The Fractality of Sentiment Arcs for Literary Quality Assessment: the Case of Nobel Laureates

Yuri Bizzoni, Pascale Feldkamp Moreira, Mads Rosendahl Thomsen, Kristoffer L Nielbo

Journal of Data Mining and Digital Humanities

EPIsciences overlay journals

The Fractality of Sentiment Arcs for Literary Quality Assessment: the Case of Nobel Laureates

Yuri Bizzoni[1,2], **Pascale Feldkamp Moreira**[1,2], **Mads Rosendahl Thomsen**[2], **Kristoffer L. Nielbo**[1]

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

Corresponding author: Yuri Bizzoni , yuri.bizzoni@cc.au.dk

Abstract

In the few works that have used NLP to study literary quality, sentiment and emotion analysis have often been considered valuable sources of information. At the same time, the idea that the nature and polarity of the sentiments expressed by a novel might have something to do with its perceived quality seems limited at best. In this paper, we argue that the fractality of narratives, specifically the longterm memory of their sentiment arcs, rather than their simple shape or average valence, might play an important role in the perception of literary quality by a human audience. In particular, we argue that such measure can help distinguish Nobel-winning writers from control groups in a recent corpus of English language novels. To test this hypothesis, we present the results from two studies: (i) a probability distribution test, where we compute the probability of seeing a title from a Nobel laureate at different levels of arc fractality; (ii) a classification test, where we use several machine learning algorithms to measure the predictive power of both sentiment arcs and their fractality measure. Lastly, we perform another experiment to examine whether arc fractality may be used to distinguish more or less popular works within the Nobel canon itself, looking at the probability of higher GoodReads' ratings at different levels of arc fractality. Our findings seem to indicate that despite the competitive and complex nature of the task, the populations of Nobel and non-Nobel laureates seem to behave differently and can to some extent be told apart by a classifier. Moreover, the probability of Nobel titles having better ratings appears higher at different levels of arc fractality.

Keywords

computational narratology; sentiment analysis; fractal analysis; literary quality assessment; stylometry

I INTRODUCTION

The question of what defines the perception of literary quality is probably as old as narrative itself. However, the ability to process and analyze large quantities of literary texts and to perform complex statistical experiments on them (Moretti [2013]) has recently made new ways of studying this question possible, opening up also for a renewed debate. Moreover, the concept of literary quality is controversial in literary studies, where it seems that distrust of the very idea of literary quality has become a default inclination. Words like "classic" and "canon" are perceived as markers of a passé attitude (Guillory [1995]), especially when canons are thought to represent nothing but entrenched interests (von Hallberg [1983]). The centrality of discussions of social capital, representativity and exclusion in the canon-debate has contributed to marginalizing the question of literary quality (van Peer [2008]). Nevertheless, literary cultures continue to uphold and establish various "canons" in practice, through, for example, reading groups, school anthologies, 'classics' series, and literary awards; and the resilience of canonical works

through time suggests that there certain features or characteristics continuously appreciated by readers. Excessively contextual views do not explain how either the reasons for canonizing, nor why some works remain highly acclaimed across languages, cultures, and time. While achieving a comprehensive understanding of literary quality or establishing a singular metric for canonicity or literary prestige may remain elusive, there is inherent value in exploring the explanatory potential of some well-described textual features, that may be integrated into more complex models of perceived literary quality. In this paper, we examine how the persistence of dynamic sentiment arcs, as measured by the Hurst exponent, can provide above chance-level classification of works by Nobel and non-Nobel laureates. It is particularly important that the measure should be able to rule out the chance of finding a work by a Nobel laureate outside of a limited range in the Hurst exponent. We designed two experiments to examine the difference in Hurst exponent between Nobel and non-Nobel titles, looking firstly at the distribution of Hurst values between the groups, and the probability of finding a Nobel title within certain intervals of Hurst score. Secondly, we examine the pontential of models trained on the Hurst score of texts for classifying Nobel and non-Nobel titles in our corpus. In a third experiment, we further examined the Nobel canon in terms of GoodReads' ratings, using them as an alternative proxy for literary appreciation to test the potential of the Hurst exponent for distinguishing marginally more popular novels within the Nobel canon itself.

II RELATED WORKS

2.1 Sentiment Arcs of Narratives

Various studies have leveraged computational methods to try and predict "literary quality", reader appreciation or success, looking at both text-extrinsic features such as the visibility of an author, what publishing house publishes the book, or the gender of the author (Wang et al. [2019], Lassen et al. [2022]), as well as text-intrinsic features. Especially text-intrinsic features continue to pose a challenge and are explored variously because of the great range of textual features that can be possibly considered. Moreover, the perceived quality of a literary work is arguably the result of the interplay of many elements at different levels of literary texts (style, plot, topics), where some are virtually impossible to study by computational methods (e.g. metaphors or images). Predominantly, studies that seek to predict a kind of "literary quality" from textual features have relied on a set of classical stylistic features, such as sentence-length (Koolen et al. [2020], Maharjan et al. [2017]), percentages of word classes (Koolen et al. [2020]) or n-gram frequencies of texts (van Cranenburgh and Koolen [2020]). Methods focusing on series' dynamics have also been applied to such stylistic features: Mohseni et al. [2021], for example, conducted fractal analysis on classical stylistic features, showcasing the potential of this method for telling between canonical and non-canonical literary texts, which appear to exhibit different dynamics at the stylometric level.

