Weak correlations happen out of these two areas - Wikipedia's rank correlates with Sci-fi awards, but not with the more mediatized Pulitzer prize, the award which, together with the Nobel, correlates with GoodReads' best books of the 20th century. However, these do not correlate with each other. Furthermore, the number of ratings of GoodReads and Audible shows correlations with Opensyllabus, the Norton anthology, and the Penguin Classics series.

Secondly, if we disregard the Nobel prize, which correlates with "canon" proxies such as Opensyllabus, the awards do not overlap much with one another, and do not display strong correlations with other categories. Beyond the mentioned correlations of the Pulitzer and Nobel with the GoodReads' list of best books of the 20th century, awards – and especially genre-awards – do not appear to correlate with other proxies. This lack of correlation is relevant, especially as it means that long-listed works of genre-literature appear to have no strong presence in resources like the Norton anthology or in the GoodReads' Classics list, indicating the strong presence of general fiction in these resources. However, it is still possible that the awards elicit a particular range of ratings in terms of GoodReads' ratings or libraries holdings without eliciting a detectable correlation. Also, not surprisingly, genre-fiction awards do not overlap with more literary awards (such as the Pulitzer, National Book Award, and the Nobel). At the same time, the Pulitzer and National Book Award do converge. The awards of Romantic fiction and Fantasy are the most removed, showing little convergence other proxies.

In sum, we could hypothesize that we are seeing the difference between two types of quality modeling, one that corresponds to crowd-based measures (GoodReads, Audible) and one that relates to more expert-based measures (Opensyllabus, Norton). The first category includes only measures based on counting votes - the number of people who rated a book and the average values of all users' ratings. Instead, The second category appears to be lists defined by small groups of experts that exclude or include titles, even if that group, as in the case of the GoodReads' Classics, may be lay readers.

It is notable that what we have called the "in-between" measure of library holdings correlates more strongly with crowd-based proxies (GoodReads, Audible). The correlations range from slight to robust with GoodReads' and Audible's rating count and GoodReads' average ratings. That is, books that many people rate or listen to on those platforms also tend held by many libraries. In this sense, the group consisting of "canon" lists appear like a product of the idiosyncrasies of small expert groups, to be overcome when many annotators are actually in the picture.

However, note that the second "island" of correlations does include GoodReads' classics list and, to an extent, the GoodReads' best books of the 20th century, two lists constituted through the votes of thousands or tens of thousands of individual users. Also, if the second group's selections were completely idiosyncratic and independent from each other, they would not correlate with each other, yet show evident convergence. Finally, the "expert-based" status of *Opensyllabus* can be questioned, given that it is the collection of several independent college choices, and is, in that sense, closer to the library holdings.

Thus, no clear distinction between these two clusters can be based on the method of selection (expert-based versus crowd-based), but may be based, rather, in the form of perceived canonicity or literariness that tells the second group from the first. In other

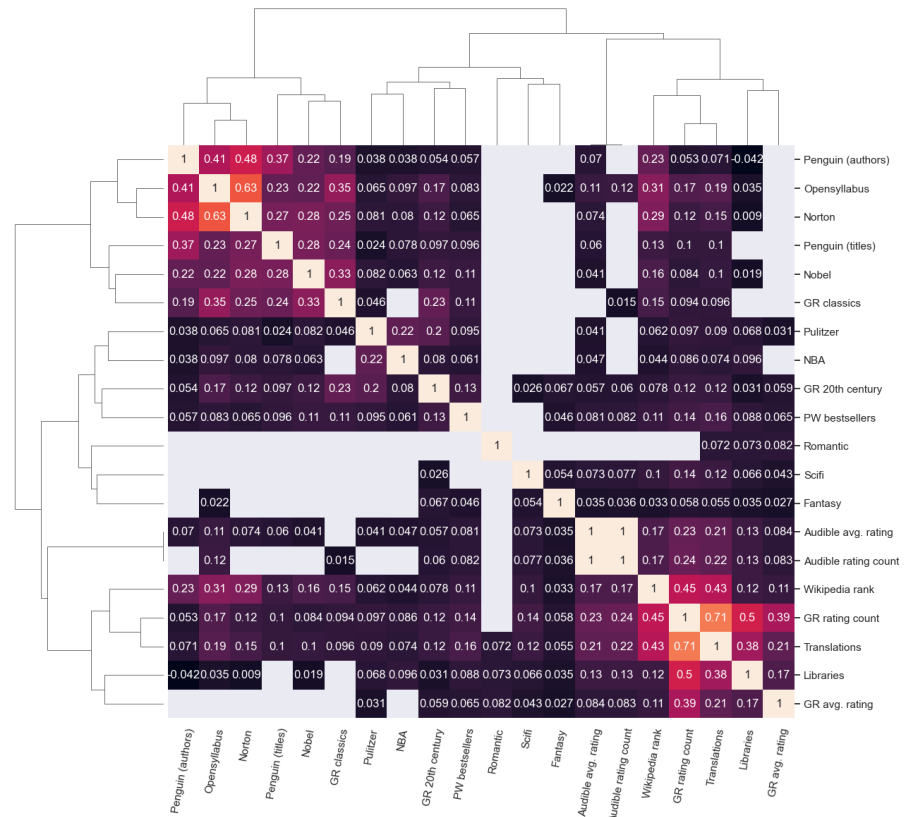


Figure 4: Again, correlations between discrete and continuous measures of literary quality (Spearman correlation), this time with non-significant correlations masked (p-value < 0.05)

words, what we are seeing might be two different “faces” of the concept of literary quality 444  
 that may be perceived by the same reader. An observation supporting that there should 445  
 be two main “perceptions” of quality is that the users of GoodReads seem not to give 446  
 the highest ratings to the titles of the Norton anthology. Still, when GoodReads users 447  
 constitute lists of “classics” and “20th century best”, they converge with the anthology 448  
 on similar ground. 449

### 5.2 Network 450

As we have seen, continuous proxies of literary quality, such as GoodReads’ ratings and 451  
 library holdings seem to correlate. However, a visualization of their convergence shows 452  
 that the correlation may not be strictly linear (Fig. 5). 453

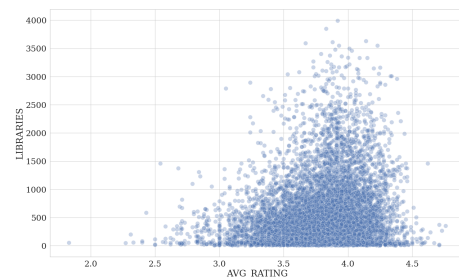


Figure 5: Scatterplot of library holdings vs. avg. rating of all titles with a threshold of 5 ratings.

