ZfdG

Zeitschrift für digitale Geisteswissenschaften

Fachartikel aus:
Berenike Herrmann / Maria Kraxenberger (Hg.): Weder Fail noch Lobgesang. Nichteindeutige Wertung von Literatur im digitalen Raum (= Zeitschrift für digitale Geisteswissenschaften / Sonderbände, 6). 2025. DOI: 10.17175/sb006

Titel:
The Goodreads' ›Mediocre‹: Assessing a Grey Area of Literary Judgements

Autor*in:
Pascale Feldkamp Moreira

Kontakt: pascale.moreira@cc.au.dk
Institution: Center for Humanities Computing, Aarhus University
GND: 1337465054 ORCID: 0000-0002-2434-4268
Contribution (CRediT): Writing – review & editing | Writing – original draft | Visualization | Conceptualization | Formal analysis | Investigation | Methodology | Data curation

Autor*in:
Yuri Bizzoni

Kontakt: yuri.bizzoni@cc.au.dk
Institution: Center for Humanities Computing, Aarhus University
GND: 1337465356 ORCID: 0000-0002-6981-7903
Contribution (CRediT): Writing – review & editing | Writing – original draft | Visualization | Conceptualization | Formal analysis | Investigation | Methodology | Validation

Autor*in:
Mia Jacobsen

Kontakt: miaj@cas.au.dk
Institution: Center for Humanities Computing, Aarhus University
GND: 135043678X ORCID: 0009-0003-3720-3418
Contribution (CRediT): Data curation | Conceptualization | Writing – review & editing

Autor*in:
Mads Rosendahl Thomsen

Kontakt: madsrt@cc.au.dk
Institution: Comparative Literature – Institute for Culture and Communication, Aarhus University
GND: 142913014 ORCID: 0000-0002-4975-6752
Contribution (CRediT): Validation | Resources | Supervision | Project administration | Funding acquisition

Autor*in:
Kristoffer L. Nielbo

Kontakt: kln@cas.au.dk
Institution: Center for Humanities Computing, Aarhus University
GND: 1100316361 ORCID: 0000-0002-5116-5070
Contribution (CRediT): Validation | Methodology | Resources | Supervision | Project administration | Funding acquisition

DOI des Beitrags:
10.17175/sb006_002

Nachweis im OPAC der Herzog August Bibliothek:
1928557961

Erstveröffentlichung:
20.11.2025

Letzte Überprüfung aller Verweise:
09.04.2025

Format:
PDF ohne Paginierung, Lesefassung

GND-Verschlagwortung:
Literarische Wertung | Literaturrezeption | Mittelmäßigkeit | Online-Community | Literaturwissenschaft

Empfohlene Zitierweise:
Pascale Feldkamp Moreira / Yuri Bizzoni / Mia Jacobsen / Mads Rosendahl Thomsen / Kristoffer L. Nielbo: The Goodreads' ›Mediocre‹: Assessing a Grey Area of Literary Judgements. In: Berenike Herrmann / Maria Kraxenberger (Hg.): Weder Fail noch Lobgesang. Nichteindeutige Wertung von Literatur im digitalen Raum (= Zeitschrift für digitale Geisteswissenschaften / Sonderbände, 6). Wolfenbüttel 2025. 20.11.2025. HTML / XML / PDF. DOI: 10.17175/sb006_002

Pascale Feldkamp Moreira / Yuri Bizzoni / Mia Jacobsen / Mads Rosendahl Thomsen / Kristoffer L. Nielbo

The Goodreads' ›Mediocre‹: Assessing a Grey Area of Literary Judgements

Abstract

Computational studies of literature have embraced statistical and social science methods, enabling studies that estimate the success of literary texts using proxies of literary success or reader appreciation. Among such proxies, Goodreads is a particularly popular resource, as it aggregates readers' opinions into one quantifiable scale. In predicting literary success, studies tend to focus on titles that are considered the ›very best‹, often compared to the ›very worst‹. However, ›mediocre‹ values in proxies of literary quality such as the rating on Goodreads, represents a complex grey area. As average ratings do not indicate unanimity, middle values of average rating might obscure a polarised rating behaviour. The question that emerges pertains to the nature of such ›mediocre‹ ratings: do they represent a simply tepid reception of some titles, or embody a readership divided into polarised factions?

To interrogate the nuanced nature of ›mediocre‹ ratings, we conducted an empirical analysis on a dataset drawn from the Chicago corpus, which comprises 9,000 novels published in the United States between the years 1880 and 2000. From this corpus, we extracted a subset of 2,150 novels that occupy the middle quartile of Goodreads average ratings, specifically those falling within a range of 3.72 to 3.91 in the Goodreads' rating scale (1–5). To gauge the presence of works that may be classified as successful, prestigious, or canonical within this subset, we employed additional proxies of literary appreciation for cross-validation. This multi-dimensional approach aims to shed light on whether these middle-rated works signify a lukewarm collective reception on Goodreads, or if they mask a rating behaviour that is, in fact, polarised. The results of this investigation aim at an enhanced understanding of the complexities inherent in quantified assessments of literary quality. Our empirical analysis reveals that approximately 30 % of the designated ›mediocre‹ category overlaps with alternative metrics of literary quality – that is, novels in the mediocre category are indexed in other proxies of literary excellency – while also manifesting a statistically significant higher rating count on Goodreads compared to the remaining titles in the category.

Our data suggests that this ›mediocre‹ cohort can be taxonomically classified into three distinct subgroups: i) titles with fewer ratings that garner tepid evaluative responses, ii) titles that are controversial, displaying divergent evaluations between Goodreads and other literary quality proxies, and iii) titles that provoke highly polarised opinions, manifesting substantial divergence in rating distributions not only among other proxies, but on the Goodreads platform itself. Intriguingly, we observed a positive correlation between rating count and the standard deviation of rating distribution for a subset of quality titles with high rating counts, a pattern conspicuously absent in the non-quality cohort. Our observations on the Goodreads' ›mediocre‹ underscore the exigency of a more nuanced, perspectivist methodology in employing proxies for literary quality. Such an approach would provide a more robust framework for predicting and modelling reader appreciation in a manner that accommodates its inherent complexities.

Die computergestützte Literaturwissenschaft hat sich statistische und sozialwissenschaftliche Methoden zu eigen gemacht, die es ermöglichen, den Erfolg literarischer Texte anhand von Näherungswerten für den literarischen Erfolg oder die Wertschätzung der Leser*innen zu ermitteln. Unter diesen Proxies ist Goodreads eine besonders beliebte Ressource, da es die Meinungen der Leser*innen in einer quantifizierbaren Skala zusammenfasst. Bei der Vorhersage des literarischen Erfolgs konzentrieren sich Studien in der Regel auf die ›besten‹ Titel, die oft mit den ›schlechtesten‹ verglichen werden. Die mittelmäßigen Werte in Proxies für literarische Qualität, wie die Bewertungen auf Goodreads, stellen jedoch eine komplexe Grauzone dar. Da durchschnittliche Bewertungen nicht auf Einstimmigkeit hindeuten, können mittlere Werte der Durchschnittsbewertung ein polarisiertes Bewertungsverhalten verschleiern. Die Frage, die sich stellt, bezieht sich auf die Natur solcher ›mittelmäßigen‹ Bewertungen: Repräsentieren sie einfach eine laue Rezeption einiger Titel oder stellen sie die Meinung einer Leserschaft dar, die in polarisierte Fraktionen gespalten ist?

Um den Charakter ›mittelmäßiger‹ Bewertungen zu untersuchen, haben wir eine empirische Analyse eines Datensatzes aus dem Chicago-Korpus durchgeführt, welcher 9.000 Romane umfasst, die zwischen 1880 und 2000 in den Vereinigten Staaten veröffentlicht wurden. Aus diesem Korpus haben wir eine Teilmenge von 2.150 Romanen extrahiert, die sich im mittleren Quartil der Goodreads-Durchschnittsbewertungen befinden, d. h. im Bereich von 3,72 bis 3,91 in der Goodreads-Bewertung (Skala 1–5). Um das Vorhandensein von Werken innerhalb dieser Untergruppe zu beurteilen, die als erfolgreich, prestigeträchtig oder kanonisch eingestuft werden können, haben wir zusätzliche Indikatoren für die literarische Wertschätzung zur Kreuzvalidierung verwendet. Dieser mehrdimensionale Ansatz soll Aufschluss darüber geben, ob die mittelmäßig bewerteten Werke für eine laue kollektive Rezeption auf Goodreads stehen, oder ob sie ein tatsächlich polarisiertes Bewertungsverhalten verdecken. Die Ergebnisse dieser Untersuchung zielen auf ein besseres Verständnis der Komplexität von quantifizierten Bewertungen literarischer Qualität ab. Unsere empirische Analyse zeigt, dass etwa 30 % der Romane in der Kategorie ›mittelmäßig‹ Überschneidungen mit alternativen Metriken für literarische Qualität aufweisen – d. h. sie finden sich auch in anderen Proxies für literarische Qualität – und weisen zudem eine statistisch signifikant höhere Bewertungszahl auf Goodreads auf als die übrigen Titel derselben Kategorie.