More recent work has tested the potential of alternative features of the text or narrative, accessed through sentiment analysis (Alm [2008], Jain et al. [2017]). Sentence- or paragraph-level sentiment analysis of a literary text provides a simple and intuitive representation of a narrative's shape, and is applied as a proxy for meaningful aspects of the reading experience (Drobot [2013], Cambria et al. [2017], Kim and Klinger [2018], Brooke et al. [2015], Jockers [2017]). The resulting representation is referred to as a sentiment arc and has been used in a range of studies that model and evaluate narratives in terms of literary genre (Kim et al. [2017]), plot archetypes and their archetypes dynamics (Reagan et al. [2016]), and lastly, reader preferences and perceived quality (Bizzoni et al. [2022]). In his seminal work, Matthew (Jockers [2017]) paved the way for using rule-based sentiment analysis to extract story arcs. While Jockers' work

was criticized on a number of key points (annieswafford [2015]), it empowered literary scholars and narratologists with tool that could, among other things, estimate abstract genre categories, e.g., comedy vs. tragedy. Following this early work, several studies used sentiment analysis combined with several smoothing and filtering techniques for narrative extraction. (Reagan et al. [2016]) used a custom dictionary to show that a small set of 'story shapes' underlies the perceived heterogeneity of literature. (Jianbo Gao et al. [2010]) introduced the use of adaptive filtering in order to improve the accuracy and interpretability of story arcs, which provided a theoretical framework for studying the story arc *fractality* and linking it to fundamental dynamic properties of narratives (Gao et al. [2016]).

2.2 Fractality

There are several aspects of a sentiment arc that can be quantified. While the most popular until now has probably been its "overall" narrative shape or flow (Reagan et al. [2016], Maharjan et al. [2018], Kim et al. [2017], Mohammad [2011]), few studies have examined potential link between the *progression* or *patterns* of arcs and how such narratives are perceived. The study of fractals (Mandelbrot and Ness [1968], Mandelbrot [1982, 1997]), mainly applied to long series (Beran [1994], Eke et al. [2002], Kuznetsov et al. [2013]) offers a new way of looking into the properties of narrative and literary texts, exploring their degree of predictability or self-similarity ([Cordeiro et al., 2015]), which follows links with fractal properties already found in visual arts and music (McDonough and Herczyński [2023], Brachmann and Redies [2017]).

2.3 Literary quality

Literary quality is a disputed issue, in which the role of text internal and text external factors are up for debate. Moreover, the definition literary quality is complicated by the diversity of individual reader preferences or types (Riddell and van Dalen-Oskam [2018]), and individual readers may certainly also change their opinion about a text (Harrison and Nuttall [2018]). The idea that readers' perception of what is pleasant or engaging could be found in complex statistical patterns has given rise to a series of attempts to approach literary quality using quantitative models that examines a wide range of textual features (Archer and Jockers [2017], van Cranenburgh and Bod [2017], Maharjan et al. [2017], Wang et al. [2019]). Yet one main challenge for such studies is that of defining an "oracle" of quality itself, for which there may be no one perfect proxy. As proxies of literary quality, users' book ratings on the GoodReads platform are especially popular (Jannatus Saba et al. [2021], Maharjan et al. [2017], Porter [2018]), while other studies have defined quality via canon-lists (Mohseni et al. [2022]), sales figures (Wang et al. [2019]), or by whether or not books have won literary awards (Febres and Jaffe [2017]). While literary awards and GoodReads' ratings may be seen as representatives of the two opposite ends of the spectrum – one representing the taste of the (often academic) few, and the other that of many readers – it has also been shown that such quality or canonicity proxies may exhibit large overlaps (Walsh and Antoniak [2021b]).

The issue of choosing a suitable proxy for assessing literary quality mirrors the predicament that individual readers face when confronted with a vast volume of literature to read and evaluate on their own – a situation that may be seen as the reason for the creation of literary awards in the first place (Underwood [2019]). Yet few endeavors have been made, to the best of our knowledge, to employ high level literary awards as a primary metric to approximate a work's quality, where most studies rely on popular votes as represented by GoodReads' ratings. In the present study, we attempt to utilize the most prestigious global literary award, the Nobel Prize for Literature, as an exclusive criterion for selecting works of high literary quality. We then test the potential of the Hurst exponent for distinguishing between more or less popular works as

represented by GoodReads' rating within this already highly prestigious canon.

III DATA

3.1 A corpus of Nobel laureates

It is essential to recognize that the approach of choosing a prestigious literary prize as a quality proxy represents a purposeful extremization, as no literary award can be considered an all-encompassing oracle of quality. Controversies regarding the Nobel committee's selection, both in terms of recipients and non-recipients of the prize, arise almost annually (Duval [2005]). The Nobel Prize for Literature has been variously criticized: for being arbitrary or political (Epstein [2012]), and for gendered or Eurocentric biases of the appointing committee of (Lindfors [1988]). However, prestigious literary awards can serve as imperfect markers for a specific type of quality, and consequently, it would be intriguing to ascertain whether a "signal" that distinguishes Nobel-winning texts from a control group can be identified on a broader scale than that of individual books or authors. In this sense, the Nobel Prize in Literature remains a significant indicator of a certain calibre of literary accomplishment, given its international recognition and stringent selection process, and it offers a concrete and quantifiable standard for literary quality. As such, we seek to explore the possible correlations between some aspects of literary texts and their literary quality by using Nobel laureates' works as a benchmark, using the dynamics of the novels' sentiment arcs to evaluate their overall narrative structures, and we establish a control group composed of non-Nobel-winning literature to compare and contrast the findings. We do not claim that the Nobel Prize is the ultimate measure of literary quality, but we do consider it a convenient, albeit imperfect, tool for assessing a certain standard of literary excellence in a quantifiable manner, especially when it is further filtered, as in our case, by high representation of the laureates' titles in libraries. Ultimately, our intent is to explore whether the sentimental trajectories of narrative texts recognized by the Nobel committee reveal any specific patterns that distinguish them from other works, thereby contributing to a more nuanced understanding of what might constitute "literary quality".

Unfortunately, a comprehensive and digitized collection of works by Nobel-winning authors has not yet been established. To conduct our investigation, we relied on a contemporary corpus of literary texts, the Chicago Corpus, assembled by Hoyt Long and Richard Jean So. This corpus encompasses 9,089 novels published in the United States between 1880 and 2000, and was compiled on the basis of the number of library holding each title worldwide, favoring higher holding-numbers. The corpus features significant works by U.S. Nobel laureates, influential pieces from mainstream literature, and notable contributions to the wide spectrum of the so-called "genre literature", from Mystery to Science Fiction (Long and Roland [2016]).[1]

The Chicago Corpus serves as a valuable resource for our study, as it encompasses an extensive range of literary works, allowing for a more robust analysis of the relationship between sentiment arcs and literary quality. All in all, it offers an expansive and representative sample of the Anglophone literary scene over a century. By including Nobel Prize-winning texts alongside other notable pieces of literature, we aim to identify any discernible patterns or characteristics that may distinguish these esteemed works from their counterparts within the broader literary landscape.