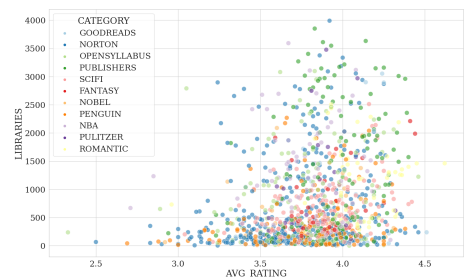
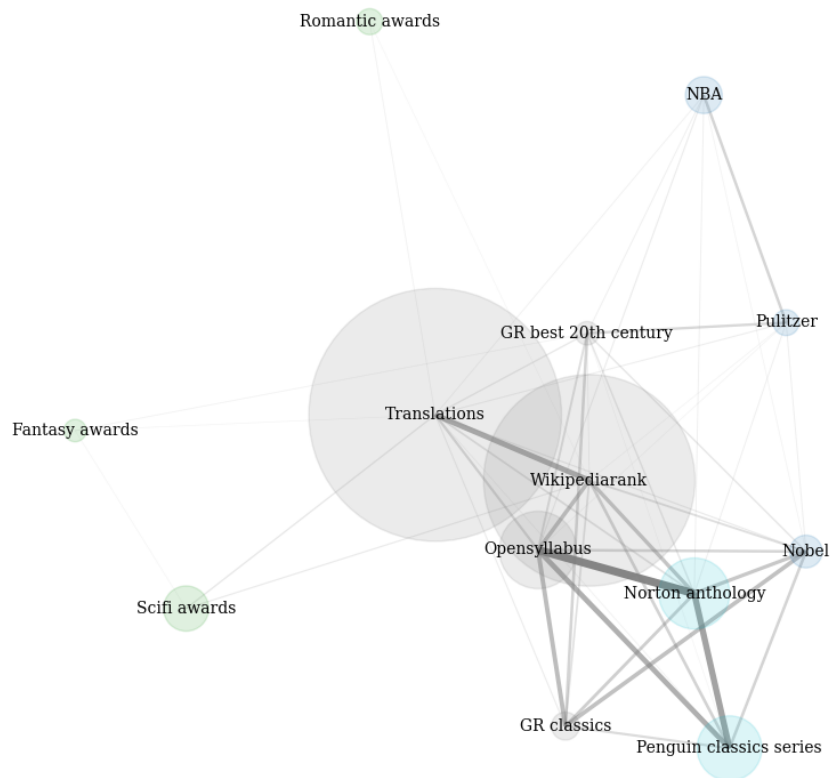


Figure 6: Scatterplot of library holdings vs. avg. rating of titles contained in one of the quality proxies.

Indeed, the interrelation between different proxies may be difficult to gauge when 454  
 looking at correlation coefficients and visualizations. Proxy interrelations are better 455  
 visualized in the literary quality standard landscape when visualized as a network, 456  
 where each node represents one proxy and each edge the correlation (i.e., for discrete 457  
 lists, the overlap) between proxies. 458



**Figure 7:** Network of literary quality proxies with edge-width and opacity based on the correlation coefficient between proxies (Spearman correlation), excepting the corpus-wide categories of GoodReads' ratings. We apply a coefficient threshold of 0.05 for edges being visualized. Positions are likewise determined by correlation between proxies, using the Fruchterman-Reingold force-directed algorithm for positioning. The sizes of the nodes are determined by the number of titles in each proxy. Colors are used to indicate similar types of awards: literary awards, genre-fiction awards, book-series/anthology.

As was also apparent in the correlation matrix (Fig. 3), longlists of genre-fiction awards 459  
 tend to be far removed from other proxies, with a slight correlation between Fantasy 460  
 and Scifi-awards, which might be explained by the thematic overlap between these 461  
 genres. The disconnection between more "literary proxies" like the Penguin Classics 462  
 series and the Norton Anthology may also be affected by relabelling of genre-fiction 463  
 in literary markets. Genre tags may act like implicit quality judgments themselves: 464  
 prestigious horror is often relabelled "gothic" or "literary fiction" and doesn't even 465  
 run for genre-awards (think of, i.a., Bram Stoker and Mary Shelley). Genre-labelling 466  
 is a complex issue, where various cultural factors and market forces may play in. 467  
 For example, works by women authors are often labeled or re-labeled into less prestigious 468  
 genres, such as 'Romantic fiction' over 'literary novel' (Groos 2000). 469



In our network, books listed in the *Index Translationum* show a strong correlation with author's in our Wikipedia-page-rank data, and also have a large actual overlap: 52.7 percent of translated books are books by authors in our Wikipedia-page-rank data, and 75.3 percent of books by authors in our Wikipedia-page-rank data are also in the *Index Translationum*-list of translated works. While literary awards, National Book Award and Pulitzer do show some overlap, the cluster of most related proxies seems to be the more expert-based expert-based type of proxy: especially Opensyllabus, Norton Anthology, and the Penguin Classics series form a distinct triangle in the network. Books that are in one of these three proxies also tend to be in the other, which is particularly interesting in this case, since the underlying selection mechanisms of these the three seem distinct, split between institutional and commercial affiliations. Nevertheless, their selection still converges on some shared perception of quality of titles. Furthermore, the divergence of awards from the remaining proxies, as well as the divergence between award-types of general (National Book Award, Pulitzer) and genre-fiction is even more apparent in the network, while the Nobel prize shows stronger convergences with the mentioned triad of more canonical, expert-based proxies, indicating its difference from the other prestigious awards.