Unsere Daten deuten darauf hin, dass diese ›mittelmäßige‹ Kohorte taxonomisch in drei verschiedene Untergruppen eingeteilt werden kann: i) Titel mit wenigen Bewertungen, die laue Reaktionen hervorrufen, ii) Titel, die divergierende Bewertungen zwischen Goodreads und anderen literarischen Qualität Proxies aufweisen, und iii) Titel, die stark polarisierte Meinungen hervorrufen und erhebliche Divergenzen in der Bewertungsverteilung nicht nur zu anderen Proxies, sondern auch auf der Goodreads-Plattform selbst aufweisen. Interessanterweise

beobachteten wir eine positive Korrelation zwischen der Anzahl der Bewertungen und der Standardabweichung der Bewertungsverteilung für eine Untergruppe mit einer hohen Anzahl von Bewertungen, ein Muster, das in der ›Nicht-Qualitätskohorte‹ auffällig fehlt. Unsere Beobachtungen zum Goodreads-›Mittelmaß‹ unterstreichen die Notwendigkeit einer nuancierten, perspektivischen Methodik bei der Verwendung von Ersatzwerten für literarische Qualität. Ein solcher Ansatz bietet einen soliden Rahmen für die Vorhersage und Modellierung der Wertschätzung von Leser*innen und der damit verbundenen Komplexität.

1. Introduction

The fields of computational linguistics and computational literary studies have in recent years converged on the development of a host of powerful new methods and tools for analysing text on a large scale. While computational linguistics and Natural Language Processing (NLP) research has increasingly researched literary corpora and modelled the way texts impact readers, computational studies of literature have embraced computational linguistics and NLP methods, allowing for the growth of more complex quantitative literary research. Since literary studies deal with social and cultural elements inherent to literary, reading, and book history as well as canon-formation, quantitative approaches have also borrowed systems from computational sociology, statistical and social science methods, enabling literary research on a scale that would be unfeasible ›by hand‹. [1]

Beyond upping the scale, computational methods have also made a shift in the object of literary studies research: from studies in literary history predominantly tracing the trajectories and influences of canonical works and authors, toward bringing into focus what Moretti has called ›the great unread‹: the forgotten 99 % of literary history.[1] Since no one can access (let alone process) everything that has ever been published, quantitative literary studies apply theoretical frameworks of corpus linguistics to try to create limited collections that are representative, i. e., reasonably mirror the whole population on a smaller scale. In traditional corpus linguistics, for example, a large diachronic collection of documents from different written and spoken domains could be considered a reasonable representation of a language in a given time span, so that the spreading of a new term at a given point might be reasonably considered to mirror the spreading of that word in the real world. [2]

A similar approach may be taken to the study of literature, where one may track the emergence and spread of concepts or stylistic devices across corpora,[2] and even study the popularity of a given work with readers, comparing its reception to that of other works in the corpus, extrapolating and correlating popularity to, for example, particular textual features. Still, representativeness is paramount: if that corpus consists only of bestselling sci-fi, we have gained insight only into what features contribute to popularity in this given niche population. [3]

While the ability to process and analyse large quantities of literary texts and perform complex statistical experiments on them has recently made new ways of studying literary appreciation possible, the question of how to assess (and recreate) literary success is probably as old as narrative itself. The question of quality has naturally been prominent in literary criticism, but its significance has often been eclipsed in the scholarly discussion. Disciplinary shifts, such as the debate on canon representativity and exclusion,[3] methodological shifts, such as moving the focus from evaluation towards interpretation in 20th century literary criticism,[4] as well as a modern expansion of literature's conceptual boundaries to encompass texts with new experimental forms (i. a., hypertext fiction), or such that are ideologically opposed to the notion of ›pleasing‹ the reader,[5] have all played a role in making words like ›literary quality‹, or ›classics‹ appear to belong to the »precritical [4]

[1] Moretti 2000.
[2] Jockers 2013; Moretti 2007.
[3] Van Peer (ed.) 2008.
[4] Hagen et al. 2018.
[5] Wellek 1972.

era of criticism itself«.[6] Nevertheless, literary cultures continue to establish and uphold estimations of literary excellence in practice, such as through literary awards, classics book series, or anthologies. It seems a disparity has arisen between a more modern literary studies inattention to, or »denial of quality«,[7] and the multitude of literary judgements prevailing within literary cultures – a disparity that is brought to a point in the context of more recent computational literary studies which show a consensus on quality judgements among readers at the scale of large numbers.[8]

[5] Most studies of literary success or quality have attempted to test links between reader appreciation and textual features: both stylistic features,[9] and more narrative characteristics, such as shape and dynamic of narratives' sentiment arcs extracted by sentiment analysis.[10] In one sense, traditional and computational literary studies have similar tendencies: like traditional literary scholarship, computational studies of literature often focus on ›the excellent few‹ – the ›exceptional books‹ that do well among readers[11] – and seek to understand their exceptionality on the textual level.

[6] Computational studies that seek to predict reader appreciation often set up the task as one of classification, dividing their corpus into ›successful‹ and ›unsuccessful‹ titles, using a threshold value of some proxy of reader appreciation, like the number of downloads on project Gutenberg.[12] Others, who seek to model the continuous scale of ›more or less successful‹ titles, often focus on discussing and contrasting the textual characteristics of the ›very good‹ against the ›very bad‹.[13] However, few studies take up the discussion of what is often a very large part of any corpus: the mediocrely estimated titles, neither decidedly good nor bad. This group is also important when seeking to assess the different ways literary cultures judge works. We would assume that the titles that are considered ›great‹ or ›bad‹ vary across proxies of literary quality. For example, the titles on bestseller lists are not necessarily those most often assigned in college syllabi. It should therefore not be a surprise if what is considered ›mediocre‹ varies greatly across different standards and tastes. As such, what is considered ›mediocre‹ may vary depending on what ›quality proxy‹ studies use to approximate the notion of ›literary success‹ or ›quality‹.

[7] 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, and are a translation of constructs, like a literary device, or concepts, like literary quality, into measurable variables. When speaking of a ›quality proxy‹, we mean a specific operationalization of quality or reader appreciation among many. For example, one might differentiate between literary ›fame‹ and ›popularity‹, since fame, such as that of James Joyce's *Ulysses* does not necessarily mean that the book is widely read. These two 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.[14] Consequently, the construction of proxies also assists in compelling us to precisely define more ambiguous concepts like ›quality‹.

[6] Guillory 1991, p. 36.
[7] Wellek 1972, p. 37. This does not mean that there is no ›perception of quality‹ within literary scholarship itself. Studies such as Porter (2018) have made distinct scholarly canons apprehensive by also looking at the con- and divergences between titles that receive scholarly attention and those popular on sites like Goodreads.
[8] Archer / Jockers 2017; Feldkamp et al. 2024; Maharjan et al. 2017; Xindi Wang et al. 2019; Porter 2018.
[9] Van Cranenburgh 2016; Crosbie et al. 2013; Ganjigunte Ashok et al. 2013; Koolen et al. 2020; Maharjan et al. 2017.
[10] Bizzoni et al. 2023b; Maharjan et al. 2018; Reagan et al. 2016.
[11] Kovács / Sharkey 2014; Maity et al. 2019; Manshel et al. 2019; Walsh / Antoniak 2021.
[12] Ganjigunte Ashok et al. 2013.
[13] Bizzoni et al. 2023a, Bizzoni et al. 2023c.
[14] At the time of writing this study, *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*, which have above 8 million ratings and average ratings above 4.3.

[8] For studies seeking to assess factors contributing to reader appreciation, selecting a resource to use as such a proxy for reader appreciation or literary success is one of the great challenges, if not the greatest. Studies often use a single proxy for literary quality as a golden standard, which may not adequately capture the diverse – and, sometimes, opposed – preferences of various types of audiences.[15] Among various proxies for quality, such as a text's presence in established literary canons[16], whether or not it was longlisted for awards,[17] sales numbers,[18] ratings on the large online platform Goodreads are widely used.[19] Specifically, the average ratings or stars of Goodreads are the most widely used, although Goodreads actually has two dimensions: stars representing various types of literary evaluations reduced to the 1–5 point ›stars‹ scale, and the rating count, representing how often a title is rated, indicating its circulation or fame among platform users.

[9] In the present paper we suggest that a closer inspection of ›mediocrely‹ or ›indecisively‹ evaluated texts may help studies to better assess the corpus, as well as the proxies of reader appreciation themselves, as it allows for a better understanding of the judgement behaviour and nature of the resource. Since it is the most widely used resource, we examine Goodreads ratings and rating counts as a proxy for literary ›mediocrity‹, specifically taking the ›mediocre‹ Goodreads' ratings as an ideal subset for exploring the evaluative process within the proxy. Our question pertains to the nature of Goodreads mediocre ratings: do they signify a simply ›lukewarm‹ or tepid reception from readers, or are they the result of a readership divided into polarised opinions about specific titles? We describe this ›mediocrely rated‹ Goodreads subset in terms of how the titles are evaluated outside of Goodreads, and examine rating behaviour, specifically the distribution of raters' evaluations on the 1–5 point scale, to explore the distinct types of literary ›mediocrity‹ formed on the Goodreads platform.

2. Goodreads as a proxy for ›quality‹

[10] Goodreads is a popular online social platform for readers that allows users, among other things, to comment, recommend, and review a book. It is, according to Nakamura (2013), a ›social cataloguing site‹, which links to other social networks (Facebook, Twitter / X, Instagram, and LinkedIn), and where social networking may, for some users, be just as important as book cataloguing and reviewing.[20] With its more than 90 million users, Goodreads arguably offers an insight into reading culture ›in the wild‹.[21] 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.[22] Whether it may be followed that sites like Goodreads represent a ›democratisation‹ of literary criticism may, however, be questioned, not least because we see the continuation of established patterns on the platform. For example, works that are often assigned on college syllabi are also perceived as ›canonic‹ or ›classic‹ on Goodreads.[23] Moreover, while the site had mostly anglophone users at the outset – which was presumably reflected in the type of books rated (and highly rated) – the rapid expansion of the site makes it difficult to gauge user demographics, though it structurally must be presumed to reflect certain tilts, i. e., preferences of readers that review books online vs. those that do not.