As shown in Table 1, the laureate group comprises 18 authors and 85 titles, while the control group compasses 738 authors and 1312 titles. Both the laureate group and the control group

[1] Numerous quantitative literary analyses have employed this corpus, ([Underwood et al., 2018, Cheng, 2020]), which can be inspected at https://textual-optics-lab.uchicago.edu/us_novel_corpus.

	N. Authors	N. Titles
Whole corpus	7000	9089
Nobel group	18	85
Control group	738	1312

Table 1: Overall titles and authors in the corpus, number of Nobel laureates and dimensions of the control group.

offer a broad range of American and non-American authors, enabling an extensive exploration of their possible correlations with textual features. However, the U.S. Nobel laureates represented in the corpus constitute a significant portion of the overall group of laureates, including authors such as John Galsworthy, Sinclair Lewis, William Faulkner, Ernest Hemingway, John Steinbeck, Saul Bellow, and Toni Morrison, while the selection of works by non-U.S. writers, such as Knut Hamsun, Samuel Beckett, and Nadine Gordimer, remains more limited.

It is important to note that the corpus has been meticulously curated, and comprises high-quality fiction from authors who have received other accolades, such as the National Book Award, including Don DeLillo, Joyce Carol Oates, and Philip Roth; as well as important works of genre-fiction, such as by Tolkien or Philip K. Dick. As such, we do not anticipate that the works of Nobel laureates will be entirely distinct from the rest of the corpus in terms of literary quality.

It is also worth noting that the corpus predominantly consists of Anglo-Saxon literature, with both the Nobel laureates and their control group primarily comprising Anglophone writers. This bias inevitably situates the entire analysis within the context of a well-defined "Anglocentric" canon. While this imbalance does not inherently undermine our experiments nor jeopardize our analysis, as it is crucial to bear it in mind when interpreting the results and caution should be exercised when extrapolating the findings to the context of a global literary field.

3.2 GoodReads

Beyond the Nobel prize, for the last part of our experiments, we used GoodReads' average scores (see Section 4.3). Launched in 2007, GoodReads is a popular social media platform designed for book lovers and readers to discover, review, and share their thoughts. With its 90 million users, GoodReads offers a valuable insight into reading culture "in the wild" (Nakamura [2013]), cataloging books from various genres (Walsh and Antoniak [2021a]), and deriving ratings from a heterogeneous pool of readers in regard to their background, gender, age, native language and reading preferences (Kousha et al. [2017]). Goodreads' scores represent the average rating given by the users who rated a particular book. These scores range from 0 stars (indicating low appreciation) to 5 stars (indicating high appreciation). The average score provides a general indication of the book's reception, while conflating types of literary appreciation, i.a., satisfaction, enjoyment, and evaluation. While it is crucial to acknowledge that these scores do not present an absolute measure of literary quality either, they offer a valuable perspective on a work's overall popularity among a diverse population of readers. Therefore, Goodreads' scores provide a complementary lens through which we can examine the potential relationship between the dynamics of sentiment arcs and the broader popularity of a book within the reading community.

Few Nobel titles have an average GoodReads' rating above 4 (fig. 1), and none above 4.3. The mean rating for Nobel-titles is 3.8, and the majority of Nobel titles have a GoodReads' rating

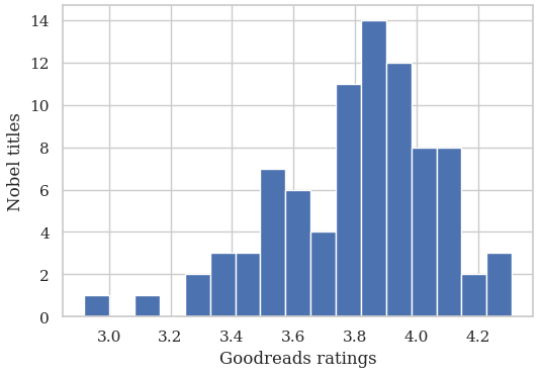

Figure 1: Distribution of GoodReads' ratings in our corpus of Nobel titles.

between 3 and 4 (75%). Consistent with the status of these works, GoodReads' ratings are relatively high for the platform. [2]

3.3 Sentiment lexicon

We rely on the VADER (Valence Aware Dictionary and sEntiment Reasoner) model and its sentiment lexicon (Hutto and Gilbert [2014]) in NLTK's implementation (Bird [2006]) to derive the sentiment arcs of the novels (see also Section 4.1). We chose VADER's sentiment lexicon for its popularity, as it has proven the weapon of choice for a large number of previous works. VADER is specifically designed to analyze sentiment in both standard and sub-standard contemporary written English language - including informal expressions, casual vernacular and slang - a characteristic that makes it apter for contemporary fiction than for more specialized domains. This characteristic sets it apart from other models that may struggle with more informal or idiomatic language, which is frequently present in literary texts. Additionally, VADER's lexicon accounts for various linguistic functions, such as negations, intensifiers, and degree modifiers, which contributes to a more refined analysis of the text and enhances the depth and precision of the text analysis, enabling a more nuanced understanding of the emotional trajectories present in the novels. Finally, its computational efficiency and ease of use make it an appealing choice to work with relatively large corpora. Therefore, by employing VADER for sentiment analysis, we aim to offer an accurate and comprehensive evaluation of the sentiment arcs within our chosen corpus, shedding light on any potential correlation between these arcs and perceived literary quality.

IV METHODS AND EXPERIMENTS

4.1 Sentiment arcs

A sentiment arc typically refers to a simple 1d representation of a literary work's sequentially dependent slices (ex., sentences or paragraphs). The arc is extracted using either rule- or learning-based tools for sentiment analysis and classification. Because narratives and their derived arcs are inherently noisy and nonlinear, Gao et al. [2016] suggested applications of filtering techniques for noise reduction in order to extract the global narrative trends. Wavelet approaches typically used for noise reduction, are, however not ideal for detrending nonlinear series. Instead, Jianbo Gao et al. [2010] developed a nonlinear adaptive filtering technique, which is more effective than wavelet approaches for determining trends in nonlinear time series. Several studies have demonstrated the usefulness of applying adaptive filtering to sentiment

[2] We might compare this to a 2018 dataset of nearly 2M GoodReads' ratings (Jannesar [2018]) In this corpus, mean rating is significantly lower: 2.9.

arcs, especially in the context of estimating an arc's fractality (Hu et al. [2021]).