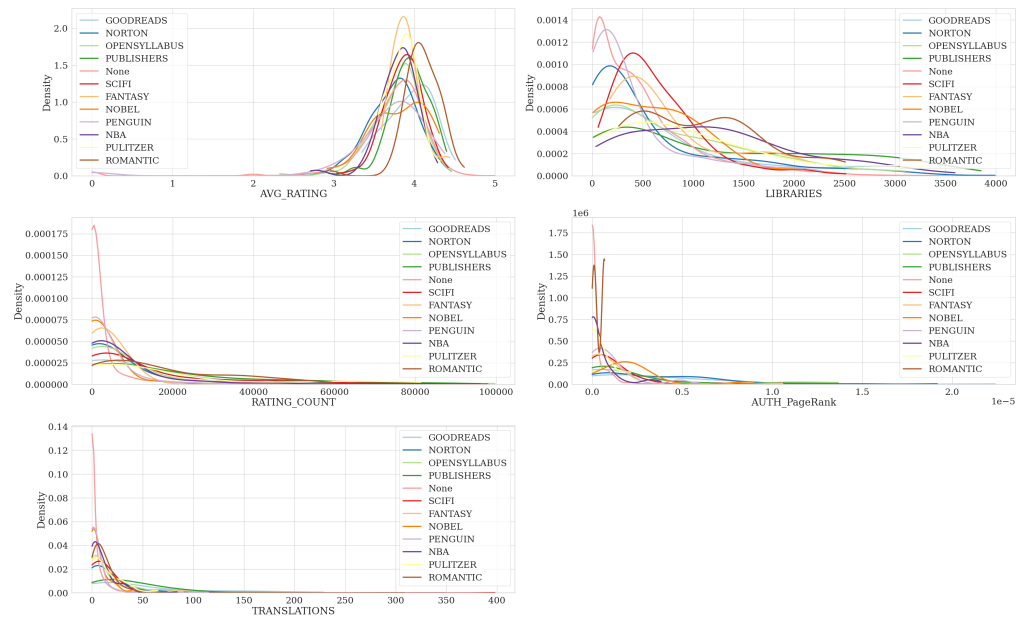
### 5.3 Intersection

	GR avg. rating	GR rating count	Library holdings	Translations	WAP rank
Corpus average	3.75	14246.36	535.74	6.58	0.000058
Opensyllabus	3.78	109831.81	738.05	25.22	0.000423
Penguin classics*	3.72	57105.42	463.54	16.18	0.000334
Penguin classics (titles)	3.76	194615.08	496.74	43.14	0.000418
Norton	3.74	74424.81	687.75	22.09	0.000402
GoodReads' classics	3.82	4307090.65	501.37	57.11	<b>0.000869</b>
GoodReads' best books of the 20th century	4.04	<b>992225.89</b>	998.41	<b>98.02</b>	0.000439
Nobel	3.81	119078.32	811.09	32.04	0.000558
NBA	3.83	62071.08	1266.10	17.28	0.000111
Pulitzer	3.91	135290.26	<b>1498.77</b>	33.98	0.000176
Scifi awards	3.88	73716.60	701.81	13.81	0.000135
Fantasy awards	3.92	164753.12	804.28	18.27	0.000158
Romantic awards	<b>4.09</b>	31595.07	1078.24	11.69	0.000037
Bestsellers	3.94	120453.92	1290.56	43.03	0.000222

**Table 2:** Intersectional values: mean continuous quality-measure per discontinuous proxies. Bold font indicates the highest mean within the selection of proxies. Note that the Wikipedia rank (WAP) has been multiplied by 100, because of the generally low values.

Correlations are not the only way of checking whether two categories converge: our continuous values (library holdings, GoodReads' average ratings and rating count, translation and Wikipedia page rank) may be used to distinguish between discrete proxies. For example, Pulitzer prize winners might elicit consistently higher GoodReads' ratings than the corpus average. In this example, we would propose that GoodReads' ratings exhibit a "convergence" with the Pulitzer resource. Similarly, it may be that one type of award has systematically higher ratings and more library holdings than other books, indicating an affinity to the perception of quality affecting library holdings. In other words, there may not be a correlation between but still a convergence of two categories. Examining proxy intersections in this way, we look at the distribution of continuous proxy-values of each discrete proxy, comparing this distribution to titles in our corpus that are not contained in any of our selected quality proxies.

When visualizing the distribution of titles of different categorical proxies in terms of our the continuous proxies (rating count, translations, etc.), we see that titles included



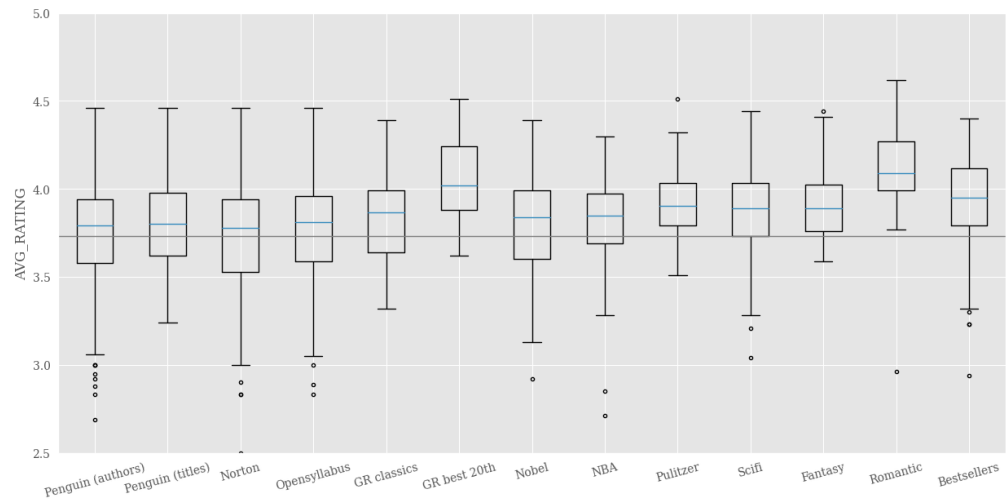
**Figure 8:** Kernel density estimate (KDE) plots of the distributions of measures per quality proxy. Note that rating count values above 100,000 have been filtered out for the purpose of visualization. “None” represents titles that are not in either of the proxies.

in categorical quality proxies generally have a longer tail and may have different distributions than titles not contained in any proxy of quality (“None” in Fig. 8). Looking at GoodReads’ average rating and library holdings, books included in categorical proxies seem to have smoother slopes in comparison to the rest of the corpus, whereas in terms of rating count, Wikipedia Author-page Rank and translations, we see a much higher amount of works in either proxy having very low values, with a long tail of few outliers at very high values. Measures such as rating count tend to exhibit a log-type distribution.

