

Weak Negative Correlation between the Present Day Popularity and the Mean Emotional Valence of Late Victorian Novels

Stefan Veleski^a

^aMasaryk University, Žerotínovo náměstí 617/9, 601 77, Brno, Czech Republic

Abstract

Despite the recent upswing of computational research on Victorian novels, it has largely overlooked insight from cultural evolution and the cognitive sciences. This study aims to contribute to this incipient scholarship by testing the hypothesis that novels containing content with a lower mean emotional valence are more likely to trigger recommendation-based transmission chains, and as a result tend to have greater cultural longevity. This study performs a correlation analysis between the mean sentiment and the contemporary popularity (using the number of user ratings from Goodreads) of a selection of late Victorian novels published in the United Kingdom between 1891 and 1901, taken from Project Gutenberg (n=846). Moreover, the study looks into the implications of this correlation for the differences between novels that were bestsellers at the time of publication and those that can be considered canonical today (that have recently had Broadview, Oxford University, or Penguin Press editions). The results show a weak negative correlation between the present day popularity and the mean emotional valence of the novels, which nevertheless holds true for both the bestselling and canonical novels. Moreover, canonical novels tend to have a lower mean emotional valence than the bestsellers.

Keywords

cultural evolution, sentiment analysis, Victorian novels, cultural longevity, bestsellers, canon

1. Introduction

In recent years, the increased availability of cultural data has led to an explosion of hypothesis-driven computational research of Victorian fiction. The Victorian period lends itself very well to computational approaches—it is situated in the “goldilocks” zone of simultaneously being recent enough to resemble the contemporary publishing industry, while being entirely in the public domain. Moreover, the literary forms that dominated the market in the Victorian period are still highly popular today, and so are many of the works published in this period. For the most part, computational research of Victorian fiction has sought to empirically retrace hypotheses by traditional literary scholars, or to conduct exploratory analyses revealing “hidden” cultural dynamics that have previously only been subject to conjecture. In order to do so, researchers have “operationalized” [30], i.e. quantified and measured, a large repertoire of textual features of Victorian fiction. For example, research has targeted the levels of abstraction[20], redundancy [2], topic distribution[44], dialogic liveliness [43], and so forth.


CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands

✉ veleski@phil.muni.cz (S. Veleski)

📞 0000-0002-6097-0864 (S. Veleski)

© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

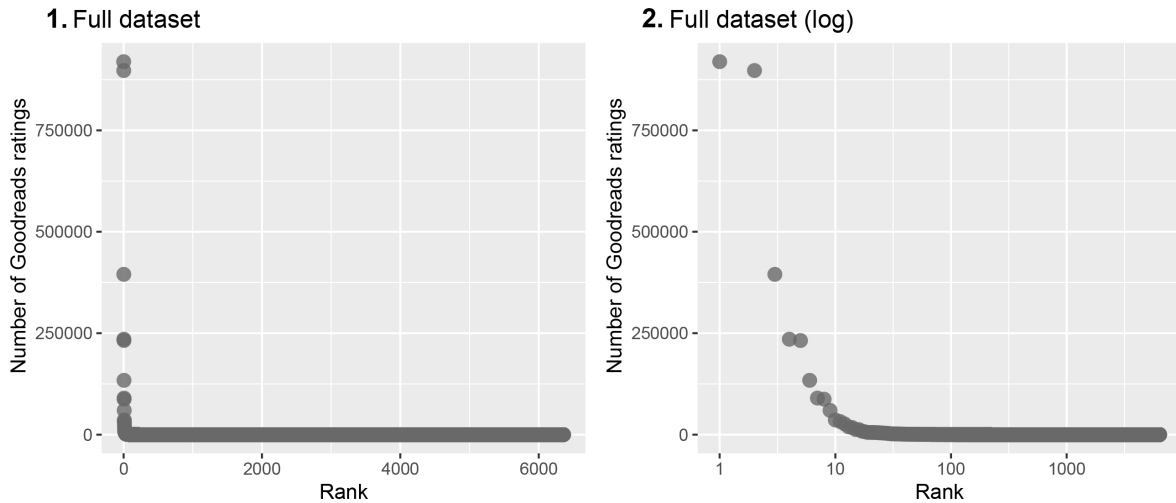


Figure 1: Power law distribution of the present day popularity of late Victorian novels. This plot was made using the “ggplot2” package [49] in R (version 3.6.3). All subsequent analyses and visualizations were made using this programming language.

However, very little of this research relies on hypotheses informed by cultural evolution, a broad interdisciplinary theoretical framework seeking to uncover large scale evolutionary dynamics behind cultural change [26], although a few promising attempts in this direction [42] have been made.

Insight derived from cultural evolution can be very valuable for the ongoing canonicity debate in the field of literary studies, and it can help shift the debate away from abstract notions of power structures, to the more quantifiable notion of cultural longevity—what causes some cultural products to endure over the years, while others to disappear [31]? Considering the fact that 120 years stand between now and the Victorian period, the diachronic dynamics of “cultural survival” and “cultural death” for late Victorian novels can be traced relatively easily. Figure 1 shows the present day popularity of late Victorian novels published between 1891 and 1901 taken from the “At the Circulating Library” website [3] (more about the datasets in the Data section).