4.2 Dynamic properties

We refer to a story arc as fractal if it displays self-similarity (i.e., sentiment fluctuation patterns at faster time scales resemble sentiment fluctuation patterns at slower time scales) and scale-invariance (i.e., measurement of the sentiment fluctuations is independent of time scale) (Riley et al. [2012]). To determine if a story arc is fractal, we examine the relationship between the time scale of fluctuation patterns and their measurement. Detrended fluctuation analysis (DFA) is the most widely used technique for estimating the Hurst parameter (Kantelhardt et al. [2001]), which provides information about this relationship. Assuming a stochastic process $X = X_t : t = 0, 1, 2, ...$, with stable covariance, mean μ and σ^2, the arc's autocorrelation function for $r(k), k \geq 0$ is:

$$r(k) = \frac{E\left[X(t)X(t+k)\right]}{E\left[X(t)^2\right]} \sim k^{2H-2}, \text{ as } k \to \infty$$

where H is the Hurst exponent that can be interpreted using the following heuristic: For $H > 0.5$ the arc is fractal, that is, characterized by long-range temporal correlations such that increases are followed by further increases and decreases by decreases. For $H = 0.5$ the time series only has short-range correlations; and when $H < 0.5$ the time series is anti-persistent such that increments are followed by decreases and decreases by increments. Elsewhere it has been argued that Hurst exponent provides an index of how coherent a literary work's moods, feelings, and attitudes will be perceived by a reader, and by extension, a measure of perceived quality (Hu et al. [2021]). Furthermore, the 'good' novel should display fractality (i.e., $H > 0.5$), but it should avoid being highly fractal ($H \sim 1$). The fundamental intuition behind this argument is that a good reading experience is facilitated by a narrative that is neither too predictable ($H \sim 1$) nor too unpredictable ($H = 0.5$).

While DFA is a popular choice, it can result in discontinuities at the boundaries of adjacent arc slices (e.g., sentences). Such discontinuities can be detrimental when the arc contains trends (Hu et al. [2001]), non-stationarity (Kantelhardt et al. [2002]), or nonlinear oscillatory components (Chen et al. [2005], Hu et al. [2009]). Adaptive fractal analysis (AFA), which is used in this paper, combines adaptive filtering (Jianbo Gao et al. [2010]) with fractal analysis in order to provide a more robust alternative to DFA (Gao et al. [2011b], Tung et al. [2011], Gao et al. [2016]). AFA consists of the following steps: first, the original process is transformed to a random walk process through first-order integration $u(n) = \sum_{k=1}^{n}(x(k) - \overline{x}), n = 1, 2, 3, ..., N$, where $\overline{x}$ is the mean of $x(k)$. Second, we extract the global trend $(v(i), i = 1, 2, 3, ..., N)$ through the nonlinear adaptive filtering. The residuals $(u(i) - v(i))$ reflect the fluctuations around a global trend. We obtain the Hurst parameter by estimating the slope of the linear fit between the residuals' standard deviation $F^{(2)}(w)$ and w window size.

A $1/f$ fractal process has a power-law decaying spectral density and it can therefore not be adequately modelled by standard techniques for time series analysis, such as an ARIMA model or a Markov process, because they have distinctly different spectral densities. A $1/f^{2H+1}$ process where $0 < H < 1$ is a non-stationary random-walk process, the differentiation of which is a covariance stationary stochastic process with mean μ, variance σ^2, and autocorrelation function (Cox [1984]).

To estimate the long-term memory of sentiment arcs we combine non-linear adaptive filtering with fractal analysis in adaptive fractal analysis or AFA (Gao et al. [2011a], Tung et al. [2011]).

4.3 Quality judgments

As we also discuss in Section II, defining optimal data for quality is an open challenge. In this series of experiment we use a double take to approximate the perception of literary quality: the Nobel Prize in Literature on one side (heavily expert-based); and GoodReads' ratings on the other (completely crowd-based). As we mentioned before, the Nobel Prize in Literature is the world's most renowned literary prize, which is being awarded based on literary quality as the members of the Swedish Academy define it. Their selection is based on nominations from hundreds of academics from around the world and the selection of winners take place after a careful reading of the works of at least five shortlisted authors. While there has been much debate on whether individual candidates have been worthy winners, there is no doubt about that the Nobel laureates are generally highly recognized by experts in literary criticism. GoodReads was launched in 2007 and is the world's largest database of literary reviews and is very indicative of both the amount of interest there is in single works and authorships and how a wide variety of lay readers value them. As such, GoodReads is complementary to the list of Nobel laureates, and the average score also provide for a more nuanced picture of valuation in contrast to the binary system of being awarded a Nobel prize or not.

Another difference between the list of Nobel laureats and the GoodReads scores is that the Nobel laureates have been selected over a period of more than 100 years, while the GoodReads scores have been compiled since 2007 and hence reflects contemporary readers perspective on literary preference and hence perceived quality. The many reviews that supplement the scores give insight into criteria of valuation but it is outside of this study to take these into account.

4.4 Predicting a Nobel

We report the findings of two experiments conducted in this study:

1. Instead of directly assessing the predictive capability of narrative sentiment arcs and their Hurst exponent, we examined their distribution in both Nobel-winning and non-Nobel-winning populations to determine if the two groups exhibit differences in average Hurst value.
2. To directly evaluate the predictive power of the Hurst exponent, we implemented a series of classifiers to ascertain if sentiment arcs and their associated Hurst scores could offer some level of predictive accuracy in determining whether a given text is likely authored by a laureate.