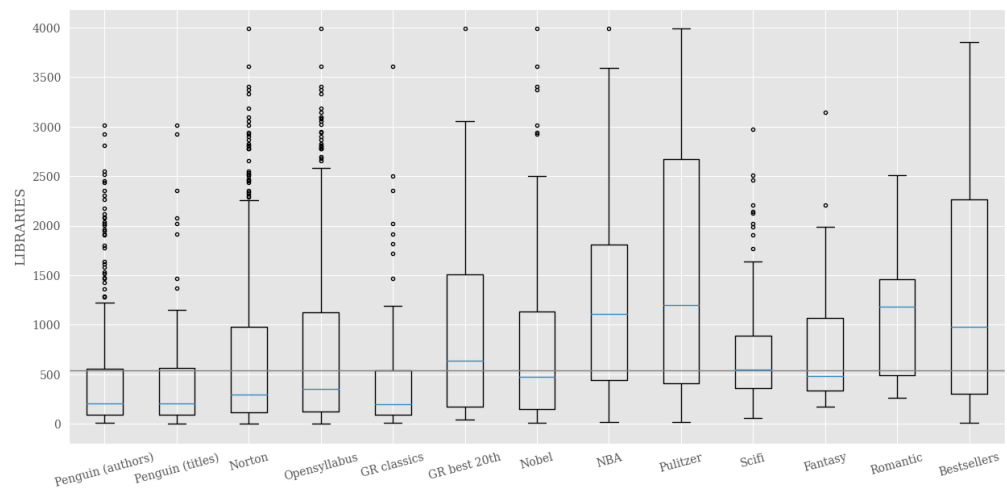
Moreover, different categorical proxies peak at different values within the continuous proxies. For example, the distribution of books that have won a Romantic literary award seem to peak at a higher value of GoodReads’ average rating, having also the highest mean average rating of any proxy (Tab. 2).<sup>16</sup> Titles in GoodReads’ Classics, Nobel prize, Opensyllabus and Norton Anthology are represented more evenly across values of Wikipedia Author-page Rank, which may be expected as we also saw that these proxies seem to be closely related in our network. It indicates that these base their selection on some shared perception of quality, which may also prompt their authors to have more prominent Wikipedia pages. Interestingly, the plot showing distributions over library holdings shows a somewhat opposite tendency: here, genre-fiction tends to place at higher values, so that Sci-fi, Fantasy and Romantic fiction, for example, peak at higher values, and have high mean library holdings numbers (Tab. 2). In general, the two “islands” of quality observed in our correlation matrix (Fig. 3) can be observed in the colors that peak in the different quadrants, genre fiction in some, what we could call more “higher brow” or canonical literature in others.

16. Note that the odd distribution of Romantic titles in the plots with library holdings and Wikipedia Author-page Rank rank may be an effect of the small number of titles. It may be that one author who has higher canonicity is responsible for the peak at the higher end in both plots.

Visualizing the mean values of each discrete proxy in terms of continuous proxies further aids in gauging the differences between these quality perspectives (Fig. 9-13).



**Figure 9:** Boxplot of average GoodReads rating for discrete categories. The grey line indicates the corpus average rating.



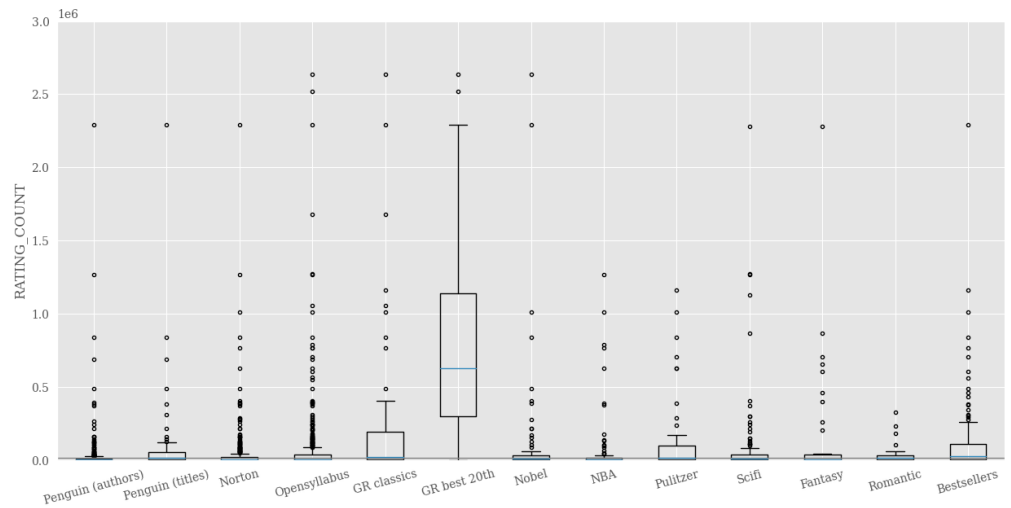
**Figure 10:** Boxplot of average number of library holdings for discrete categories. The grey line indicates the corpus average holdings.

GoodReads’ best books of the 20th century appear to have the highest average GoodReads’ ratings, closely followed by Hugo and Pulitzer titles, while the Norton and Opensyllabus titles record the lowest average ratings (Tab. 9). Overall, Opensyllabus’ and Norton Anthology titles score consistently lower with respect to any other category in terms of their GoodReads’ average ratings as well as their number of libraries holdings (10).

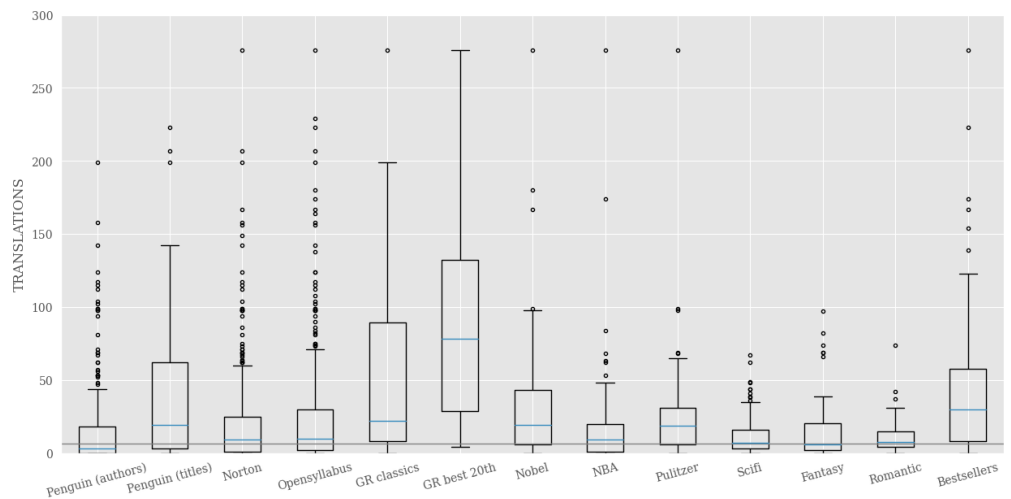
GoodReads’ best books of the 20th century is the only proxy that stands out in terms of GoodReads’ rating count (Fig. 11). Note that rating count is a problematic proxy because of it’s non-normal distribution, with very few titles at very high values, which is why we see a low corpus mean with many outliers for each proxy as well as long whiskers for the GoodReads’ best books of the 20th century category.