It is immediately apparent that the contemporary popularity of late Victorian novels is characterized by a power law distribution, or extreme inequality. Similar distributions of popularity are actually quite common, occurring in a wide range of cultural traits [7, 27], but in this case, the inequality is further amplified by the elapsed time since the original publication of the novels.

Research suggests that over the last century there has been a widespread decrease of the duration and intensity of attention that cultural products have been able to attract around themselves, most likely due to the ever increasing number of cultural products available [13, 25]. Why should the modern reader invest time and energy into reading 120 year old books, when the present day publishing industry has so many new novels to offer (not to mention the large variety of other media in existence today)? As a certain amount of decay in the attention that cultural products are able to attract is to be expected, it is the longevity of the “canon” that is anomalous, and not the “cultural death” of the majority of the works in the archive.

Some studies indicate that value-neutral heuristics, i.e. random copying, can lead to power

law distributions of some cultural traits [7], but the non-linear nature of the success of late Victorian novels over time (most late Victorian bestsellers are virtually forgotten today), indicates that there must have been certain selection pressures at play. Information cascades, otherwise known as the Matthew Effect or cumulative advantage [48], seem to be a more likely reason—especially considering the growing empirical evidence indicating that there often are certain thresholds that the “quality” of the content of cultural products needs to meet before information cascades are triggered [40, 41, 31]. Therefore, the computational analysis of the textual contents of late Victorian novels should cast some light on the power law distribution of their popularity in the present day. This article will focus on one of the most frequently investigated content biases in cultural transmission—emotional bias.

Although some studies suggest positivity bias in language in general [17, 24], there is a large body of research indicating a general negativity bias in information processing [21, 5, 39, 46], most often linked to the “smoke detector” principle—negatively valenced stimuli are less likely to be ignored as they are more likely to signal negative developments [33]. Moreover, research suggests that negatively valenced information is more likely to attract the attention of individuals [36], and narrow their attentional focus [10, 11], while positively valenced emotion widens the breadth of attentional allocation [38]. Research on emotion in cultural transmission suggests a bias for negatively valenced emotion, for example in rumors [47] and stories [6]. Research has also shown that stories containing disgust (which is unambiguously a negatively valenced emotion) fare better in cultural transmission than those that do not [19, 18].

There have also been several diachronic studies that have traced the cultural dynamics of sentiment in a variety of cultural products (song lyrics [16, 15, 12] and Anglophone literature [32, 1]). All of these studies suggest a decline of the mean emotional valence of cultural products over the last century, but all except one [12] connect the decline of the mean emotional valence to a decline of words related to positive emotion, instead of to an increase of the words related to negative emotion. Out of these, the three studies that use cultural evolution as a theoretical framework [1, 32, 12] mostly attribute this decline to cultural drift, or the authors’ unbiased copying of their immediate predecessors.

However, none of these studies have dealt with cultural longevity, but with the overall trend of the sentiment of different areas of cultural production across time. Therefore, they have focused on the social learning heuristics of authors rather than the audience or the readership—the producing rather than the receiving side in this cultural dynamic. In order to deal with the cultural longevity of novels, cultural transmission needs to be reframed in the context of novels. Most research in cultural evolution focuses on the transmission of “fluid” narratives, such as songs, folk tales, urban legends, which undergo changes during the process of transmission. In “fluid” narratives transmission entails retelling the narrative in its entirety, while more “monolithic” narratives (e.g., novels) have a fixed form, which cannot be transmitted in its entirety due to its length. So what exactly does the process of cultural transmission for novels consist of? The answer is simple—recommendations, or advice on whether the novel is worth reading, along with semantically condensed versions of the original narratives transmitted by word-of-mouth, written transmission, or any of their proxies in the digital age.

My hypothesis is that sentiment is more important in recommendation-based cultural transmission of novels than the author to author influence chains that drive the dynamics of sentiment levels of cultural products across time. Therefore, lower mean emotional valence of the contents of the relevant novels would increase the likelihood that recommendation-based transmission chains would be triggered. Tracing the content of the actual transmission that causes exposure to the cultural product is impossible, but what we can do is compare the modern day