[15] Bizzoni et al. 2023b; Manshel et al. 2019; Porter 2018.
[16] Mohseni et al. 2022.
[17] Bizzoni et al. 2023b.
[18] Archer / Jockers 2017; Xindi Wang et al. 2019.
[19] Bizzoni et al. 2023a; Jannatus Saba et al. 2021; Maharjan et al. 2017.
[20] Nakamura 2013; Thelwall / Kousha 2017.
[21] Nakamura 2013.
[22] Thelwall / Kousha 2017; Walsh / Antoniak 2021.
[23] Steiner 2008; Walsh / Antoniak 2021.

[11] Moreover, problems and pitfalls with Goodreads data should be acknowledged, such as the existence of fake or sponsored reviewing, influences of Goodreads algorithms on users' reading and reviewing behaviour, and the general anglophone and Western literary tilt and biases of the platform.[24] Still, some aspects of Goodreads reviewer behaviour support the idea that Goodreads represents an alternative resource for studying literary evaluation next to newspaper and scholarly literary criticism. For example, Verboord (2011) has shown that Goodreads reviews distribute more equally across genres and exhibit more attention to mass-market paperbacks and titles by female authors than newspaper reviews.[25] While Goodreads ratings and rating count do not present an absolute measure of literary popularity or quality – or in the context of the present study, literary ›mediocrity‹ – they do offer a valuable perspective on a title's overall reception among a diverse population of readers with preferences that vary from expert and newspaper critics. In other words, we should remember that Goodreads reflects preferences of a certain audience however heterogenous the user demographic – it is one ›quality proxy‹ among many.

[12] Despite their diversity, different proxies of quality may display significant overlaps.[26] Moreover, if a book comes to be included in one proxy, this might correlate to an increase in another: for example, studies have shown that winning an award may mean that a title gains in scholarly prestige or popularity.[27] Similarly, Kovács and Sharkey (2014) found that winning an award corresponds to an initial increase in book ratings on Goodreads, however, a winning books' average rating tends to decline relative to books that were long-listed for an award but did not win.[28] They suggest that the decline in average rating is partly an effect of higher popularity, whereby a title also attracts readers that are predisposed to *dislike* it. We may presume then that a higher number of ratings could correlate with a more polarised rating behaviour (i. e., a higher number of very low ratings).

[13] Goodreads' average ratings represent the averaged ratings of all users for a title on a Likert scale implemented as the number of stars on the platform interface, ranging from 1 (low appreciation) to 5 stars (high appreciation). While the average score provides a general indication of a title's reception, it is problematic because it conflates types of literary appreciation, i. a., satisfaction, enjoyment, and evaluation to one scale, forcing users to cast their literary evaluations to a mono-dimensional scale. Moreover, averaging Likert scale data entails a conversion of ordinal to interval data, reducing uneven judgements to one seemingly unambiguous score, which is a well-known issue.[29] The averaged score potentially conflates genre-specific value judgements, which may obscure important differences in rating behaviour across audiences. For example, readers of sci-fi may be inclined to give a generally higher average rating on Goodreads, something that we do not take into account when using average rating as a quality metric. Additionally, the average score may be the result of numerous middle-ratings or many ratings at either of the extremes. The simplicity of the Goodreads resource is also why it is often employed in computational studies seeking to assess literary quality or success: it offers a streamlined approach to a problem that frequently proves too complex for quantitative analysis. While the Goodreads proxy thus has the benefit of aggregating readers' opinions into one quantifiable scale, it also creates a grey area in the middle. Mediocre ratings on Goodreads may be averages of numerous extreme positive and negative ratings,[30] which poses challenges for researchers both when interpreting and trying to build systems to predict ratings, especially when they only have access to the average ratings without information on rating distribution.

[24] Walsh / Antoniak 2021.
[25] Verboord 2011.
[26] Manshel et al. 2019; Walsh / Antoniak 2021.
[27] Manshel et al. 2019.
[28] Kovács / Sharkey 2014.
[29] Bishop / Herron 2015.
[30] Kai Wang et al. 2019.

3. Data

3.1 The Chicago Corpus

As the present study is part of a larger research inquiry into modelling literary qualities, we specifically chose to work with *The Chicago Corpus* for its accessibility of full texts, thereby enabling future textual analytics within the broader scope of the ongoing research. [14]

	No. of titles	No. of authors	No. of q. titles	Avg. GR rating	Avg. GR rating count
All	8991	3124	2414	3.77	14365.27
Subset	2150	1142	678	3.82	9638.1
Q. titles	678	351		3.82	26767.76

Table 1: Overview of titles in the corpus. Note that the subset indicates titles in the mediocre subset, i. e., titles that were given a mediocre rating (see section 3.2), and Q. titles indicates titles included in some quality proxy outside of Goodreads (see section 3.3).

The Chicago Corpus spans close to 9,000 novels published in the US (1880–2000), and is a unique corpus for computational analysis both in terms of size and diversity.[31] The corpus was compiled on the basis of how many libraries hold a copy of a novel worldwide, the selection preferring more circulated works in terms of library holdings. As such, it is a valuable resource, encompassing an expansive and representative sample of the (mainly) Anglophone literary scene over a century, spanning from mass market fiction (i. a., mystery and detective novels and series) to works by authors who received the Nobel Prize and other distinctions, such as the National Book Award (i. a., Don DeLillo, Joyce Carol Oates, and Philip Roth). Within this low/high-brow divide, we also find important works of genre-fiction (i. a., by J. R. R. Tolkien or Philip K. Dick). Moreover, even if all 9,000 novels sustain high enough library holdings numbers to be included in the corpus, the corpus contains a large portion of what we may see as more obscure and forgotten novels if we compare to how known the books are on, for example, Goodreads. 2,192 titles in the corpus (or ca. 25 % of the corpus) have a Goodreads rating count below 50, and 813 titles (or 9 % of the corpus) have a rating count below 10. Nevertheless, with their high library holding numbers, these titles did enjoy a high circulation and were in relatively high demand in libraries and among library-goers at some point in time. [15]

It is also worth noting that the corpus has an Anglophone bias, which inevitably situates the entire analysis within the context of Anglophone and Western literary culture. While this does not inherently undermine our analysis, it is crucial to consider when interpreting the results, and caution should be exercised when extrapolating the findings to the context of a global or another literary field. [16]

3.2 The Goodreads ›mediocre‹ subset

To analyze ›mediocre‹ Goodreads ratings, we define a ›mediocre group‹ within the Chicago corpus as titles that constitute the middle 25 % based on average Goodreads ratings. In this context, ›mediocre‹ does not suggest poor quality or a title's standing on the Goodreads platform, but rather reflects the average rating compared to the entire corpus. To distinguish the middle 25 % of our corpus, we distinguished titles that lie between the 37.5th and 62.5th percentile of the distribution of average Goodreads ratings in our corpus (see fig. 1). These percentiles divide the corpus distribution into parts, with 25 % of the corpus (the ›mediocre‹ group) occupying the middle segment (between the percentiles): 37.5 % of the corpus falls below it and 37.5 % falls above it. The middle segment comprises titles representing the middle 25 % of the corpus in [17]

[31] Generally, studies on literary quality have relied on corpora of < 1,000 books (i. a., Ganjigunte Ashok et al. 2013; Koolen et al. 2020).

terms of their Goodreads ratings. We thus compile a ›mediocre‹ subset using the titles representing the middle 25 % of the corpus in terms of their Goodreads ratings, which is equivalent to having a Goodreads' average rating between 3.72 and 3.91. This ›mediocre group‹ comprises 2,150 titles in total.[32]

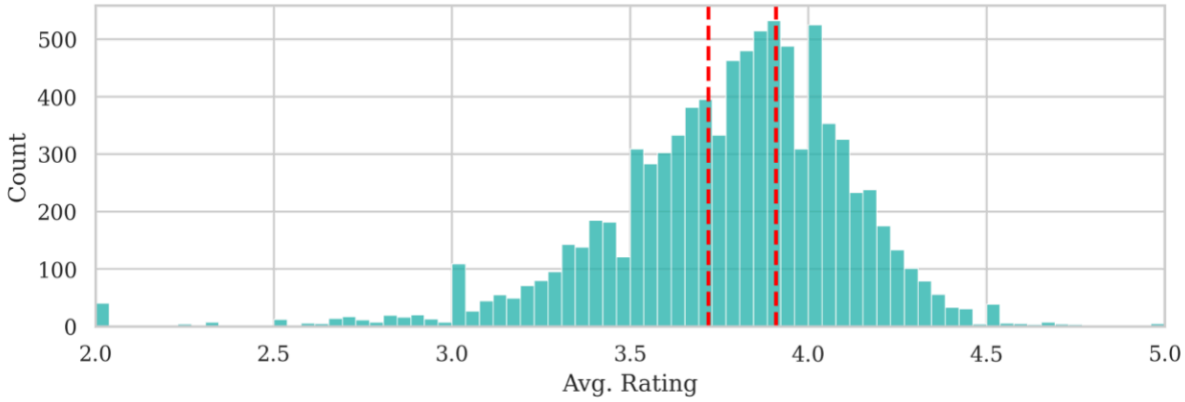

Figure 1: Histogram of Goodreads ratings of the Chicago Corpus, where the ›mediocre‹ subset is indicated. [Chart: Feldkamp et al. 2025]