Translation numbers and Wikipedia Author-page Rank are the two continuous measures that appear similar in the sense that titles longlisted for awards tend to score low in

conference version



**Figure 11:** Boxplot of rating count of discrete categories. The grey line indicates the corpus average rank.



**Figure 12:** Boxplot of average translation numbers for discrete categories. The grey line indicates the corpus average number.

comparison to, for example, GoodReads' Classics titles. Again, there is a difference between general fiction awards (National Book Award, Pulitzer) and genre-fiction awards, where titles longlisted for genre-fiction awards tend to place lower. It is interesting that for these two plots (Fig. 12, 13), the user-generated lists GoodReads' Classics and best books of the 20th century score high, with a subtle difference between the two plots. When looking at translation numbers, we see that GoodReads' best 20th century books score higher than GoodReads' Classics, and that bestsellers are also one of the proxies with higher mean translation numbers. Conversely, when looking at the Wikipedia Author-page Rank, we see that GoodReads' Classics have a higher mean than the best 20th century books, and that the Nobel titles, as well as the more expert-based measures that showed the strongest affinities in our network (7) also have a higher mean in comparison to when looking at translation numbers. Considering each of these boxplots together, overall, we observe the following patterns:

1. Titles longlisted for awards, both general fiction and genre-awards, tend to have

- higher average GoodReads' rating and library holdings. 553
- 2. The proxies we found to be strongly correlated in the "island" of our correlation matrix representing more "canonical" fiction (Fig. 3), Opensyllabus, Norton, and GoodReads' Classics, tend to have lower average GoodReads' ratings and library holdings. 554 555 556 557
- 3. There is a partial convergence between vote-based continuous scores and discrete categories. While translation numbers and Wikipedia Author-page Rank seem to ascribe higher values to more "canonical" fiction, GoodReads' users and library holdings they seem to have a higher appreciation for awards and genre-fiction, and a lower appreciation for the canon. 558 559 560 561 562

We clearly note a distinct variation among quality proxies, with an inclination of proxies of similar affiliation type – i.e., institutional, intellectual, commercial – to exhibit analogous behavior. Especially awards appear less aligned to other proxies of literary quality in terms of correlation (Fig. 3, 7). Nevertheless, titles longlisted for awards in our corpus enjoy a higher appreciation among users of GoodReads and a higher circulation in libraries. This agrees with the approach of Manshel et al. (2019), who consider awards an distinct form of quality proxy Manshel et al. 2019. 563 564 565 566 567 568 569

Looking at the different types of awards, we seem to confirm Bourdieu's intuition that the literary field is polarized: our genre-award proxies appear far removed from other proxies (including more general literary awards, see Fig.7). Yet they have higher average GoodReads' ratings and library holdings than, for example, the more institutionally oriented Norton Anthology. These characteristics would situate titles of genre-awards roughly at the place of the "popular novel" in Bourdieu's map of the literary field, which also aligns with the study of the prestige versus popularity of genre fiction by Porter 2018. In contrast, a proxy like the Norton Anthology, may be situated more toward the "intellectual" and "bourgeois" poles of Bourdieu's map, considering it is part of the inter-linked triangle of proxies observed in our network (Fig. 7), of which Opensyllabus has an institutional status. The clear divergence between proxies like the Norton Anthology 570 571 572 573 574 575 576 577 578 579 580

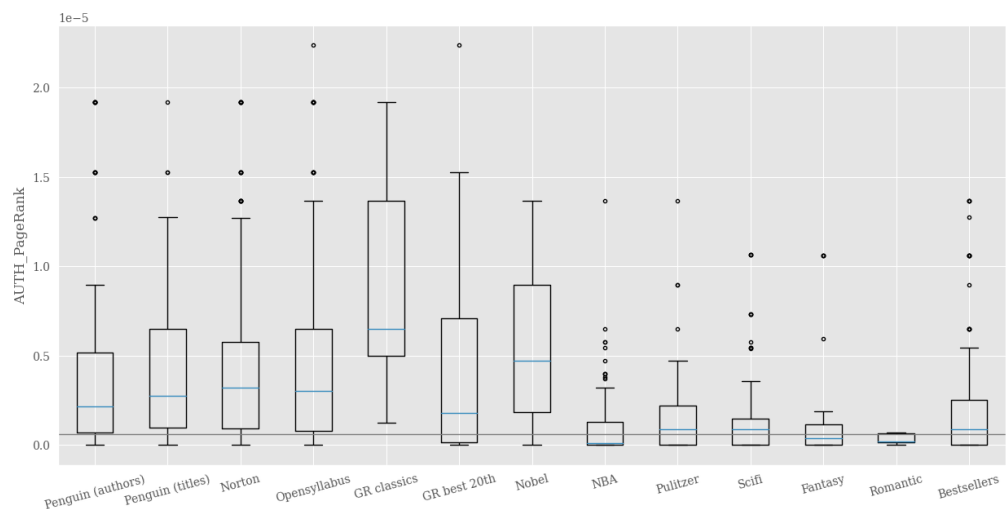


Figure 13: Boxplot of average Wikipedia AP rank for discrete categories. The grey line indicates the corpus average rank.

and genre-fiction awards may be explained by differences in style and topic of books, 581  
 but studies have also suggested that different types of audiences appreciate books at 582  
 different levels of readability (Bizzoni et al. 2023). Thus, the divergence may also have 583  
 to do with socio-cultural factors like population literacy, where more “readable” works 584  
 are preferred at the level of larger audiences, and more institutionally acclaimed works, 585  
 such as those included in the Norton Anthology less so, partly because of difficulty at 586  
 the sentence level. 587

Following Bourdieu, we might contrast actors behind the general fiction award proxies as 588  
 “intellectual audiences” against those behind genre-fiction awards as a “mass audience” 589  
 (Fig. 1). However, it is important to note we do not find audiences to be as polarized 590  
 or distinct as Bourdieu suggested. Rather, proxies seem to transverse their actor-type 591  
 affiliations. For instance, while bestsellers and Opensyllabus have dissimilar actors 592  
 underlying them – institutional versus market-oriente – bestsellers had the strongest 593  
 correlation with Opensyllabus as seen in Fig. 3. These findings imply the potential 594  
 existence of two overarching types of “quality perception,” which overlay and interlink 595  
 proxies underpinned by divergent actors or audiences. This insight emerges from the 596  
 observation of two “islands” when looking at correlations (3), but also from looking 597  
 into the differential favoring of each of the continuous measures contained in the first 598  
 “island”. When exploring discrete proxies in terms of the continuous ones, we saw 599  
 that GoodReads’ ratings and library holdings on one side, and translation numbers 600  
 and Wikipedia page-rank on the other were more similar in the way they value, for 601  
 example, longlisted titles for genre-awards. This suggests that actor or audience-based 602  
 distinctions might not fully capture the intricate dynamics of appreciation judgments in 603  
 the literary field. 604