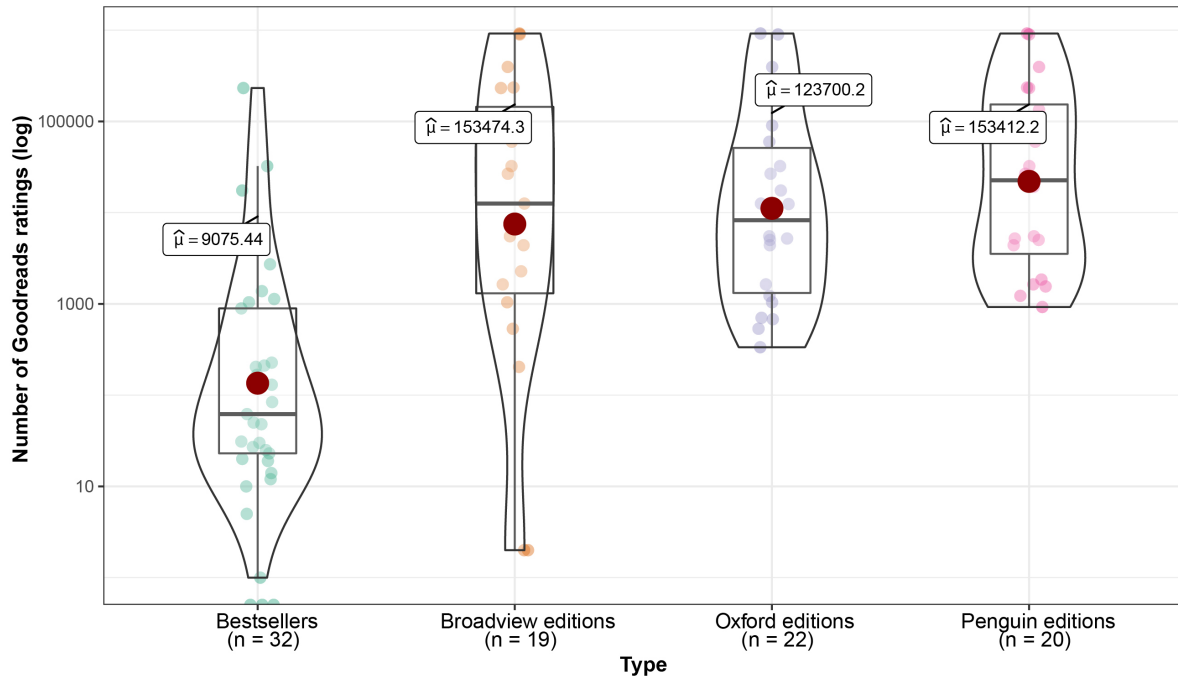


Figure 2: Distribution of the present day popularity of late Victorian bestsellers compared to novels published in recent critical editions by major publishing houses. The boxplot was created using the “ggstatsplot” package in R [34].

popularity of the cultural product and its contents. If the data clearly shows that the novels with extraordinary cultural longevity tend to contain more negatively valenced content than their forgotten contemporaries, then this would suggest that it must have contributed towards the “cultural survival” of the relevant novel.

The secondary hypothesis of this study concerns the difference of mean emotional valence between bestsellers and “canonical novels”. Late Victorian bestsellers were initially extremely popular, but as Figure 2 clearly shows, their longevity is significantly lower than that of the “canonical” novels (more about the datasets in the Data section). Was the mean sentiment of these novels a factor in the “cultural death” of the bestsellers? Considering the literature of the topic, it would be expected that the content of these novels has a higher mean emotional valence than the “survivors”, i.e. “canonical” novels.

2. Data

All of the information about the novels published in the United Kingdom between 1891 and 1901 and their authors was taken from the “At the Circulating Library” website [3] before the June 2020 update. The website indexes various narrative forms (collections of short stories and poems, plays), so only novels were retained in the final dataset (n=6358). In order to perform a sentiment analysis of the novels, their full texts were needed, so their names were matched against those of the full dataset of the Project Gutenberg website (numbering around 60000 novels in total). Only 846 novels of the “At the Circulating Library” dataset had their full texts