In both instances, we carefully designed the non-Nobel-winning class (or control group) to closely resemble the context of the Nobel population. For each book written by a Nobel Prize-winning author, we selected all novels published between one year before and one year after its publication date, considering these as the "control group" for that particular book. The control groups for all books by a single author collectively function as the control group for that author, while all control groups combined form the overall control group for the Nobel laureate population. Our rationale for this approach aligns with the Nobel Prize's selection method, which considers contemporary candidates. Table 2 offers a detailed overview of this process, elucidating our methods of control group selection and comparison.

4.4.1 Probability distribution

In the first experiment, our primary focus was to determine whether the Nobel-winning population exhibited a distinct Hurst score distribution compared to the control group, and whether

Nobel	N. titles	Control
S. Beckett	1	32
S. Bellow	5	228
W. Churchill	4	125
W. Faulkner	15	332
J. Galsworthy	9	105
W. Golding	2	6
N. Gordimer	2	3
K. Hamsun	1	1
E. Hemingway	7	170
R. Kipling	3	19
D. Lessing	3	34
S. Lewis	8	137
T. Morrison	5	192
A. Munro	1	2
J. Steinbeck	15	81
R. Tagore	1	19
S. Undset	2	32
P. White	1	0
Total	85	1518

Table 2: Number of titles per Nobel and control group. Notice that the control group's total number is higher than the one reported in Table 1 since one title can figure in more than a subgroup.

this difference was statistically significant on a large scale. To further investigate this hypothesis, we categorized our corpus into Hurst classes (e.g., all titles with Hurst scores of 0.51, 0.52, etc.) and analyzed the likelihood of encountering a title from a Nobel laureate within each of these classes.

Given the challenge of working with a heavily imbalanced dataset, as the number of control authors far exceeds Nobel-winning authors in any given class, we calculated probabilities using a sub-sampled portion of the control group equivalent in size to the Nobel-winning group. This ensured that both populations had an equal total count. To mitigate the impact of potential random variations in sub-sampling from the majority class and to enhance the representativeness of our comparison, we repeated the random majority class sub-sampling 100 times and calculated the average probability for each Hurst class. Consequently, for each Hurst class, we computed the probability of encountering a title by a Nobel laureate and the average probability of encountering a title by a control author, as determined from multiple subsamples.

4.4.2 Machine Learning

In the second experiment, we trained four different classifiers:

- **Quadratic Discriminant Analysis** classifier (Bose et al. [2015]): a generative model that is particularly apt to classify data when the decision boundaries are non-linear;
- **Gaussian Naive Bayes** classifier (Chan et al. [1982]): we chose this model particularly for its ability to handle small and complex training data;
- **Random Forest** classifier (Ho [1995]): this algorithm is well suited to make fine-grained predictions on data that are not necessarily linearly divisible;
- **Decision Tree** classifier, which has the benefits of being simple and able to handle relatively small datasets (Swain and Hauska [1977]).

As features, we used the Hurst score and a condensed version of the sentiment arc for each novel.

The large difference in our classes' sizes represents an additional difficulty. The sparsity of Nobel titles makes training on the dataset as is a seemingly meaningless task, since classifiers systematically ignore or misrepresent the minority class. To contrast that dataset's imbalance, we tried three resampling techniques:

- **Random** subsampling: this is the easiest resampling technique, and it simply means that we randomly drew from the majority class a number of data points equal to the size of the minority class, as we did in Section 5.1;
- **Near Miss** subsampling ([Mani and Zhang, 2003, Bao et al., 2016]), specifically the so called *Near-Miss 1* method: this is a more sophisticated undersampling technique based on the distance between items from the majority and items from the minority class, where the elements from the majority class with the smallest average distance to three minority class examples are selected for comparison. In this way, the algorithm selects datapoints that are closest to the decision boundary;
- **SMOTE** upsampling (Chawla et al. [2002]), a upsampling technique widespread in machine learning, often used in cases of severely imbalanced datasets ([Liu et al., 2019, Rustogi and Prasad, 2019]). SMOTE has the considerable benefit of creating not simple duplicates of the observed datapoints, but rather slightly different synthetic datapoints, increasing the ability of a classifier of modeling a minority class.

4.5 Telling between Nobels

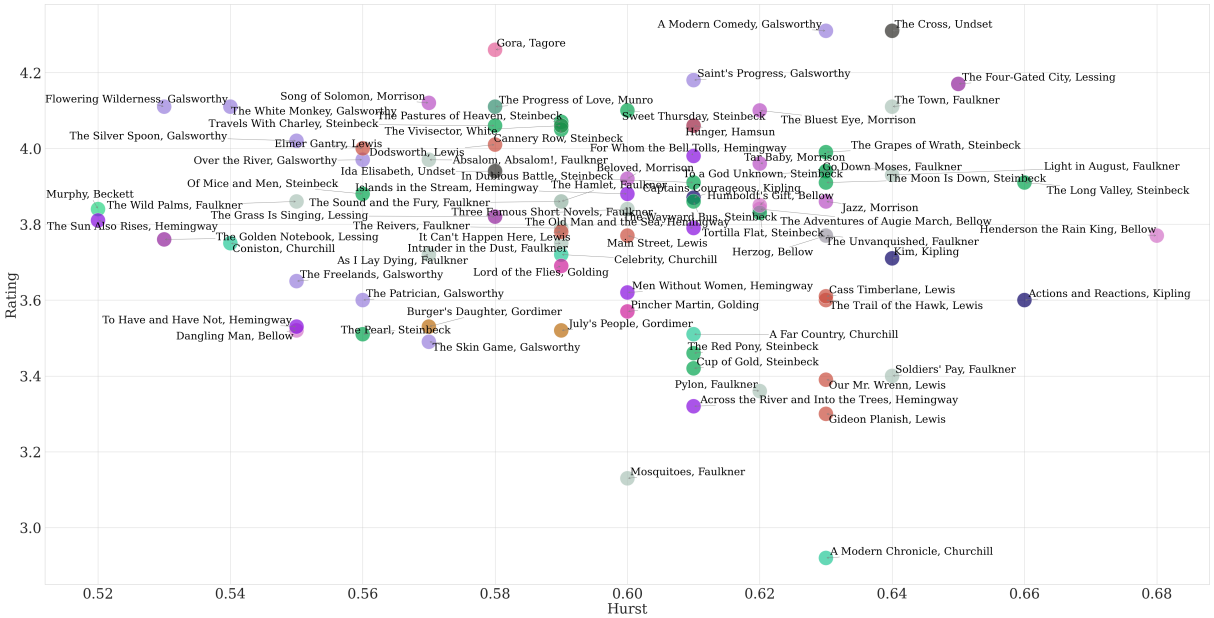

Figure 2: GoodReads' rating for Nobel titles at given values of Hurst. Titles by the same authors have the same colors.