When looking at proxies in terms of the distinction between expert-based or crowd- 605  
 based, we do see vote-based or what we could characterize as “crowd-based” proxies 606  
 cluster in terms of correlation: Audible average ratings with GoodReads’ average ratings, 607  
 as wells as libraries, translation numbers and Wikipedia Author-page Rank, of which 608  
 the latter may, in part, represent tastes of lay-readers (see section 3.3.3). However, 609  
 continuous crowd-based proxies also differ: GoodReads’ ratings and library holdings 610  
 numbers assign higher values to some proxies, like awards, which proxies like Wikipedia 611  
 Author-page Rank does not. Wikipedia Author-page Rank is also the proxy which mostly 612  
 strongly bridges the two “islands” in our correlation matrix, exhibiting correlations with 613  
 both “islands” (Fig. 3), which may explain its different behaviour and which may more 614  
 properly situate it between expert-based and crowd-based type of proxies. As such, we 615  
 may use the distinction between expert-based and crowd-based proxies heuristically, 616  
 though it seems that more complex judgements based on different quality “perceptions” 617  
 contribute to the clusters we have observed. 618

## 6. Conclusion and Future Works 619

Generally, we seem to observe two types of “quality perception”, or two faces of the 620  
 concept of quality, emerge through the differences and surprising convergences of the 621  
 host of proxies considered in the present study. 622

There appears to be a perception of titles’ canonicity in expert-based proxies like Open- 623

syllabus that does not converge much with the popularity of a title on crowd-based resources like GoodReads. In this sense, we validated and expanded Walsh and Antoniak 2021's study, as we too observed the convergence of different canonicity proxies, including those compiled on GoodReads by large numbers of unqualified readers. This suggests the presence of two distinct modes of evaluating quality, which can mirror two macro-classes of reader types (Riddell and Dalen-Oskam 2018) or can be even accessible to individual readers as they navigate different dimensions of assessment.

This duality is reminiscent of several similar dichotomies theorized in previous works: C. Koolen et al. 2020's distinction of literariness and enjoyability, Porter 2018 and Manshel et al. 2019's distinction between prestige and popularity, and naturally of Bourdieu 1993's two axes of institutionalized vs popular art. Yet, the duality that emerges from our data is nuanced and does not represent a polar opposition, but rather fuzzy islands between different proxies. Bestseller lists agree with canonical groups and with GoodReads' metrics, and the distinctness of titles included in longlists for genre awards might even indicate a possible third – or many – different perceptions of quality, which may be connected to various extra- and intra-textual features.

This is not surprising: indeed, as we mentioned in the beginning, every literary judgment is unique insofar as it is based on idiosyncratic or internalized interpretations of the text, various expectations suggested by the genre of a title, its publication date, textual features, the cover, etc. For example, one type of book may be more demanding to read and likely set the expectation bar of readers higher, genre-codes influence readers quality judgements or attract types of readers, and so on. The consensuses among readers found in recent computational studies, which suggest that textual features inform quality judgements (i.a., Bizzoni et al. 2021; Dalen-Oskam 2023; Maharjan et al. 2017; Wang et al. 2019) should therefore be interpreted with an eye to the type of proxy used in the particular study.

More complicated is the possible influence of social structures and power dynamics (Bennett 1990; Casanova 2007; Guillory 1995; Moretti 2007) on quality judgments: it is possible that we see the effect of crowd-based types of proxies being more diverse in terms of gender, reviewer background, etc. so that they appear to form a different "perception" of quality. This would not explain, however, why what we would understand as a crowd-based type of proxy, the bestseller list, seems to correlate with expert-based proxies. Examining the characteristics of titles at the textual level in conjunction with considerations of various quality proxies – but also considering likely biases influencing literary judgements – would help shed further light on the complex issue of measuring literary qualities. Nevertheless, what we have called two main "perceptions of quality" in this study cannot be completely idiosyncratic since two main groups of proxies do correlate and seem to converge on similar grounds, despite differences in their nature.

Various limitations inhere to the selection of quality proxies and to the quality proxies themselves, and it should be noted that various other proxies could be collected, among others, sales figures. Moreover, different literary cultures may vary in their ways of assessing quality, while this study is clearly situated in an Anglophone and American context. In terms of challenges in assessing the quality proxies themselves, for example, it is possible that GoodReads represents a contemporary audience so that canonical literature, assessed over decades or centuries, does not precisely align with their tastes.

In future studies, we suggest a closer inspection of possible biases, such as the publication 669  
 dates of titles, as well as gender or race biases influencing literary judgements. We also 670  
 suggest a stronger focus on the interplay between textual features and different types of 671  
 quality proxies. For example, assessing the importance of readability for different types 672  
 of proxies, which is an often underrated metric that may, among other things, likely 673  
 account for the demise of certain avant-garde works over time, as well as the difference 674  
 in preference between types of audiences. 675

## 7. Data Availability 676

Data can be found here: [https://github.com/centre-for-humanities-computing/c 677](https://github.com/centre-for-humanities-computing/chicago_corpus)  
[hicago\\_corpus](https://github.com/centre-for-humanities-computing/chicago_corpus). 678