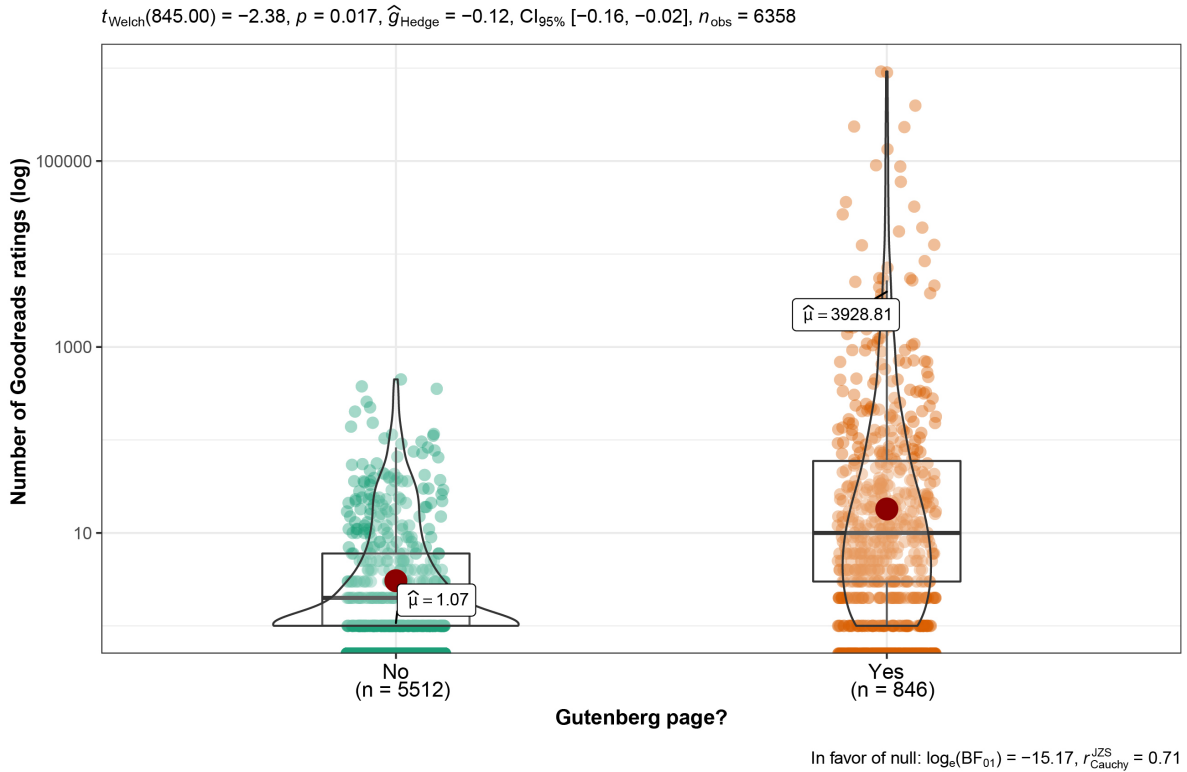


Figure 3: Distribution of the novels that have and do not have their full text available on Project Gutenberg. This boxplot was also created using the “ggstatsplot” package in R [34].

available on the Project Gutenberg website, which in itself is quite telling about the inequality of the present day popularity of late Victorian novels. However, a better, quantifiable proxy for the contemporary popularity of the novels is Goodreads, which allows the users to rate the books on a scale from 1 to 5 and to provide reviews. The insight we can get from the value of the ratings is often flawed, especially for those books that have few ratings, but the number of ratings does give us an approximate idea of the present day popularity of each of these novels. The distribution of present day popularity of the books that have and those that do not have Gutenberg pages is visualized on Figure 3.

The sample of novels that have Gutenberg pages encompasses the entirety of the popularity spectrum, while those that do not are heavily concentrated towards the bottom. 5039 of the 5512 novels that lack Gutenberg pages do not have any ratings on Goodreads as well, while that number is 280 out of 846 for the novels that have Gutenberg pages. We need to bear this bias towards popularity of the full texts dataset in mind when we look at the results of the analysis.

After identifying the Gutenberg IDs of the novels, the “gutenbergr” package [37] was used to download the full texts of the novels. Although the package does remove some of the paratext, the parts of the novel that are not part of the intradiegetic world, i.e. table of contents, advertisements at the end of the novels, were manually removed. The full texts files of the novels are available in the Appendix.

The data on which novels were bestsellers at the time of publication was taken from Bassett and Walter’s study of late Victorian bestseller lists throughout the UK published by “The

Bookman” magazine [4]. Just like in the full dataset, books originally written in another language and works of fiction that are not novels were removed, as well as works of non-fiction, which were surprisingly common in late Victorian bestseller lists. Data about the canonical status of novels was understandably harder to acquire, especially considering how semantically loaded the term is. As a proxy for the semantically complex “canonical status”, this article uses novels that have recently been published by three major publishing houses—Broadview, Oxford University and Penguin Press. As there is a significant overlap of the novels published by these three major publishing houses, the main correlation analysis considers these novels as a single group. Just like the general information about the novels and the authors of the novels published between 1891 and 1901, this data too was taken from the “At the Circulating Library” website [3]. In the Discussion section, several alternatives to this approach are provided.

3. Method

The sentiment analysis of the novels was performed using the “syuzhet” package in R, developed by Matthew Jockers [22]. The default, “syuzhet” sentiment dictionary was used. This sentiment dictionary was compiled by the Nebraska Literary Lab and contains 10748 words in total (7161 positive and 3587 negative ones), each of them with a sentiment value ranging from -1 to 1. The “syuzhet” package first divides the text into sentences and then matches the sentiment of the words in each sentence to their sentiment value in the “syuzhet” sentiment dictionary. Then, the package adds up the sentiment value of each word on a sentence level. As this paper is interested in the general emotional valence of the full text of the novels, the mean of the sentiment value of all the sentences in the novel was calculated.