3.3 Other quality proxies

To estimate the number of books within this subset of books, we collected various other proxies of literary appreciation to distinguish ›high quality‹ titles within our ›mediocre‹ subset. These are generally distinguished by type of proxy – top-down, bottom-up or in-between – and further by type of affiliation of agents that define the proxy: institutional, intellectual or commercial. These affiliation-categories should not be considered clearly distinct or mutually exclusive, but as an aid for heuristically conceptualising the proxies. Editors of the Norton Anthology, for instance, may rely on their (intellectual and / or institutional) expertise of literature but also consider (commercial) popular demand, but especially academic (institutional) demand – for example, what literature is often included in college syllabi and what should thus be extant in an anthology that is frequently used by students and in classrooms.[33] [18]

As we sought to consider a wide array of proxy types to include as many different perspectives on what is considered quality as possible, we selected specific proxies of quality within the three overarching categories: predominantly intellectual (e. g. longlists for literary awards), institutional / canon-making (e. g. OpenSyllabus, which collects college syllabi), commercial (e. g. bestsellers), and measures that fall in-between these categories, such as number of translations of a work registered in UNESCO's *Index Translationum*. Which books are translated is a complex issue not clearly connected exclusively to market forces nor expert judgements. As these latter, in-between measures are not binary (either you are longlisted or not) but continuous (1, 2, 3, etc. translations), we included titles that were among the 10 % top-scoring in terms of these three continuous proxies into our ›high quality‹ group. This ›quality group‹ within our ›mediocre‹ subset comprises 678 titles. [19]

[32] We use the average rating and rating count for our corpus that were collected in December 2022.
[33] Ragen 1992.

Quality proxy type		**Number**
Miscellaneous		
Library Holdings (top 10 %)		900
Translations (top 10 %)		968
Author-page Rank (top 10 %)*		909
Canon		
Opensyllabus*		479
Norton Anthology*		401
Penguin Classics series		76
Goodreads' Classics*		62
Awards	**Specific award**	
General, Literary Awards	- The National Book Award - The Nobel Prize - The Pulitzer Prize	247
Sci-fi Awards	- The Hugo Awards - J. W. Campbell Award - Locus Sci-fi Award - The Nebula Awards - Philip K. Dick Award - Prometheus Award	180
Fantasy Awards	- British Fantasy Award - Locus Fantasy Award - Mythopoeic Award - World Fantasy Award	40
Horror Awards	- Bram Stoker Award - Locus Horror Award	19
Mystery Awards	The Edgar Awards	10
Romantic Awards*	- Rita Awards - RNa Awards	54
Commercial		
Publisher's Weekly Bestsellers		139

Table 2: Proxies used for assessing ›quality‹ titles in our subset. *These proxies are author-based due to the scarcity of data or nature of the proxy (e. g. author-page rank or the Nobel Prize in Literature).

4. Method

As we seek to gain a better insight into types of titles, rater behaviour and judgements of the ›mediocre‹ Goodreads subset, we apply measures that can reasonably be expected to differentiate groups within the subset. These are, firstly, external proxies of literary quality which indicate whether a title is highly appreciated outside the Goodreads platform (table 2). Secondly, with the Goodreads data, we examine whether these ›quality titles‹ within the ›mediocre‹ subset differ from other titles in terms of rating count, examining whether the ›quality‹ titles are more popular (i. e., more frequently rated) in comparison to the [20]

›non-quality‹ titles in our subset. Thirdly, we examine the distribution of ratings of the ›quality‹ titles in our subset, to assess whether some titles in the subset are subject to different rating behaviour and, specifically, whether titles that are ›quality‹ titles judged ›quality titles‹ by external proxies (outside of Goodreads) are subject to a polarised rating behaviour within the Goodreads platform itself.

[21] To assess the polarisation of the raters of individual titles, we calculated the *standard deviation* (SD) of their rating distribution, based on the assumption that fewer ratings around the mean indicates more polarized rater behaviour. Moreover, following Maity et al. (2019), we calculate the *Shannon entropy* of rating distributions as a proxy measure to assess the polarisation of raters.[34] If values are more ›polarized‹, they should also be more ›spread out‹ on the rating scale (1–5), which entropy should capture.

[22] To inspect how SD captures polarisation of reception of titles – essentially a ›sanity check‹ of our measures – we plotted the distribution of ratings of individual titles. For assessing and visualising, we keep a threshold of rating count (10,000). Firstly, because very low rating count may exhibit a ›polarised‹ readership because judgements have not converged at the larger scale or random patterns not corroborated by other raters. Secondly, to visualise more recognizable titles for the sake of our readers. Generally, we found both SD and entropy to distinguish the same titles at the highest and lowest end of the spectrum: the list and order of titles on the low and high end of SD are almost identical to those at the low and high end of rating entropy.

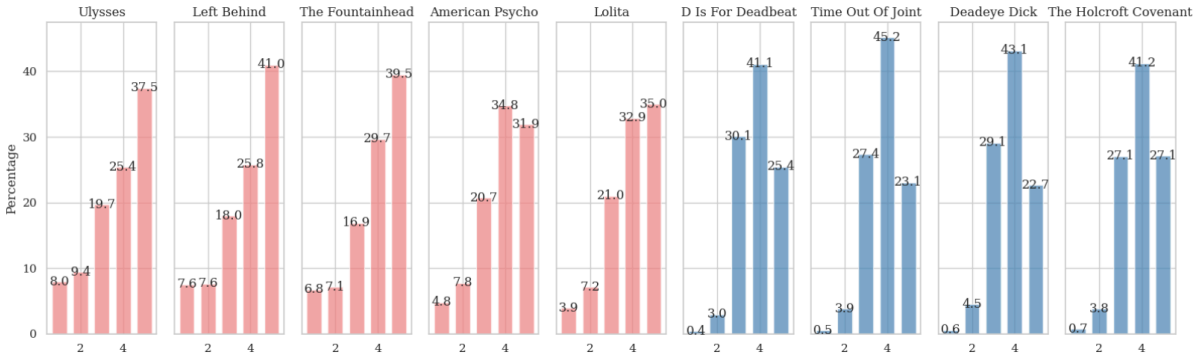

Figure 2: Titles among top 20 highest in SD to the left (SD descending) and titles among the 20 lowest in SD to the right (SD ascending). Note that titles to the right (blue) seem to have a slightly more ›normal‹ distribution, while rating distributions of titles to the left (pink) have more of a U-shape, with more low ratings (of 1 and 2) and very high ratings (of 5). [Chart: Feldkamp et al. 2025]

[23] From this inspection, it is clear that SD and entropy of rating distribution can adequately capture how controversial titles are: titles on the high end of SD and entropy (fig. 2, left side) are known controversial titles. Either because they are stylistically demanding (e. g., James Joyce's *Ulysses*), or controversial in terms of the themes and worldview they treat (e. g., Vladimir Nabokov's *Lolita* or Tim LaHaye and Jerry B. Jenkins's *Left Behind*). Titles at the low end of these measures (fig. 2, right side) are more mainstream literature: widely read and not known to be very debated. In fact, 7 out of the 20 titles lowest in SD are novels by Agatha Christie (note that we are still looking at titles above the set threshold in rating count, i. e., above 10,000 ratings).

[24] Moreover, testing the collinearity of standard deviation and entropy by using a Spearman correlation, we found that these two measures are very highly correlated in our subset (coefficient = 0.9, $p < .01$). As such, we focus on reporting values for SD in the following, as values for entropy are very similar.

[34] Maity et al. 2019. We calculated entropy using the Neurokit package (Makowski et al. 2021): https://neuropsychology.github.io/NeuroKit/functions/complexity.html#entropy-shannon

5. Results

5.1 Rating count of the quality / non-quality groups

In comparing titles in the ›mediocre‹ subset that appear in other proxies of literary quality (i. e., one of the proxies in table 2), we find that there is a difference in the frequency that titles in these groups are rated (fig. 3, 4, see also table 1), even if there is no difference in average rating between the groups (fig. 3, table 1). [25]

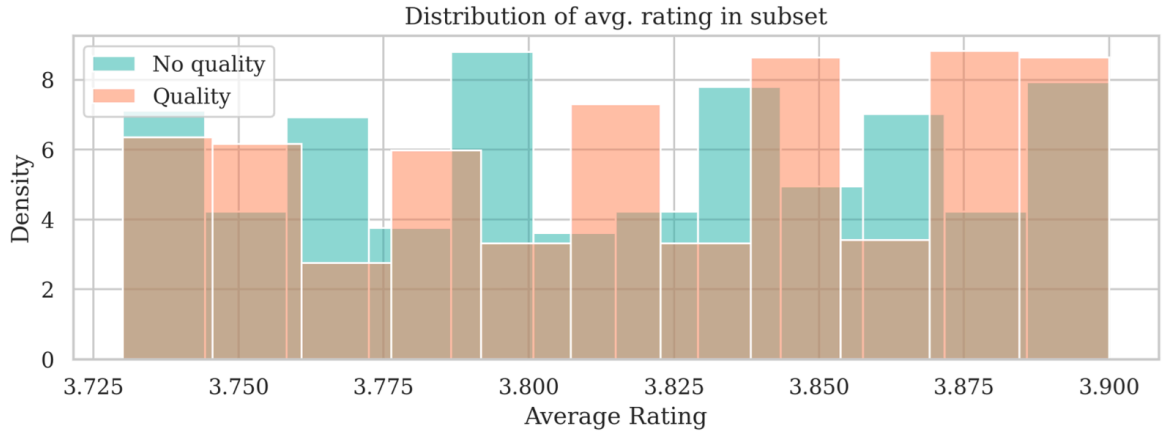

Figure 3: Histogram of Goodreads rating in ›mediocre‹ subset. [Chart: Feldkamp et al. 2025]