## 8. Author Contributions 679

**Pascale Feldkamp:** Analysis, writing, review & editing 680

**Yuri Bizzoni:** Analysis, writing, review & editing 681

**Ida Marie S. Lassen:** Methodology, project administration 682

**Mads Rosendahl Thomsen:** Methodology, review & editing, project administration 683

**Kristoffer L. Nielbo:** Methodology, project administration 684

## References 685

- Algee-Hewitt, Mark, Sarah Allison, Marissa Gemma, Ryan Heuser, Franco Moretti, 686  
 and Hannah Walser (2018). *Canon/archive : large-scale dynamics in the literary field*. 687  
 Vol. Stanford Literary Lab. Pamphlets of the Stanford Literary Lab 11, 14. [itlab.stanford.edu/LiteraryLabPamphlet11.pdf](https://l 688</a><br/>
<a href=). 689
- Alter, Alexandra, Elizabeth A. Harris, and David McCabe (July 2022). "Will the Biggest 690  
 Publisher in the United States Get Even Bigger?" In: *The New York Times*.  [\[trust-trial.html\]\(https://www.nytimes.com/2022/07/31/books/penguin-random-house-simon-schuster-anti 693\). 693](https://ww 691</a><br/>
<a href=)
- Archer, Jodie and Matthew L Jockers (2017). *The bestseller code*. London: Penguin books. 694
- Bennett, Tony (1990). *Popular Fiction: Technology, Ideology, Production, Reading*. Routledge. 695
- Bizzoni, Yuri, Pascale Moreira, Nicole Dwenger, Ida Lassen, Mads Thomsen, and Kristof- 696  
 fer Nielbo (2023). "Good Reads and Easy Novels: Readability and Literary Quality 697  
 in a Corpus of US-published Fiction". In: *Proceedings of the 24th Nordic Conference on 698*  
*Computational Linguistics (NoDaLiDa)*. Tórshavn, Faroe Islands: University of Tartu 699  
 Library, 42–51. <https://aclanthology.org/2023.nodalida-1.5>. 700
- Bizzoni, Yuri, Telma Peura, Kristoffer Nielbo, and Mads Thomsen (2022a). "Fractal Sen- 701  
 timents and Fairy Tales-Fractal scaling of narrative arcs as predictor of the perceived 702  
 quality of Andersen's fairy tales". In: *Journal of Data Mining & Digital Humanities*. 703  
[10.46298/jdmdh.9154](https://doi.org/10.46298/jdmdh.9154). 704



- Bizzoni, Yuri, Telma Peura, Kristoffer Nielbo, and Mads Thomsen (2022b). “Fractality of sentiment arcs for literary quality assessment: The case of Nobel laureates”. In: *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*. Taipei, Taiwan: Association for Computational Linguistics, 31–41. <https://aclanthology.org/2022.nlp4dh-1.5>.
- Bizzoni, Yuri, Telma Peura, Mads Rosendahl Thomsen, and Kristoffer Nielbo (2021). “Sentiment Dynamics of Success: Fractal Scaling of Story Arcs Predicts Reader Preferences”. In: *Proceedings of the Workshop on Natural Language Processing for Digital Humanities*. Silchar, India: NLP Association of India (NLPAI), 1–6. <https://aclanthology.org/2021.nlp4dh-1.1>.
- Bjerck Hagen, Eric, Christine Hamm, Frode Helmich Pedersen, Jørgen Magnus Sejersted, and Eirik Vassenden (2018). “Literary Quality: Historical Perspectives”. In: *Contested Qualities*. Ed. by Knut Ove Eliassen, Jan Hovden, and Øyvind Prytz. Fagbokforlaget, 47–74.
- Bloom, Harold (1995). *The Western Canon: The Books and School of the Ages*. First Riverhead Edition. New York, NY: Riverhead Books.
- Bourdieu, Pierre (1993). *The field of cultural production: essays on art and literature*. Ed. by Randal Johnson. Columbia University Press.
- Casanova, Pascale (2007). *The World Republic of Letters*. Harvard University Press.
- Cranenburgh, Andreas van and Rens Bod (Apr. 2017). “A Data-Oriented Model of Literary Language”. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguistics, 1228–1238. <https://aclanthology.org/E17-1115>.
- Dalen-Oskam, Karina van (June 2023). *The Riddle of Literary Quality*. ISBN: 978-90-485-5814-8. <https://www.aup.nl/en/book/9789048558148/the-riddle-of-literary-quality> (visited on 04/12/2024).
- Febres, Gerardo and Klaus Jaffe (2017). “Quantifying literature quality using complexity criteria”. In: *Journal of Quantitative Linguistics* 24.1, 16–53. [10.1080/09296174.2016.1169847](https://doi.org/10.1080/09296174.2016.1169847).
- Groos, Marije (2000). “Wie schrijft die blijft? Schrijfsters in de literaire kritiek van nu”. In: *Tijdschrift voor Genderstudies* 3.3.
- Guillory, John (1995). *Cultural Capital: The Problem of Literary Canon Formation*. University of Chicago Press.
- Harrison, Chloe and Louise Nuttall (2018). “Re-reading in stylistics”. In: *Language and Literature* 27.3. SAGE Publications Ltd, 176–195. [10.1177/0963947018792719](https://doi.org/10.1177/0963947018792719).
- Hube, Christoph, Frank Fischer, Robert Jäschke, Gerhard Lauer, and Mads Rosendahl Thomsen (2017). *World Literature According to Wikipedia: Introduction to a DBpedia-Based Framework*. <http://arxiv.org/abs/1701.00991>.
- Jannatus Saba, Syeda, Biddut Sarker Bijoy, Henry Gorelick, Sabir Ismail, Md Saiful Islam, and Mohammad Ruhul Amin (2021). “A Study on Using Semantic Word Associations to Predict the Success of a Novel”. In: *Proceedings of \*SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, 38–51. <https://aclanthology.org/2021.starsem-1.4>.
- Jockers, Matthew L (2015). *Syuzhet: Extract sentiment and plot arcs from text*.
- Karlynn, Danny and Tom Keymer (n.d.). *Chadwyck-Healey Literature Collection*. [http://collections.chadwyck.com/marketing/products/about\\_ilc.jsp?collection=nfc](http://collections.chadwyck.com/marketing/products/about_ilc.jsp?collection=nfc).