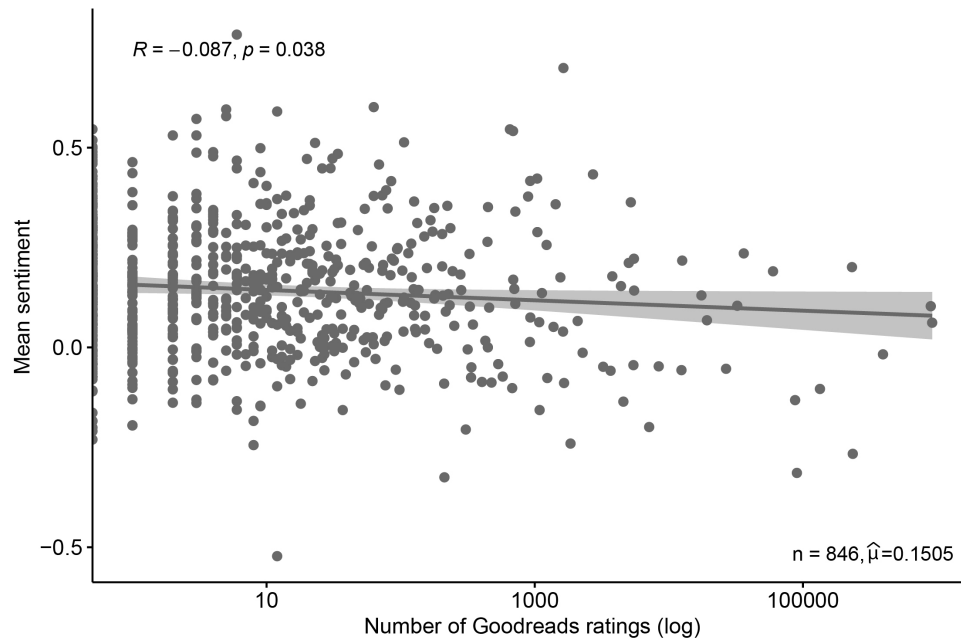
After the mean sentiment of the novels whose full texts were available on Project Gutenberg (n=846) was determined, a correlation analysis of the number of Goodreads ratings and the mean sentiment of the novels was performed. In addition, Welch’s unequal variances t-test was done on the mean emotional valences of the groups of bestselling and canonical novels in order to see whether or not the difference in the means of the sentiment of novels belonging to the two groups is due to chance.

4. Results

The analysis shows a weak negative correlation between the mean emotional valence of the texts of the novels in the Gutenberg dataset and their present day popularity (Figure 4.1.). The analysis of the canonical and bestselling novels (Figure 4.2.) shows that the trend holds up for them as well, although the smaller sample sizes affect the statistical significance of these analyses. Moreover, the mean emotional valence of each of these two subgroups tell quite an interesting story. The group of bestselling novels have a significantly more positive emotional valence than the group of canonical novels, $t_{Welch}(49.19)=2.35$, $p=0.023$.

The almost identical mean emotional valence of the full dataset (0.1505) and that of the books that are neither canonical or were bestsellers (0.1500), further highlight the difference of the mean sentiment of these two groups.

1. Correlation plot (full dataset)



2. Correlation plot (faceted by status)

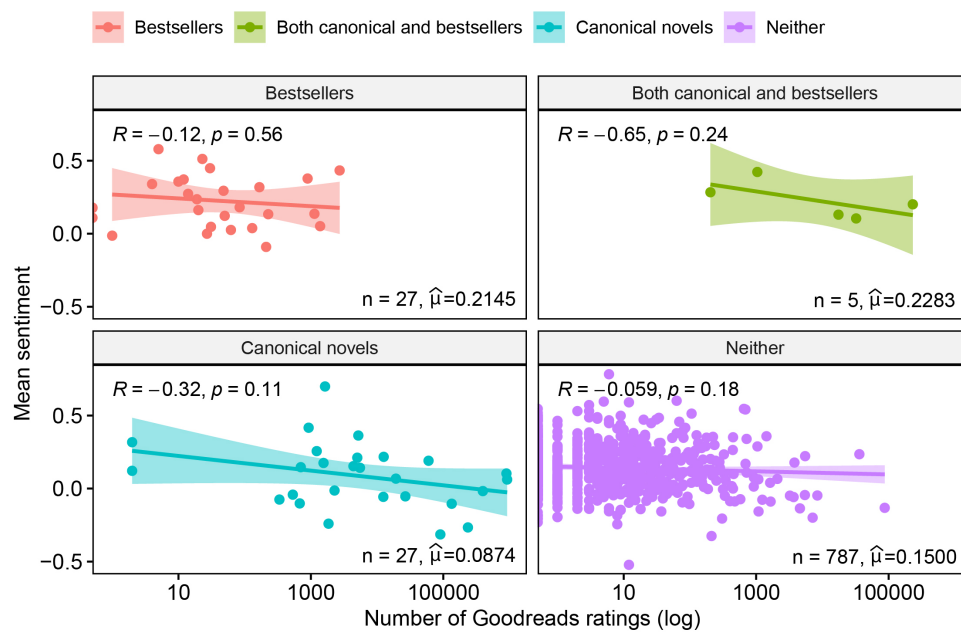


Figure 4: Correlation of sentiment for the full dataset (plot 1) and divided according to status (plot 2). Both correlation plots were made using the “ggpubr” package [23].

5. Discussion

Only a small part of the variation in the present day popularity of late Victorian novels can be explained by their mean emotional valence. Novels are highly complex cultural products, and so is their transmission—many different variables, both content and context related, tug

in opposite directions, thus complicating the insight that might be gotten from such analyses. Other features of the content of the novels (e.g. the presence of particular stylistic features, contemporary cultural artifacts and historical information, as well as over-saturation with topics related to unrelatable contemporary cultural contexts) and contextual factors (prestige, success, conformity biases) most likely account for the rest of the observed variation, and future studies can look into these dynamics.

It is quite possible that most of the observed correlation is owing to the fact that the most popular novels (which are also a part of the “canon”) belong either to the Gothic or science fiction genres, which is an interesting fact in itself. A further study might also explore whether these genres do indeed tend to predominately contain negatively valenced content, and why these two genres appear to be favored in cultural transmission. It is likely that this is because the content of these genres is much more heavily saturated with minimally counterintuitive (MCI) concepts [14], which fare quite well in cultural transmission.