In our third experiment, we sought to assess whether Hurst may be used to discriminate between works more or less popular (based on GoodReads) within the Nobel canon itself. Studies using GoodReads ratings as proxies of quality often frame the task as a classification problem and introduce a threshold in GoodReads' rating that distinguishes successful from unsuccessful books (Maharjan et al. [2017]). However, rather than setting a more or less arbitrary threshold, our idea was to examine whether the Hurst itself may be used to distinguish such a cut-off.

Considering the difficulty of assessing the discriminating potential of Hurst in such a small corpus, we also opted against using machine learning. Yet we may still assess whether there are patterns to be observed in the distributions (fig. 2). For assessing the probability of Nobel titles' high or low rating at their given Hurst values, we defined thresholds for the "low" and "high" rating groups as the points where the difference in terms of mean Hurst exponent for each group was the most significant according to an independent t-test. Essentially, the idea was to find the point at which Hurst was best at discriminating between high and low rating groups, and to use this point as the threshold for dividing high- from low-rating titles, and then to examine the mean Hurst value of each group.

V RESULTS

5.1 Probability distribution

The difference between the distributions of Hurst scores for the Nobel and the control group is statistically significant according to several measures, as can be seen in Table 3.

	Score	**p-value**
T-test	2.57	0.01
Anova	6.63	0.01
Mann-Whitney U	55106	0.023
Kruskal-Wallis	5.166	0.023

Table 3: Difference between Nobel laureates and control group as tested by four significance measures (the first two assume that the populations have a normal distribution, the last two do not make such assumption). In all cases, the difference in Hurst score distributions is statistically significant.

A cursory qualitative examination of the results for different authors proved that these results often (but not always) correspond to what we might expect from a given title or author. For example John Steinbeck, one of the best represented writers in the corpus with 15 novels, has an average Hurst exponent of 0.598, and thus differs insignificantly from the 90 works in its control group, that score an average of 0.606, but with a more significant standard deviation (0.41 vs. 0.25). While Steinbeck's novels Hurst scores range from 0.56 to 0.64, the two novels that get by far the highest average grades on GoodReads (*Mice and Men* and *The Grapes of Wrath* with *Cannery Row* as a very distant third) both have a Hurst exponent of exactly 0.58, at the apex of the probability curve for Nobel titles. Similar observations can be made for the works for other popular Nobel laureates, such as Hemingway, with his most renowned titles (such as for example *The Old Man and the Sea* or *For whom the bell tolls*) roughly falling within what we considered a fuzzy Goldilocks interval for literary quality, while less acclaimed texts such as *To have and have not* are clearly out of it (Figure 1). Many other factors play into the success of these prominent novels, but their location in the middle of what seems to be a "Goldilocks"-zone for variability is significant, also when studied on the level of the individual authorship.

The probability of encountering a text by a Nobel laureate peaks at a different point compared to the probability of encountering a text from the control group (refer to Figure 1). The distribution of the two groups supports the hypothesis proposed by (Hu et al. [2021]) that high literary quality may be concentrated within a specific region of the Hurst continuum. In other words, there could be a particular range of Hurst values where high-quality narrative texts are most likely to be found. Of course, it is important to recognize that the probability distributions of the two groups significantly overlap; the statistical significance, while robust, does not indicate

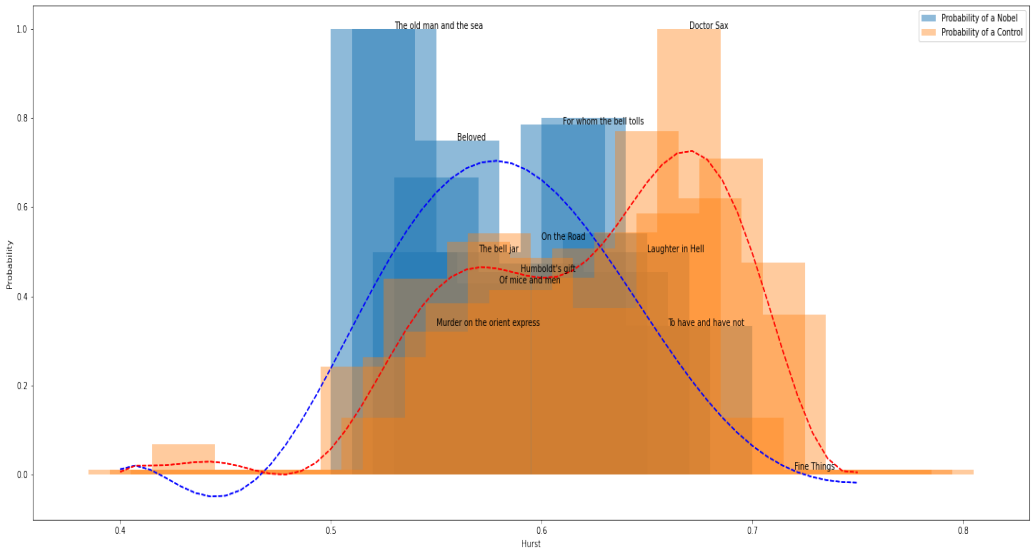

Figure 3: Probability distribution of the Nobel group and the control group. The control population's probabilities are averaged over 100 different selections. We have included some reference titles for context. It is important to note that not all works by Nobel laureates fall within the Hurst sweet spot": for instance, *The Old Man and The Sea* has a Hurst score of 0.53, while the less acclaimed *To Have and Have Not*, written by the same author, has a Hurst score of 0.69.

complete separability between the groups.