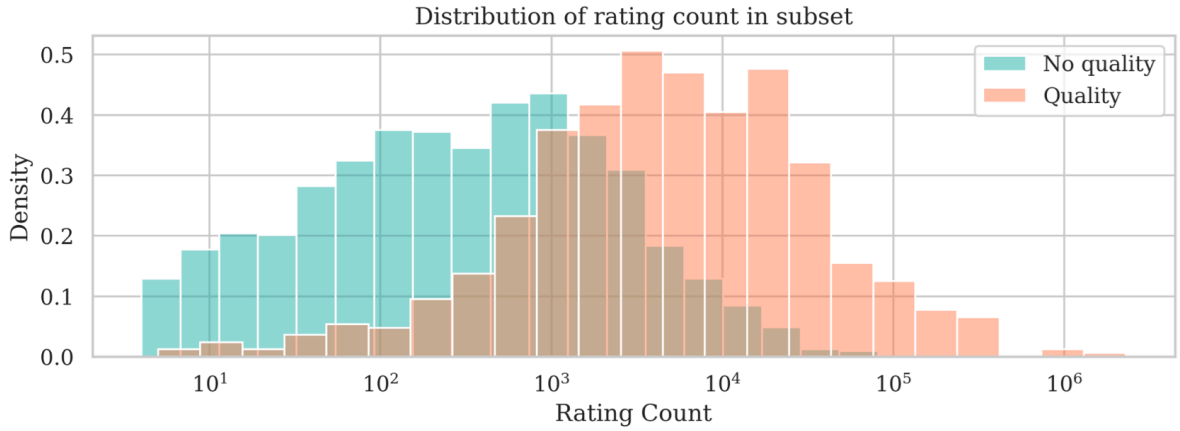

Figure 4: Histogram of Goodreads rating count in ›mediocre‹ subset. [Chart: Feldkamp et al. 2025]

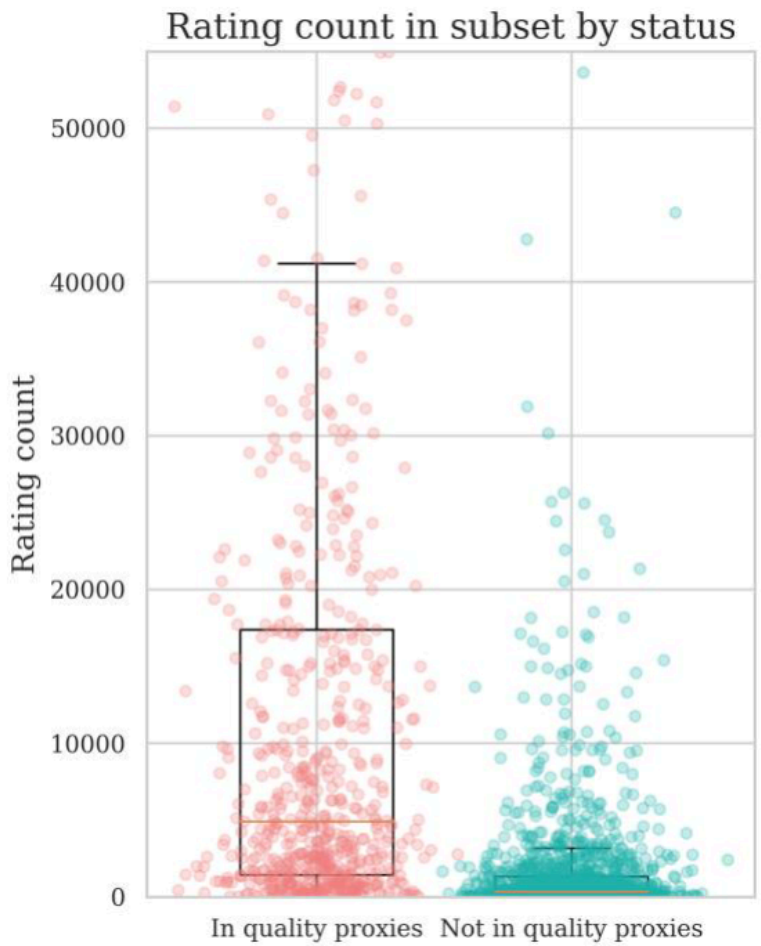

Figure 5: Boxplot showing the rating count per group. [Chart: Feldkamp et al. 2025]

	Avg. Rating count	**Std. deviation**
All	14365.27	3121538.20
Subset	9638.1	163201.66
Q. titles	26767.76	110124.60
Non-q. titles	1680.0	4685.72

Table 3: Rating count and standard deviation for the whole corpus, the ›mediocre‹ subset, as well as the quality and non-quality groups.

We find that the mean rating count of the ›quality group‹ in our subset is higher than the non-quality group, but it is also higher than the mean rating count in the whole corpus. Note that the standard deviation is, however, extremely high, indicating that the ›quality‹ group is by no means homogenous. This would suggest that the quality group contains several titles that are rated extremely often, even when compared to the ratings in the whole corpus. In this sense, several works that are considered of ›quality‹ outside of the platform Goodreads seem to garner more attention than those that are not. [26]

5.2 Comparison of rating distribution polarisation per group

Since our group of quality titles is contested in the sense that they are highly esteemed outside, but not on Goodreads (in terms of average rating), it is reasonable to hypothesise that their reception on Goodreads is not exclusively ›lukewarm‹, but that certain readers would rate them higher than others. Moreover, the [27]

higher overall levels in rating count of the quality group may also be connected to a more polarised reception as suggested in Kovács and Sharkey (2014).[35] The question is therefore whether the polarisation in the reception of these titles between Goodreads and the other proxies of quality is repeated on the Goodreads platform itself, in other words, are the Goodreads' user ratings of these titles polarised?

[28] Therefore, we inspected the distribution of ratings across the scale 1–5 stars for each title, through which we examine whether ›quality titles‹ are more ›polarised‹, i. e., whether they are rated more on the extremes of the scales than the non-quality titles are. To assess the ›polarisation‹ of these distributions, we measured the standard deviation and entropy of title's rating distributions. In general, we find that there does not seem to be a measurable difference between the quality and non-quality group in our subset (table 4). Moreover, a Spearman correlation shows weak to nonsignificant correlations between rating count and standard deviation of rating distribution in both the quality and non-quality group in our subset (table 5).

	SD mean	**SD of dist. SD**	**Entropy mean**	**Entropy SD**
Q. titles	0.958	0.069	1.889	0.077
Non-Q. titles	0.940	0.104	1.840	0.150

Table 4: SD and entropy of rating count distributions for each group, with the standard deviation reported.

[29] Note that the distribution entropy is at the same levels as in Maity et al.'s (2019) examination of the Goodreads' rating distribution of 558,563 books, where they found entropy peaking around 1.7–1.9.[36] As such, these are very common values even when looking at a bigger corpus and across the whole scale of Goodreads ratings.

5.3 Quality titles at high and low rating count

[30] In figure 2, we saw that standard deviation clearly indicates works in the quality group that have a polarised distribution. As in the case of rating count, however, standard deviation of rating distributions seems to be very heterogeneous within the quality group to the point where variance makes it difficult to distinguish from the non-quality group. Following the suggestion of Kovács and Sharkey (2014) that higher rating count is linked to higher polarisation of the raters,[37] we examine marginally more rated titles within the quality group itself. For this, we split the quality group into a group of titles with high and one with low rating count, selecting the subset mean rating count as a somewhat arbitrary threshold. The aim here is not to distinguish clear groups, but simply to examine the characteristics of some portion of quality titles that have a very high rating count.

[35] Kovács / Sharkey 2014.
[36] Maity et al. 2019, p. 219.
[37] Kovács / Sharkey 2014.

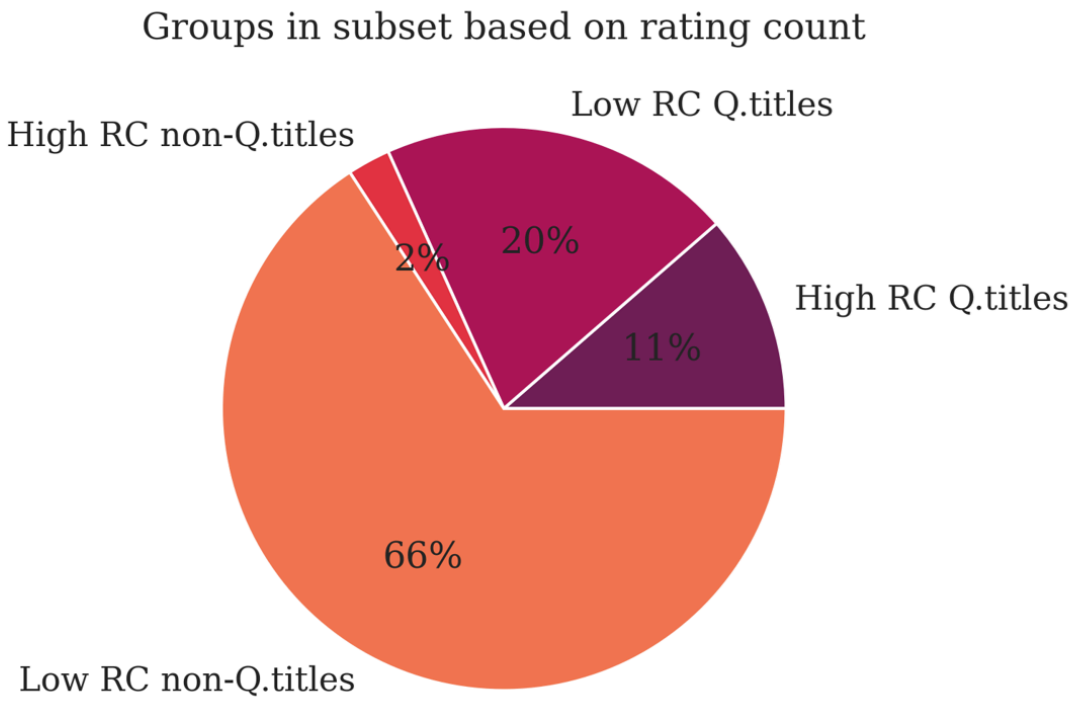

Figure 6: Pieplot showing the sizes of groups below and above subset mean rating count (RC) within quality and non-quality titles. [Chart: Feldkamp et al. 2025]