This study had several limitations that future researchers should take into consideration. For example, some of the results regarding sentiment were occasionally skewed by the authors’ usage of dialects (most notably the Kailyard school of Scottish fiction), where the authors try to mimic certain local dialects and thus produce words that do not match the standard, dictionary form. Naturally these words are missing in the “syuzhet” sentiment dictionary, meaning that they are not tallied when calculating the mean emotional valence on a sentence level.

By using the number of Goodreads ratings as a proxy for the contemporary success of late Victorian novels, this paper simplifies what research in the digital humanities has lately regarded as two distinct categories of success: popularity and prestige [45, 2, 35] (the former emerging from unsupervised reader preferences and the latter from those of an “elite” group of literary scholars and critics, which has the power to affect the degree of exposure of the general readership to a limited group of novels). The resulting dynamics are quite complicated and beyond the scope of this paper, whose population-based approach was conceptually guided by Moretti’s claim that the “social” canon shapes the “academic” canon and not the other way around [29]. Future studies can also use resources like the MLA bibliography and the Open Syllabus project in order to disentangle the complex social dynamics of literary prestige, and whether they are in any way affected by the emotional valence of the novels.

Furthermore, just measuring the mean of the emotional valence of novels as a whole could sometimes be problematic. To illustrate, *Dracula* has a mean sentiment of 0.1031, slightly higher than the mean emotional valence of canonical novels (0.0874), which is somewhat counterintuitive. The unexpectedly high mean sentiment of the novel might be owing to the fact that in addition to being a Gothic/horror novel, *Dracula* is also an epistolary novel and large portions of the text (for example Mina’s journal entries and the correspondence between her and Lucy) are filled with passages that have a high mean emotional valence which inflate the mean sentiment of the work. Therefore, future studies will need to focus on the concentration of negative emotional valence instead of only looking at the mean.

Moreover, a more modular analysis of sentiment in the cultural dynamics of longevity could provide answers to many of the questions that this study left unanswered. For example, are certain specific emotions (e.g. disgust) preferred in the recommendation-based transmission chains? Certain sentiment dictionaries like the NRC sentiment lexicon [28] break down the emotions into different semantic categories which can be useful in such analyses. Moreover, as certain studies [8, 9] indicate that emotional arousal could play a bigger part in the transmission of information than emotional valence, future studies can look into this issue as well.

References

- [1] A. Acerbi et al. “The Expression of Emotions in 20th Century Books”. In: *PLoS ONE* 8.3 (2013), pp. 1–6. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0059030.
- [2] M. Algee-Hewitt et al. “Canon/Archive. Large-scale Dynamics in the Literary Field”. In: *Pamphlets of the Stanford Literary Lab* 11.January (2016), pp. 1–13. ISSN: 2164-1757. URL: <https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf>.
- [3] T. J. Bassett. *At the Circulating Library: A Database of Victorian Fiction, 1937-1901*. 2020. URL: <http://www.victorianresearch.org/atcl/index.php>.
- [4] T. J. Bassett and C. M. Walter. “Booksellers and Bestsellers: British Book Sales as Documented by “The Bookman”, 1891-1906”. In: *Book History* 4 (2001), pp. 205–236. URL: <http://www.jstor.org/stable/30227332>.
- [5] R. F. Baumeister et al. “Bad Is Stronger Than Good”. In: *Review of General Psychology* 5.4 (2001), pp. 323–370. ISSN: 1089-2680. DOI: 10.1037/1089-2680.5.4.323.
- [6] K. Bebbington et al. “The Sky Is Falling: Evidence of a Negativity Bias in the Social Transmission of Information”. In: *Evolution and Human Behavior* 38.1 (2017), pp. 92–101. ISSN: 1090-5138. DOI: 10.1016/j.evolhumbehav.2016.07.004. URL: <http://dx.doi.org/10.1016/j.evolhumbehav.2016.07.004>.
- [7] R. A. Bentley, M. W. Hahn, and S. J. Shennan. “Random drift and culture change”. In: *Proceedings of the Royal Society B: Biological Sciences* 271.1547 (2004), pp. 1443–1450. ISSN: 1471-2970. DOI: 10.1098/rspb.2004.2746.
- [8] J. Berger. “Arousal Increases Social Transmission of Information”. In: *Psychological Science* 22.7 (2011), pp. 891–893. ISSN: 0956-7976. DOI: 10.1177/0956797611413294.
- [9] J. Berger and K. L. Milkman. “What makes online content viral?” In: *Journal of Marketing Research* 49.2 (2012), pp. 192–205. ISSN: 0022-2437. DOI: 10.1509/jmr.10.0353.
- [10] M. A. Bezdek et al. “Neural Evidence That Suspense Narrows Attentional Focus”. In: *Neuroscience* 303 (2015), pp. 338–345. ISSN: 1873-7544. DOI: 10.1016/j.neuroscience.2015.06.055. URL: <http://dx.doi.org/10.1016/j.neuroscience.2015.06.055>.
- [11] M. A. Bezdek and R. J. Gerrig. “When Narrative Transportation Narrows Attention: Changes in Attentional Focus During Suspenseful Film Viewing”. In: *Media Psychology* 20.1 (Jan. 2017), pp. 60–89. ISSN: 1521-3269. DOI: 10.1080/15213269.2015.1121830. URL: <https://www.tandfonline.com/doi/full/10.1080/15213269.2015.1121830>.
- [12] C. O. Brand, A. Acerbi, and A. Mesoudi. “Cultural evolution of emotional expression in 50 years of song lyrics”. In: *Evolutionary Human Sciences* 1.e11 (2019), pp. 1–14. ISSN: 2513-843X. DOI: 10.1017/ehs.2019.11. URL: <https://www.cambridge.org/core/journals/evolutionary-human-sciences/article/cultural-evolution-of-emotional-expression-in-50-years-of-song-lyrics/E6E64C02BDB0480DB13B8B6BB7DFF598>.
- [13] C. Candia et al. “The Universal Decay of Collective Memory and Attention”. In: *Nature Human Behaviour* 3.1 (2018), pp. 82–91. DOI: 10.1038/s41562-018-0474-5.
- [14] M. Clasen. “Attention, Predation, Counterintuition: Why Dracula Won’t Die”. In: *Style* 46.3/4 (2012), pp. 378–398.