Due to their frequency, the number of control titles exceeds the number of titles by Nobel-winning authors in any given Hurst interval. In essence, any given text has a lower probability of being authored by a Nobel laureate than by an author who has not won the award, given that the Nobel Prize can only be conferred upon a single individual each year. However, if we consider equal-sized classes for the two groups, texts with Hurst scores ranging approximately between 0.53 and 0.61 appear to have a higher probability of being written by a Nobel laureate than by a control author. Conversely, texts falling outside this range have a higher probability of being authored by a control writer than by a Nobel laureate. This observation further emphasizes that the Nobel-winning and control populations exhibit statistically different behaviors along the Hurst continuum. Figure 1 provides a visual representation of our findings.

5.2 Classification

Among the three techniques we adopted to resample our dataset, we found that randomly under-sampling the majority class does not yield particularly strong results, while Near Miss under-stampling and SMOTE oversampling both bring the models to better performances (see Figure 2). The reason for this lies probably in the fact that the difference between the two populations, while present, is quite difficult to pick up even when we control for size: after all, we are using a corpus with a large number of high quality authors that did not win a Nobel prize, so the control group is both much larger than the Nobel group and bound to have several elements similar to its members. Just randomly subsampling from the majority class to create a small group of non-Nobels to learn from makes the task very difficult, while an algorithm like Near Miss, that selects data with the least distance to the negative classe's samples, essentially selecting

learning cases that is most fruitful for the classifier to model, brings significantly better results. Finally, it's worth noting how SMOTE upsampling brings about the highest performances of the group (excluding the "All dataset" case): while this technique does not create completely dependable results, since it relies on the synthetic generation of new data points for the minority class, its effectiveness can make us more confident in postulating that a difference between the Nobel and the control populations does indeed exist.

Table 4 presents a summary of the classifier performances, with the performances of the classifiers using only sentiment arc information (without the Hurst exponent) provided in parentheses. This comparison is intriguing: the sentiment arcs alone appear to generate better-than-chance performances and, in some instances, even achieve quite high scores. Moreover, classifiers trained on a feature set that includes the Hurst exponent of the arcs consistently outperform those without access to such information.

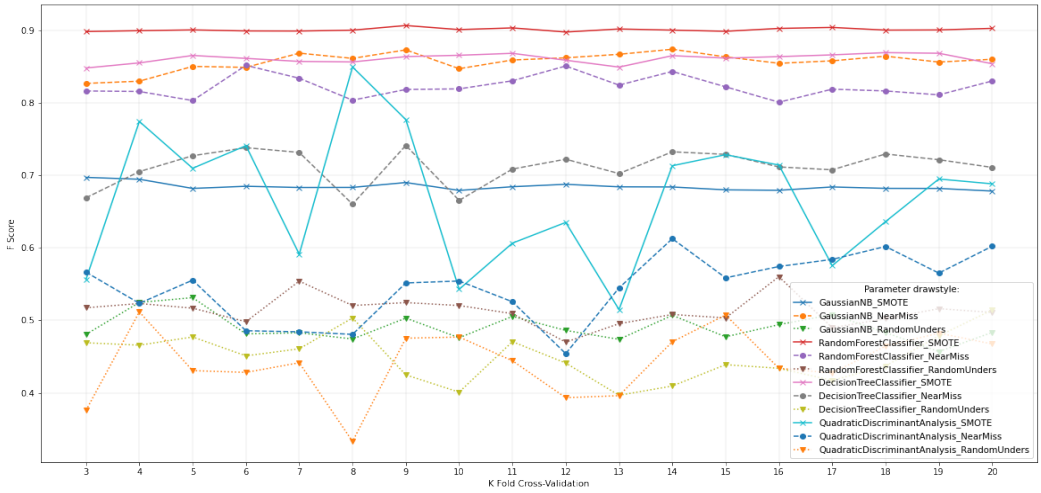

Figure 4: Classification results for our 4 classifiers under three different assumptions: random undersampling, Near Miss undersampling and SMOTE upsampling, with increasing number of folds in a K-folds cross-validation.

	Original dataset	**Random Subs.**	**Near Miss**	**SMOTE Ups.**
Quadratic Discr. An.	0.90 (0.90)	0.55 (0.51)	0.56 (0.51)	0.57 (0.50)
Gaussian Naive Bayes	0.91 (0.90)	0.52 (0.49)	0.80 (0.67)	0.67 (0.53)
Decision Tree Cl.	0.88 (0.88)	0.57 (0.52)	0.69 (0.60)	0.87 (0.82)
Random Forest	0.91 (0.90)	0.53 (0.51)	0.79 (0.62)	**0.90 (0.86)**
Average	0.90 (0.89)	0.53 (0.50)	0.71 (0.60)	0.75 (0.67)

Table 4: Weighted F scores, averaged from a 10-fold cross-validation, for four classifiers trained on different versions of the dataset. Notice how the results on the "all dataset" column are effects of the majority class being overwhelmingly larger than the minority class. In parenthesis, we add the performances when not using Hurst. The other three columns, reporting results based on resampled versions of the dataset, do not resent of the distortion.

5.3 Some Nobels are nobler than others

In our second experiment, looking at more or less successfull works within the Nobel canon, we used independent t-tests to determine thresholds at which the average Hurst of each group (high/low rating-titles) was most distinct. Independent t-tests were performed for each possible rating thresholds, however, only those thresholds where the differences in mean between the two groups was statistically significant ($p < 0.05$) are visualized (fig. 5). We then calculated the probability that each work at a given Hurst would have a high or a low rating, considering all of the low/high rating thresholds where the t-test show a statistically significant difference in mean Hurst.[3] Considering these plots (fig. 5), it appears that titles with a higher Hurst are more likely to have a lower rating.