In figure 6, we see that a much larger portion of quality titles place above the subset mean compared to the non-quality group. Still, a considerable part of the ›quality‹ group – approximately ⅓ of the group – does not have a high rating count, indicating works of what we might call ›indecisive quality‹, That is, they are valued highly beyond Goodreads (contained in other quality proxies), but they do not seem to garner much attention on Goodreads, neither in terms of rating nor in terms of higher rating count. [31]

While there are no differences between the high RC and low RC ›quality‹ groups in terms of publication date or average rating, we find different types of quality proxies more prominent in one group compared to the other (fig. 7). It seems that more canonical works are more prevalent in the high rating count compared to the low rating count ›quality‹ group. Proxies that are institutionally oriented (OpenSyllabus, the Norton Anthology) as well as the GoodRead's Classics list,[38] but also titles that won prestigious awards (NBA, the Nobel Prize) are more frequent in this group. Conversely, titles longlisted for genre-fiction awards appear more prevalent in the low rating count ›quality‹ group (fig. 7). [32]

[38] Titles of which Walsh and Antoniak (2021) found were mirroring those more frequent in college syllabi collected by OpenSyllabus.

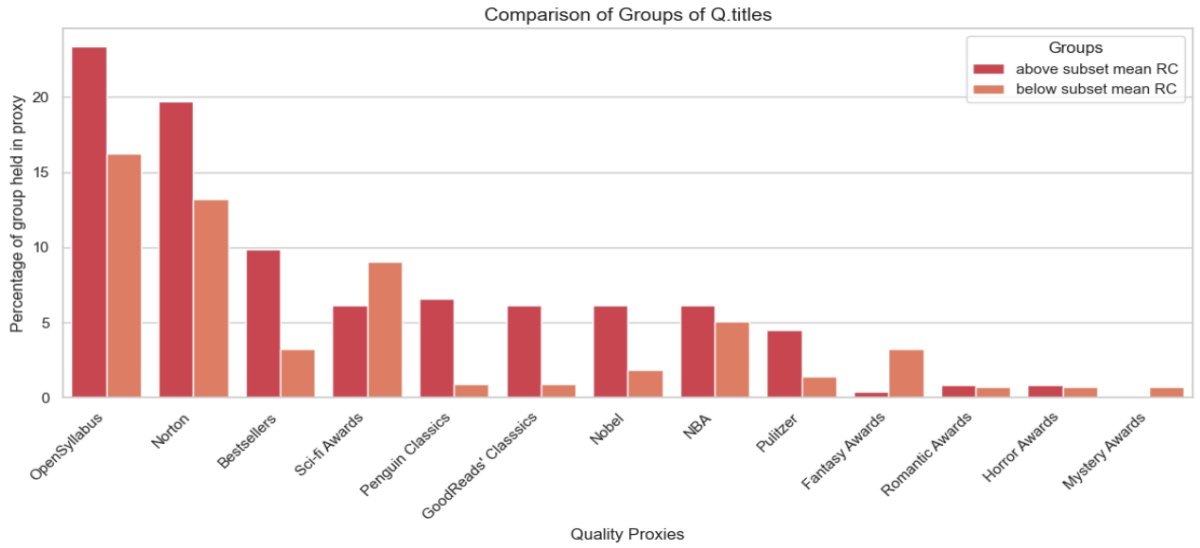

Figure 7: Barplot showing the number of titles from different proxies within the above and the below subset mean rating count (RC) groups. [Chart: Feldkamp et al. 2025]

To further assess this low rating portion of the quality group, two literary scholars manually inspected the titles it contained. They found that many of the titles appear to be less appreciated (not the most prominent) works by rather famous or canonical authors. This includes, for example, *The Little Lady of the Big House* by Jack London, who is most prominently known for works like *The Call of the Wild*, *Flush* and *The Years* by Virginia Woolf, primarily known for her works *Mrs. Dalloway* or *To the Lighthouse*; *The 42nd Parallel* by John Dos Passos, primarily known for his *Manhattan Transfer*; *Steps* by Jerzy Kosinski, primarily known for his *The Painted Bird*, as well as, for example, *Murphy* by Samuel Beckett, primarily known for his theatre plays, and so on. [33]

Conversely, in the high rating quality group, we tend to see more typically canonical works by famous authors, such as Woolf's *To the Lighthouse* and *Mrs. Dalloway*, or *The Portrait of a Lady* by Henry James, as well as works that are very known and sometimes controversial. The controversy of such titles may lie in their political aspects, like Ayn Rand's *The Fountainhead* and Tim LaHaye and Jerry B. Jenkins's *Left Behind*, or their dealing with controversial themes such as Vladimir Nabokov's *Lolita*, but also in being stylistically experimental, such as Joyce's *Ulysses* or William Faulkner's *The Sound and the Fury*, or having a renowned difficult style, such as Malcolm Lowry's *Under the Volcano*. [34]

Moreover, there appears to be a difference in the polarisation of rating distributions between the low and high rating count ›quality‹ groups (fig. 8), where we see a small successive rise from the mean standard deviation of the non-quality group, the quality group with low rating count, to the mean standard deviation of the quality group with higher rating count. By using a Spearman correlation, we find that there is a medium strength correlation between standard deviation of rating distribution in the quality group with high rating count (table 5). [35]

5.4 Quality defined on a title- vs. author-base

An important consideration following from our qualitative inspection of titles in these groups, is that our selection of proxies of literary quality may go some way in explaining why the low rating count ›quality‹ group may contain so many titles that are not very well known but are written by well-known authors. Particularly considering that some of our proxies are author-based, so that every title by an author mentioned in the [36]

proxy (e. g. a Nobel winning author) is included in the ›quality‹ group. This inclusion of titles that are not properly considered quality may also make it harder to distinguish between the quality and non-quality group. To test for this issue, we tried using exclusively title-based proxies of quality against using all proxies.

While the difference between using title and author-based proxies is small (37 titles less), and the mean rating count of the quality group only goes up slightly, to a mean rating count of 27,723.82 compared to 26,767.76 when using author-based proxies, we seem to see a slightly larger difference when examining the polarisation of rating distributions, so that quality groups selected via title-based proxies show a stronger correlation than when using all proxies (table 5). [37]

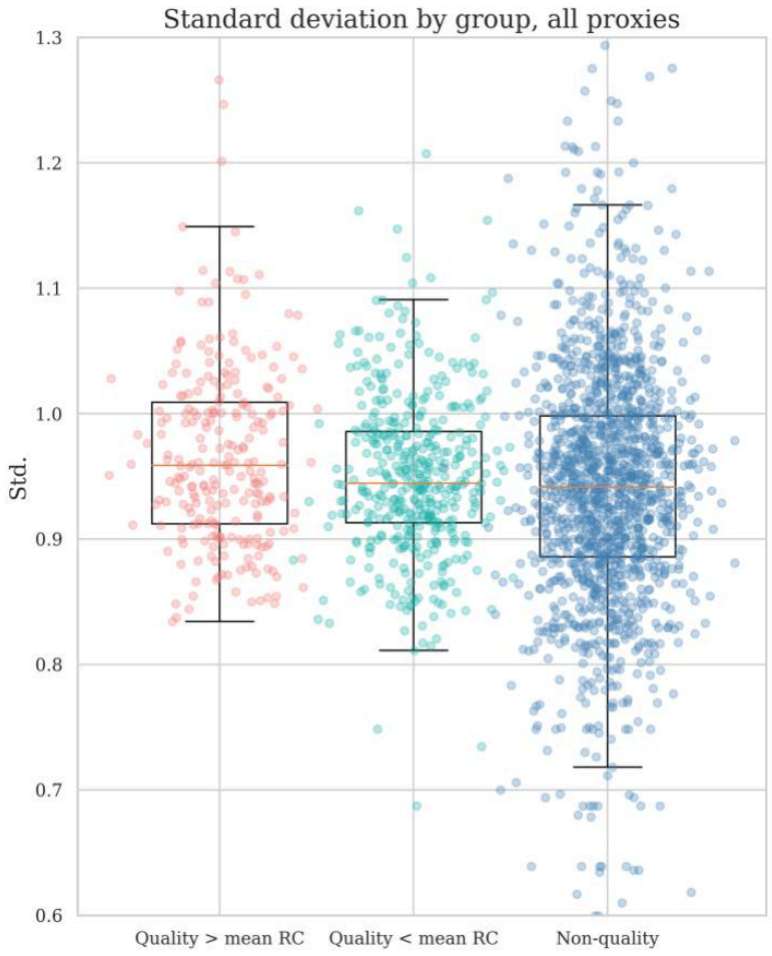

Figure 8: Boxplot showing the distribution of rating count values in the three groups. [Chart: Feldkamp et al. 2025]

	Q. titles > mean RC	**Q. titles < mean RC**	**Q. titles**
Authorbased proxies	0.30** (0.27**)	0.08 (0.06)	0.08* (0.07*)
Titlebased proxies	0.32** (0.29**)	0.09 (0.09)	0.09* (0.07*)

Table 5: Coefficients of Spearman correlations between rating count and standard deviation of rating distributions per group, when using all or exclusively title-based proxies (with entropy values in parentheses). *p < .05 **p < .01

Combining qualitative and quantitative levels of analysis, we visualise the correlation, as well as placement of individual titles within the high rating count ›quality‹ group in fig. 9. Here we tend to see controversial titles like *The Fountainhead*, or stylistically controversial titles like *Ulysses* at the extreme end of standard deviation. Less polarised works, such as *Cujo*, *Dolores Claiborne* and *Firestarter* by the widely popular Stephen King appear at middle values of standard deviation, while works of a mass market type literature such as Agatha [38]

Christie's *The Body in the Library* and *The Mystery of the Blue Train*, part two of Frank Herbert's bestselling Science Fiction series, *Dune Messiah*, and Truman Capote's popular classic *Breakfast at Tiffany's* appear at the very lowest end of standard deviation (fig. 9).