- [15] C. N. DeWall et al. “Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics”. In: *Psychology of Aesthetics, Creativity, and the Arts* 5.3 (2011), pp. 200–207. ISSN: 1931-3896. DOI: 10.1037/a0023195.
- [16] P. S. Dodds and C. M. Danforth. “Measuring the happiness of large-scale written expression: Songs, blogs, and presidents”. In: *Journal of Happiness Studies* 11.4 (2010), pp. 441–456. ISSN: 1389-4978. DOI: 10.1007/s10902-009-9150-9.
- [17] P. S. Dodds et al. “Human Language Reveals a Universal Positivity Bias”. In: *Proceedings of the National Academy of Sciences* 112.8 (June 2014), pp. 2389–2394. ISSN: 0027-8424. DOI: 10.1073/pnas.1411678112. arXiv: 1406.3855. URL: <http://www.pnas.org/content/112/8/2389.abstract>.
- [18] K. Eriksson and J. C. Coultas. “Corpses, Maggots, Poodles and Rats: Emotional Selection Operating in Three Phases of Cultural Transmission of Urban Legends”. In: *Journal of Cognition and Culture* 14.1-2 (2014), pp. 1–26. ISSN: 1567-7095. DOI: 10.1163/15685373-12342107.
- [19] C. Heath, C. Bell, and E. Sternberg. “Emotional Selection in Memes: The Case of Urban Legends”. In: *Journal of Personality and Social Psychology* 81.6 (2001), pp. 1028–1041. ISSN: 1939-1315. DOI: 10.1037/0022-3514.81.6.1028. URL: <https://doi.apa.org/doiLandin g?doi=10.1037%2F0022-3514.81.6.1028>.
- [20] R. Heuser and L.-K. Long. “A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method”. In: *Pamphlets of the Stanford Literary Lab* 4.May (2012), pp. 1–68. ISSN: 2164-1757. URL: <https://litlab.stanford.edu/Literary LabPamphlet4.pdf>.
- [21] T. A. Ito et al. “Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations.” In: *Journal of personality and social psychology* 75.4 (1998), pp. 887–900. ISSN: 0022-3514. DOI: 10.1037/0022-3514.75.4.887.
- [22] M. L. Jockers. *Syuzhet: Extract Sentiment and Sentiment-Derived Plot Arcs from Text*. 2017. URL: <https://cran.r-project.org/package=syuzhet>.
- [23] A. Kassambara. *ggpubr: “ggplot2” Based Publication Ready Plots*. 2020. URL: <https://cran.r-project.org/web/packages/ggpubr/index.html>.
- [24] I. M. Kloumann et al. “Positivity of the English Language”. In: *PLoS ONE* 7.1 (Aug. 2011). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0029484. arXiv: 1108.5192. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029484>.
- [25] P. Lorenz-Spreen et al. “Accelerating Dynamics of Collective Attention”. In: *Nature Communications* 10.1759 (Dec. 2019). ISSN: 2041-1723. DOI: 10.1038/s41467-019-09311-w. URL: <https://www.nature.com/articles/s41467-019-09311-w>.
- [26] A. Mesoudi. *Cultural Evolution: How Darwinian Theory can Explain Human Culture and Synthesize the Social Sciences*. Chicago, IL: University of Chicago Press, 2011. ISBN: 9780226520438.
- [27] Michael Tauberg. *Power Law in Popular Media*. June 2018. URL: <https://medium.com/@michaeltauberg/power-law-in-popular-media-7d7efef3fb7c> (visited on 07/26/2020).
- [28] S. M. Mohammad and P. D. Turney. “NRC emotion lexicon”. In: *National Research Council, Canada* 2 (2013).