- Koolen, Corina, Karina van Dalen-Oskam, Andreas van Cranenburgh, and Erica Nagel-  
hout (2020). "Literary Quality in the Eye of the Dutch Reader: The National Reader  
Survey". In: *Poetics* 79, 1–13. [10.1016/j.poetic.2020.101439](https://doi.org/10.1016/j.poetic.2020.101439). 752  
753  
754
- Koolen, Cornelia Wilhelmina (2018). *Reading beyond the female: the relationship between  
perception of author gender and literary quality*. ILLC dissertation series DS-2018-03. 755  
Amsterdam: Institute for Logic, Language and Computation, Universiteit van Ams-  
terdam. ISBN: 978-94-028-0951-0. 756  
757  
758
- Kousha, Kayvan, Mike Thelwall, and Mahshid Abdoli (2017). "GoodReads reviews to  
assess the wider impacts of books". In: *Journal of the Association for Information Science  
and Technology* 68.8, 2004–2016. ISSN: 2330-1643. [10.1002/asi.23805](https://doi.org/10.1002/asi.23805). 759  
760  
761
- Kovács, Balázs and Amanda J Sharkey (2014). "The Paradox of Publicity". In: *Adminis-  
trative Science Quarterly* 1, 1–33. [10.1177/0001839214523602](https://doi.org/10.1177/0001839214523602). 762  
763
- Kuijpers, Moniek M. and Frank Hakemulder (2018). "Understanding and Appreciating  
Literary Texts Through Rereading". In: *Discourse Processes* 55.7, 619–641. [10.1080/01  
63853X.2017.1390352](https://doi.org/10.1080/0163853X.2017.1390352). 764  
765  
766
- Lassen, Ida Marie Schytt, Yuri Bizzoni, Telma Peura, Mads Rosendahl Thomsen, and  
Kristoffer Laigaard Nielbo (2022). "Reviewer Preferences and Gender Disparities in  
Aesthetic Judgments". In: *CEUR Workshop Proceedings*. Antwerp, Belgium, 280–290. 767  
768  
769  
[https://ceur-ws.org/Vol-3290/short\\_paper1885.pdf](https://ceur-ws.org/Vol-3290/short_paper1885.pdf). 770
- Maharjan, Suraj, John Arevalo, Manuel Montes, Fabio A. González, and Tamar Solorio  
(2017). "A Multi-task Approach to Predict Likability of Books". In: *Proceedings of  
the 15th Conference of the European Chapter of the Association for Computational Linguis-  
tics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguis-  
tics, 1217–1227. <https://aclanthology.org/E17-1114>. 771  
772  
773  
774  
775
- Maharjan, Suraj, Sudipta Kar, Manuel Montes, Fabio A. González, and Tamar Solorio  
(2018). "Letting Emotions Flow: Success Prediction by Modeling the Flow of Emo-  
tions in Books". In: *Proceedings of the 2018 Conference of the North American Chapter of the  
Association for Computational Linguistics: Human Language Technologies: Volume 2, Short  
Papers*. New Orleans, Louisiana: Association for Computational Linguistics, 259–265. 776  
777  
778  
779  
780  
<https://aclanthology.org/N18-2042>. 781
- Manshel, Alexander, Laura B McGrath, and J.D. Porter (2019). *Who Cares about Literary  
Prizes?* <https://www.publicbooks.org/who-cares-about-literary-prizes/>. 782  
783
- Mohseni, Mahdi, Christoph Redies, and Volker Gast (2022). "Approximate Entropy in  
Canonical and Non-Canonical Fiction". In: *Entropy* 24.2, 278. [10.3390/e24020278](https://doi.org/10.3390/e24020278). 784  
785
- Moretti, Franco (2007). *Graphs, Maps, Trees: Abstract Models for Literary History*. Verso. 786
- Nakamura, Lisa (2013). "'Words with Friends': Socially Networked Reading on GoodReads".  
In: *PMLA* 128.1, 238–243. [10.1632/pmla.2013.128.1.238](https://doi.org/10.1632/pmla.2013.128.1.238). 787  
788
- Peer, Willie van (2008). "Ideology or aesthetic quality?" In: *The quality of literature: linguis-  
tic studies in literary evaluation*. Ed. by Willie van Peer. John Benjamins Publishing, 17–  
29. 789  
790  
791
- Pope, Colin (2019). *We Need to Talk about the Canon: Demographics in 'The Norton Anthology'*.  
[https://themillions.com/2019/04/we-need-to-talk-about-canons-picturing-  
writerly-demographics-in-the-norton-anthology-of-american-literature.ht  
ml](https://themillions.com/2019/04/we-need-to-talk-about-canons-picturing-writerly-demographics-in-the-norton-anthology-of-american-literature.html). 792  
793  
794  
795
- Porter, J.D. (2018). *Popularity/Prestige: A New Canon*. Vol. 17. Pamphlets of the Stanford  
Literary Lab. Stanford Literary Lab. [https://litlab.stanford.edu/LiteraryLab  
Pamphlet11.pdf](https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf). 796  
797  
798  
799

- Ragen, Brian Abel (1992). "An Unconventional Classic: The Politics of the "Norton Anthology"". In: *Christianity and Literature* 41.4, 471–479. <https://www.jstor.org/stable/44312103>. 799 800 801
- Reagan, Andrew J, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds (2016). "The emotional arcs of stories are dominated by six basic shapes". In: *EPJ Data Science* 5.1, 1–12. 802 803 804
- Riddell, Allen and Karina van Dalen-Oskam (2018). "Readers and their roles: Evidence from readers of contemporary fiction in the Netherlands". In: *PLOS ONE* 13.7. Ed. by K. Brad Wray, e0201157. [10.1371/journal.pone.0201157](https://doi.org/10.1371/journal.pone.0201157). 805 806 807
- Shesgreen, Sean (2009). "Canonizing the Canonizer: A Short History of The Norton Anthology of English Literature". In: *Critical Inquiry* 35.2, 293–318. <https://www.jstor.org/stable/10.1086/596644>. 808 809 810
- Trotsky, Leon (1974). *Class and Art: Problems of Culture Under the Dictatorship of the Proletariat*. New Park. 811 812
- Veleski, Stefan (2020). "Weak Negative Correlation between the Present Day Popularity and the Mean Emotional Valence of Late Victorian Novels". In: *Proceedings of the Workshop on Computational Humanities Research*. Amsterdam, the Netherlands: CEUR Workshop Proceedings, 32–43. <http://ceur-ws.org/Vol-2723/long44.pdf>. 813 814 815 816
- Vulture, editors (2018). *A Premature Attempt at the 21st Century Literary Canon*. <https://www.vulture.com/article/best-books-21st-century-so-far.html>. 817 818
- Walsh, Melanie and Maria Antoniak (2021). "The Goodreads "Classics": A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism". In: *Post45: Peer Reviewed*. <https://post45.org/2021/04/the-goodreads-classics-a-computational-study-of-readers-amazon-and-crowdsourced-amateur-criticism/>. 819 820 821 822
- Wang, Xindi, Burcu Yucesoy, Onur Varol, Tina Eliassi-Rad, and Albert-László Barabási (2019). "Success in Books: Predicting Book Sales Before Publication". In: *EPJ Data Science* 8.1, 31. [10.1140/epjds/s13688-019-0208-6](https://doi.org/10.1140/epjds/s13688-019-0208-6). 823 824 825
- Wellek, René (1972). "The Attack on Literature". In: *The American Scholar* 42.1. Publisher: The Phi Beta Kappa Society, 27–42. <https://www.jstor.org/stable/41207073>. 826 827