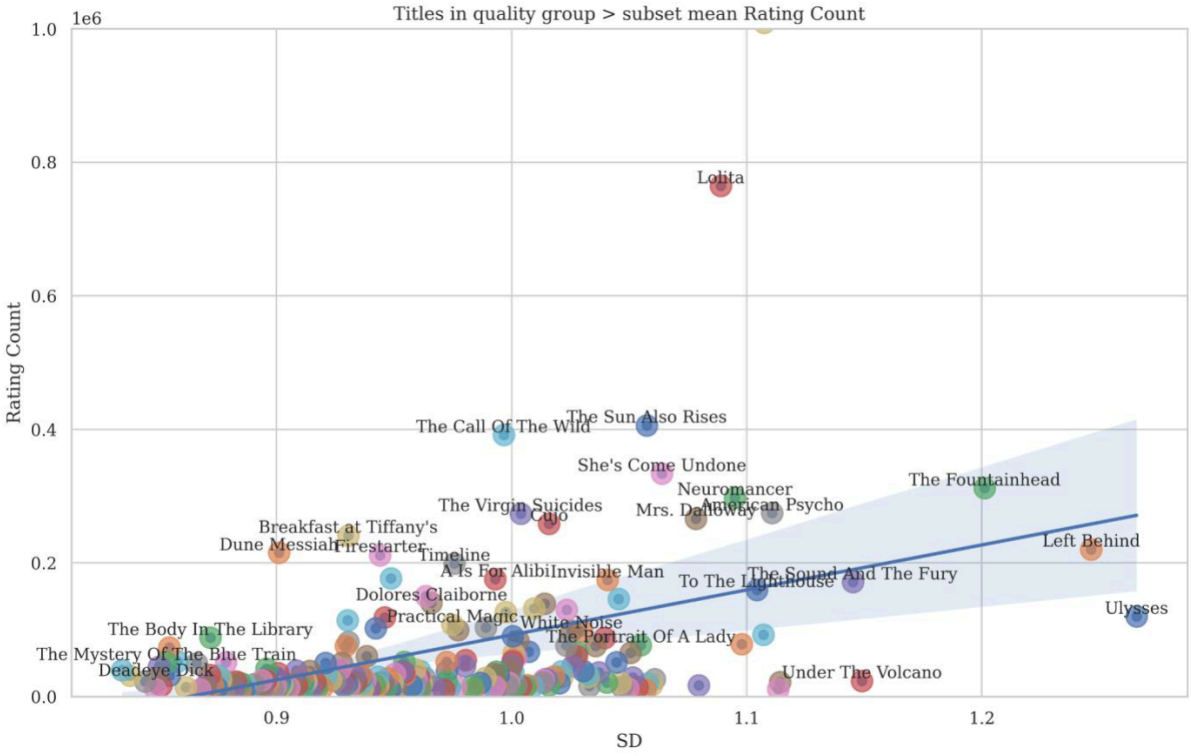

Figure 9: Titles in the quality titles above subset mean rating count group according to their rating count and SD of rating distributions. [Chart: Feldkamp et al. 2025]

6. Conclusion & future works

We have run an examination of a subset of contemporary English language novels falling into the ›least decisive‹ section of our corpus in terms of Goodreads' average ratings, namely the 25 % of titles in our corpus that fall in the range of 3.72 to 3.91 in average rating. Intriguingly, this quartile range includes a notable proportion of works that have received accolades in alternative indices of literary merit, such as inclusion in canonical lists or receipt of esteemed literary awards. For brevity, we called this the ›quality group‹, as it represents several novels highly regarded by alternative evaluative lenses. [39]

This group has on average a higher rating *count* than its ›non-quality‹ counterpart. When looking at the titles in the quality group that are above the mean rating count of the whole subset, and that push the whole group's average up, we found a weak-to-medium correlation between the rating count of the titles and the standard deviation of their ratings, as well as the entropy of the rating distribution. This pattern suggests that titles with higher readership counts exhibit greater polarisation in ratings, corroborating the tendency suggested by Kovács and Sharkey (2014).[39] In other words, there is a portion of titles in the ›quality group‹ whose ratings on Goodreads are also polarised, so that their mean rating count obscures the extreme evaluations that are their basis. [40]

[39] Kovács / Sharkey 2014.

Qualitative analysis of these polarising works revealed distinct characteristics. These titles are often notable for their stylistic experimentation (e. g. Joyce's *Ulysses*), possibly leading to mixed feelings in readers unfamiliar with or appalled by their style, or controversial in terms of their message (Rand's *The Fountainhead*), topic or point of view (Nabokov's *Lolita*). It is noteworthy that this highly-read subset predominantly comprises works considered canonical, with limited representation of genre fiction (fig. 7). Contrastingly, an investigation into the novels within the ›quality group‹ that have lower than average rating counts yielded a cluster of works by renowned or canonical authors that did not gain extensive appraisal, distinguishing them from the more polarised, widely-read titles. [41]

In summary, we have observed that the most mediocre section of Goodreads rating in our corpus comprises at least three populations: [42]

- **Absolute Mediocrity.** These are titles that the majority of readers on the platform rated as neither great nor terrible, and that were not picked by any of the other quality proxies we consulted for having outstanding merits. They were not ›good enough‹ to score higher than 3.91, nor ›bad enough‹ to score below 3.72 on Goodreads, though being included in more than a few libraries around the world (thus making the Chicago Corpus' original selection). Among such works we find, for example, the forgotten classic *Susy, a Story of the Plains* by Bret Harte set in the California Gold Rush, or the dystopian novel by Paul Auster, *In the Country of Last Things*.
- **Relative Mediocrity.** These are titles that most readers on the platform rated as neither great nor terrible, but that were considered by other quality proxies as having outstanding merits. While these titles' average score on Goodreads reflects an actual ›non-decisive‹ judgment from the reader's majority, they appear to be polarizing when we consider a wider evaluative landscape in literary culture. An example in this subgroup is *The 42nd Parallel* by John Dos Passos, i. e., a book by a famous author who is most known for other works.
- **Apparent Mediocrity.** These titles are polarising both inter- and intra- proxies, in other words, they are widely read, they are considered outstanding in other quality proxies, and they are also considered to be great by a portion of the Goodreads' community itself. Their apparent mediocrity is an effect of two very decisive readerships pulling the rope in opposite directions. Examples of these titles are Joyce's *Ulysses* or Nabokov's *Lolita*.

What is ›mediocre‹ on Goodreads is thus a particular flavour of mediocrity, composed of different subpopulations. As such, it is advisable for studies that use Goodreads ratings as a proxy for literary quality to assess the amount of their corpus that consists of such works and to use Goodreads ratings to predict quality while keeping this aspect in mind. It's notable that in this respect, the number of ratings can be more telling than the average score, as titles categorised as ›apparently mediocre‹ based on Goodreads ratings exhibited a consistently elevated volume of readership compared to titles that were genuinely indeterminate in their reception. [43]

It is important to underline that the strength of Goodreads' rating system – its reductionist approach that distils reader judgement into a mono-dimensional metric – is also its limitation. The most ›mediocre‹ titles in our selection still have very positive and negative ratings, and the non-decisive or ambivalent judgement of many reviews often emanates from a composition of perceived strengths and weaknesses perceived by the individual readers. Looking forward, our ambition is to conduct a more extensive exploration of rating polarisation within the Goodreads community. While the standard deviation and entropy of individual book rating distributions serve as satisfactory preliminary metrics for gauging polarisation, the employment of more sophisticated analytical techniques is warranted. At a more complex level, applying techniques such as aspect-based opinion mining on Goodreads' reviews could return a much more nuanced (but, again, harder to summarise) landscape on the reception of a given novel. Conversely, the presence of titles held in high regard by other evaluative mechanisms suggests that Goodreads serves as a viable alternative resource for [44]

literary assessment, corroborating the assertions of Verboord (2011).[40] Thus, future research endeavours that compare the mass appraisal methodologies of Goodreads with other literary evaluative proxies could contribute significantly to the discourse on literary evaluation.

[40] Verboord 2011.