- [29] F. Moretti. “The Slaughterhouse of Literature”. In: *Modern Language Quarterly* 61.1 (2000), pp. 207–228. ISSN: 0026-7929. DOI: 10.1215/00267929-61-1-207. URL: <https://read.dukeupress.edu/modern-language-quarterly/article/61/1/207-228/19219>.
- [30] F. Moretti. ““Operationalizing”: or, the Function of Measurement in Modern Literary Theory”. In: *Pamphlets of the Stanford Literary Lab* Pamphlet 6 (2013). ISSN: 2164-1757. DOI: 10.15794/jell.2014.60.1.001. URL: <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>.
- [31] O. Morin. *How Traditions Live and Die*. Oxford: Oxford University Press, 2016.
- [32] O. Morin and A. Acerbi. “Birth of the cool: a two-centuries decline in emotional expression in Anglophone fiction”. In: *Cognition and Emotion* 31.8 (2017), pp. 1663–1675. ISSN: 1464-0600. DOI: 10.1080/02699931.2016.1260528. URL: <https://doi.org/10.1080/02699931.2016.1260528>.
- [33] R. M. Nesse. “The Smoke Detector Principle”. In: *Annals of the New York Academy of Sciences* 935.1 (Jan. 2006), pp. 75–85. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.2001.tb03472.x. URL: <http://doi.wiley.com/10.1111/j.1749-6632.2001.tb03472.x>.
- [34] I. Patil. “ggstatsplot: “ggplot2” Based Plots with Statistical Details”. In: *CRAN* (2018). DOI: 10.5281/zenodo.2074621. URL: <https://cran.r-project.org/package=ggstatsplot>.
- [35] J. Porter. “Popularity/Prestige”. In: *Pamphlets of the Stanford Literary Lab* 17.September (2018), pp. 1–12. ISSN: 2164-1757. URL: <https://litlab.stanford.edu/LiteraryLabPamphlet17.pdf>.
- [36] F. Pratto and O. P. John. “Automatic Vigilance: The Attention-Grabbing Power of Negative Social Information”. In: *Journal of Personality and Social Psychology* 61.3 (1991), pp. 380–391. ISSN: 0022-3514. DOI: 10.1037/0022-3514.61.3.380.
- [37] D. Robinson. *gutenbergr: Download and Process Public Domain Works from Project Gutenberg*. 2019. URL: <https://cran.r-project.org/package=gutenbergr>.
- [38] G. Rowe, J. B. Hirsh, and A. K. Anderson. “Positive affect increases the breadth of attentional selection”. In: *Proceedings of the National Academy of Sciences of the United States of America* 104.1 (Jan. 2007), pp. 383–388. ISSN: 0027-8424. DOI: 10.1073/pnas.0605198104. URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.0605198104>.
- [39] P. Rozin and E. B. Royzman. “Negativity bias, negativity dominance, and contagion”. In: *Personality and Social Psychology Review* 5.4 (2001), pp. 296–320. ISSN: 1088-8683. DOI: 10.1207/S15327957PSPR0504_2.
- [40] M. J. Salganik, P. S. Dodds, and D. J. Watts. “Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market”. In: *Science* 311.5762 (Feb. 2006), pp. 854–856. ISSN: 0036-8075. DOI: 10.1126/science.1121066. URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.1121066>.
- [41] M. J. Salganik and D. J. Watts. “Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market”. In: *Social Psychology Quarterly* 71.4 (Dec. 2008), pp. 338–355. ISSN: 0190-2725. DOI: 10.1177/019027250807100404. URL: <https://journals.sagepub.com/doi/10.1177/019027250807100404>.
- [42] O. Sobchuk. “Charting Artistic Evolution : An Essay in Theory”. Ph.D. thesis. University of Tartu, 2018. ISBN: 978-9949-77-882-9.

- [43] O. Sobchuk. “The Evolution of Dialogues: A Quantitative Study of Russian Novels (1830–1900)”. In: *Poetics Today* 37.1 (2016), pp. 137–154. ISSN: 0333-5372. DOI: 10.1215/03335372-3452643.
- [44] T. Underwood. *Distant Horizons: Digital Evidence and Literary Change*. Chicago and London: The University of Chicago Press, 2019.
- [45] T. Underwood and J. Sellers. “The Longue Durée of Literary Prestige”. In: *Modern Language Quarterly* 77.3 (2016), pp. 321–344. ISSN: 0026-7929. DOI: 10.1215/00267929-3570634.
- [46] A. Vaish, T. Grossmann, and A. Woodward. “Not All Emotions Are Created Equal: The Negativity Bias in Social-Emotional Development”. In: *Psychological Bulletin* 134.3 (2008), pp. 383–403. ISSN: 0033-2909. DOI: 10.1037/0033-2909.134.3.383.
- [47] C. J. Walker and B. Blaine. “The virulence of dread rumors: A field experiment”. In: *Language and Communication* 11.4 (1991), pp. 291–297. ISSN: 0271-5309. DOI: 10.1016/0271-5309(91)90033-R.
- [48] D. J. Watts. *Everything Is Obvious, Once You Know the Answer: How Common Sense Fails Us*. New York: Crown Business, 2011.
- [49] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: <https://ggplot2.tidyverse.org>.

A. Data availability

All the datasets, text files, and code used in this article are available in this Github repository.