Bibliography

Jodie Archer / Matthew Lee Jockers: The Bestseller Code. Anatomy of the Blockbuster Novel. London 2017. [Nachweis im GVK]

Phillip A. Bishop / Robert Louis Herron: Use and Misuse of the Likert Item Responses and Other Ordinal Measures. In: International Journal of Exercise Science 8 (2015), no. 3, pp. 297–302. DOI: 10.70252/LANZ1453

Yuri Bizzoni / Pascale Feldkamp Moreira / Nicole Dwenger / Ida Marie Schytt Lassen / Kristoffer Laigaard Nielbo / Mads Rosendahl Thomsen (2023a): Good Reads and Easy Novels: Readability and Literary Quality in a Corpus of US-published Fiction. In: Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa, Tórshavn, FO, 22.–24.05.2023). 2023. PDF. [online]

Yuri Bizzoni / Pascale Feldkamp Moreira / Mads Rosendahl Thomsen / Kristoffer Laigaard Nielbo (2023b): Sentimental Matters. Predicting Literary Quality with Sentiment Analysis and Stylometric Features. In: Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis (ACL, Toronto, 14.07.2023). 2023, pp. 11–18. PDF. [online]

Yuri Bizzoni / Mads Rosendahl Thomsen / Pascale Feldkamp Moreira / Kristoffer Laigaard Nielbo (2023c): Modeling Readers' Appreciation of Literary Narratives Through Sentiment Arcs and Semantic Profiles. In: Nader Akoury / Elizabeth Clark / Mohit Iyyer / Snigdha Chaturvedi / Faeze Brahman / Khyathi Chandu (eds.): The 5th Workshop on Narrative Understanding. Proceedings of the Workshop (WNU, Toronto, 14.07.2023). Stroudsburg, US-PA 2023, pp. 25–35. PDF. [online]

Andreas Wolf van Cranenburgh: Rich Statistical Parsing and Literary Language. PhD thesis, University of Amsterdam 2016. PDF. Handle: 11245/1.543163

Tess Crosbie / Tim French / Marc Conrad: Towards a Model for Replicating Aesthetic Literary Appreciation. In: Roberto De Virgilio / Fausto Giunchiglia / Letizia Tanca (eds.): SWIM '13: Proceedings of the Fifth Workshop on Semantic Web Information Management (New York, 23.06.2013). New York 2013. DOI: 10.1145/2484712.2484720

Pascale Feldkamp / Yuri Bizzoni / Mads Rosendahl Thomsen / Kristoffer Laigaard Nielbo: Measuring Literary Qualtiy: Proxies and Perspectives. In: Journal of Computational Literary Studies 3 (2024), no. 1. DOI: 10.48694/jcls.3908

Vikas Ganjigunte Ashok / Song Feng / Yejin Choi: Success with Style: Using Writing Style to Predict the Success of Novels. In: David Yarowsky / Timothy Baldwin / Anna Korhonen / Karen Livescu / Steven Bethard (eds.): Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP, Seattle, 18.–21.10.2013) Seattle 2013. PDF. [online]

John Guillory: Canon, Syllabus, List: A Note on the Pedagogic Imaginary. In: Transition (1991), no. 52, pp. 36–54. DOI: 10.2307/2935123

Eric Bjerck Hagen / Christine Hamm / Frode Helmich Pedersen / Jørgen Magnus Sejersted / Eirik Vassenden: Literary Quality: Historical Perspectives. In: Knut Ove Eliassen / Jan Fredrik Hovden / Øyvind Prytz (eds.): Contested Qualities. Bergen 2018, pp. 47–73.

Syeda Jannatus Saba / Biddut Sarker Bijoy / Henry Gorelick / Sabir Ismail / Md Saiful Islam / Mohammad Ruhul Amin: A Study on Using Semantic Word Associations to Predict the Success of a Novel. In: Lun-Wei Ku / Vivi Nastase / Ivan Vulić (eds.): Proceedings of the 10th Joint Conference on Lexical and Computational Semantics (*SEM 2021, online, 05.–07.08.2021). Stroudsburg, US-PA 2021. DOI: 10.18653/v1/2021.starsem-1.4

Matthew Lee Jockers: Macroanalysis: Digital Methods and Literary History (= Topics in the Digital Humanities). Urbana, US-IL 2013. [Nachweis im GVK]

Corina Koolen / Karina van Dalen-Oskam / Andreas van Cranenburgh / Erica Nagelhout: Literary Quality in the Eye of the Dutch Reader: The National Reader Survey. In: Poetics 79 (2020). 15.02.2020. DOI: 10.1016/j.poetic.2020.101439

Balázs Kovács / Amanda J. Sharkey: The Paradox of Publicity: How Awards Can Negatively Affect the Evaluation of Quality. In: Administrative Science Quarterly 59 (2014), no. 1, pp. 1–33. DOI: 10.1177/0001839214523602

Suraj Maharjan / Manuel Montes-y-Gómez / John Arevalo / Fabio Augusto González / Thamar Solorio: A Multi-Task Approach to Predict Likability of Books. In: Mirella Lapata / Phil Blunsom / Alexander Koller (eds.): Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL, Valencia, 03.–07.04.2017), Volume 1: Long Papers. Stroudsburg, US-PA 2017. PDF. [online]

Suraj Maharjan / Sudipta Kar / Manuel Montes-y-Gómez / Fabio Augusto González / Thamar Solorio: Letting Emotions Flow: Success Prediction by Modeling the Flow of Emotions in Books. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT, New Orleans, 01.–06.06.2018), Volume 2 (Short Papers). Stroudsburg, US-PA 2018. PDF. DOI: 10.18653/v1/N18-2042

Suman Kalyan Maity / Abhishek Panigrahi / Animesh Mukherjee: Analyzing Social Book Reading Behavior on Goodreads and How It Predicts Amazon Best Sellers. In: Mehmet Kaya / Reda Alhajj (eds.): Influence and Behavior Analysis in Social Networks and Social Media (= Lecture Notes in Social Networks). Cham 2019, pp. 211–235. DOI: 10.1007/978-3-030-02592-2_11

Dominique Makowski / Tam Pham / Zen Juen Lau / Jan C. Brammer / François Lespinasse / Hung Pham / Christopher Schölzel / Shen-Hsing Annabel Chen: NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing. In: Behavior Research Methods 53 (2021), pp. 1689–1696. 02.02.2021. DOI: 10.3758/s13428-020-01516-y

Alexander Manshel / Laura B. McGrath / Jack D. Porter: Who Cares about Literary Prizes? In: Public Books. 09.03.2019. HTML. [online]

Mahdi Mohseni / Christoph Redies / Volker Gast: Approximate Entropy in Canonical and Non-Canonical Fiction. In: Entropy 24 (2022), no. 2. 15.02.2022. DOI: 10.3390/e24020278

Franco Moretti: The Slaughterhouse of Literature. In: MLQ: Modern Language Quarterly 61 (2000), no. 1, pp. 207–227. DOI: 10.1215/00267929-61-1-207

Franco Moretti: Graphs, Maps, Trees: Abstract Models for Literary History. New York etc. 2007. [Nachweis im GVK]

Lisa Nakamura: »Words with Friends«: Socially Networked Reading on *Goodreads*. In: PMLA / Publications of the Modern Language Association of America 128 (2013), no. 1, pp. 238–243. DOI: 10.1632/pmla.2013.128.1.238

Willie van Peer (ed.): The Quality of Literature. Linguistic Studies in Literary Evaluation (= Linguistic Approaches to Literature, 4). Amsterdam 2008. DOI: 10.1075/lal.4

Jack D. Porter: Popularity / Prestige: A New Canon (= Literary Lab Pamphlet, 6). 29.10.2018. HTML. [online]

Brian Abel Ragen: An Uncanonical Classic: The Politics of the *Norton Anthology*. In: Christianity and Literature 41 (1992), no. 4, pp. 471–479. DOI: 10.1177/014833319204100409

Andrew James Reagan / Lewis Mitchell / Dilan Kiley / Christopher M. Danforth / Peter Sheridan Dodds : The Emotional Arcs of Stories Are Dominated by Six Basic Shapes. In: EPJ Data Science 5 (2016). 04.11.2016. DOI: 10.1140/epjds/s13688-016-0093-1

Ann Steiner: Private Criticism in the Public Space: Personal Writing on Literature in Readers' Reviews on *Amazon*. In: Participations 5 (2008), no. 2. PDF. [online]

Mike Thelwall / Kayvan Kousha: Goodreads: A Social Network Site for Book Readers. In: Journal of the Association for Information Science and Technology 68 (2017), no. 4, pp. 972–983. DOI: 10.1002/asi.23733

Marc Verboord: Female Bestsellers: A Cross-National Study of Gender Inequality and the Popular–Highbrow Culture Divide in Fiction Book Production, 1960–2009. In: European Journal of Communication 24 (2012), no. 4, pp. 395–409. DOI: 10.1177/0267323112459433

Melanie Walsh / Maria Antoniak: The Goodreads »Classics«: A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism. In: Post45 (2021), no. 7. 21.04.2021. [online]

Kai Wang / Xiaojuan Liu / Yutong Han: Exploring Goodreads Reviews for Book Impact Assessment. In: Journal of Informetrics 13 (2019), no. 3, pp. 874–886. DOI: 10.1016/j.joi.2019.07.003

Xindi Wang / Burcu Yucesoy / Onur Varol / Tina Eliassi-Rad / Albert-László Barabási: Success in Books: Predicting Book Sales before Publication. In: EPJ Data Science 8 (2019). 17.10.2019. DOI: 10.1140/epjds/s13688-019-0208-6

René Wellek: The Attack on Literature. In: The American Scholar 42 (1972), no. 1, pp. 27–42. [online]

List of Figures and Tables