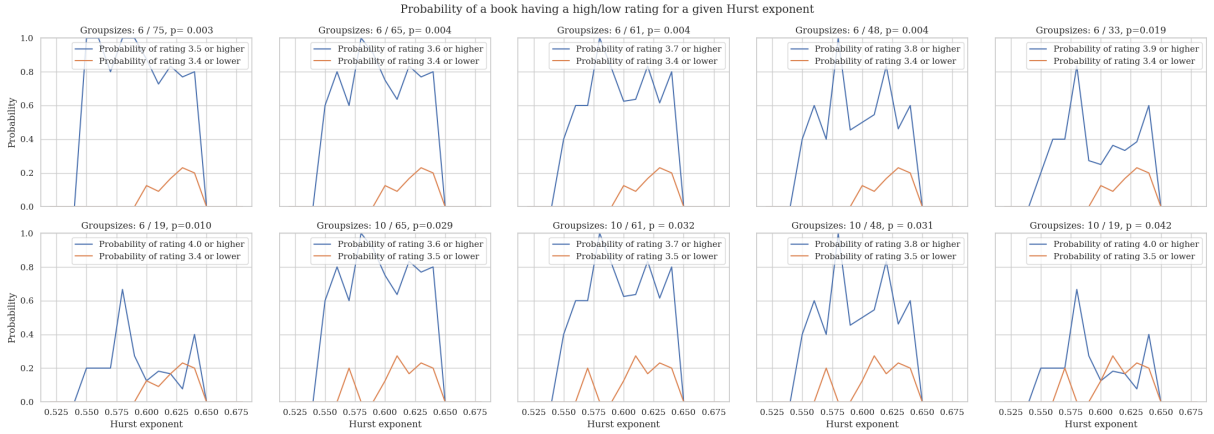

Figure 5: Probability of high and low-rating titles at certain values of the Hurst exponent. Each plot shows a different selection of thresholds for distinguishing high and low-rating groups. At the top of each plot are noted: group sizes at the given thresholds and p-values of the t-test comparing mean Hurst of groups at the given thresholds.

The t-tests show that the average Hurst for each group is always most significantly different when low rating is defined as ratings below 3.4, and high ratings as ratings above 3.5 (up until 4.0). The thresholds at which Hurst groups are most distinct based exclusively on Hurst value, and consequently, where Hurst can most strongly discriminate between high and low-rating groups is where the high rating group is defined as above 3.5, and the low rating group as below 3.4. Here, the independent t-test shows a significant difference in mean Hurst between the groups ($t = 3.724$, $p = 0.003$)(table 5). At these thresholds, low-rating titles (rating$<$3.4) exhibit a lower mean Hurst. Considering the narrow range of possible Hurst values in our data ($0.525 < H > 0.675$), the difference of 0.025 points between the groups can still be considered notable.

	Avg. Hurst
All titles	0.597
Titles rated above 3.5	0.595
Titles rated below 3.4	0.620

Table 5: Mean Hurst exponent per group.

[3] Probabilities were calculated only for values of Hurst that were shared among at least 5 titles.

VI DISCUSSION

The results in this paper is based on a limited corpus of works, mainly due to the availability of digitized works by Nobel laureates. Even so, we were able to show a relation between literary prestige (defined as Nobel status) and Hurst value. This was also evident in the absence of works of prestige outside of the Hurst "Goldilocks zone". One should be careful about concluding much from inspection of individual authorships, yet in the cases of John Steinbeck and Ernest Hemingway, both represented with multiple works, there was a further indication of a correlation between Hurst score and canonicity. We find that this is another example of the explanatory power of dynamic sentiment arc that can be supplemented with other features in a more complex model for understanding reader preferences and literary prestige.

In our third experiment, which sought to use the Hurst exponent to distinguish more or less popular works *within* the Nobel canon, we found that titles with higher levels of Hurst ($0.595 < H < 0.65$) appear more likely to have a lower rating. This finding is similarly aligned with the results of Bizzoni et al. [2022], that also defined popularity by GoodReads' rating and indicated a Hurst "sweet spot" for more popular fairy tales with a Hurst below 0.58. These findings may indicate that sentiment arcs with a Hurst higher than 0.59 start to become excessively linear or predictable.

VII CONCLUSION AND FUTURE WORKS

Despite the limitations inherent in the corpus used, we have made advances in laying a foundation for a novel approach to understanding literary quality, that keep at the center the dynamic properties at the sentiment level of the literary work. We believe that a model that take into account how the literary work provides the conditions for a particular reading experience is necessary in order to better grasp why some works consistently are valued highly. While our exploratory approach cannot stand alone and has clear limitations, the relatively strong signal given by the Hurst rating in order to predict critical appraisal is an indication that it could be a strong building block in a more elaborate model for understanding perceived literary quality. Our hope is in the future to expand this work by incorporating a more diverse and representative array of global literary works. We will also build a model that includes other features, e.g. readability and topics, that can hopefully refine the fundament provided by the dynamic sentiment arcs.

References

Ebba Cecilia Ovesdotter Alm. *Affect in* text and speech*. University of Illinois at Urbana-Champaign, 2008.

annieswafford. Problems with the Syuzhet Package, March 2015. URL https://annieswafford.wordpress.com/2015/03/02/syuzhet/.

Jodie Archer and Matthew Lee Jockers. *The bestseller code*. Penguin books, London, 2017. ISBN 978-0-14-198248-9.

Lei Bao, Cao Juan, Jintao Li, and Yongdong Zhang. Boosted near-miss under-sampling on svm ensembles for concept detection in large-scale imbalanced datasets. *Neurocomputing*, 172:198–206, 2016.

Jan Beran. *Statistics for Long-Memory Processes*. Chapman and Hall/CRC, New York, 1 edition, October 1994. ISBN 978-0-412-04901-9.

Steven Bird. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, pages 69–72, 2006.

Yuri Bizzoni, Telma Peura, Kristoffer Nielbo, and Mads Thomsen. Fractal sentiments and fairy tales-fractal scaling of narrative arcs as predictor of the perceived quality of andersen's fairy tales. *Journal of Data Mining & Digital Humanities*, 2022.

Smarajit Bose, Amita Pal, Rita SahaRay, and Jitadeepa Nayak. Generalized quadratic discriminant analysis. *Pattern Recognition*, 48(8):2676–2684, 2015.

Anselm Brachmann and Christoph Redies. Computational and Experimental Approaches to Visual Aesthet-

ics. *Frontiers in Computational Neuroscience*, 11:102, November 2017. ISSN 1662-5188. doi: 10.3389/fncom.2017.00102. URL http://journal.frontiersin.org/article/10.3389/fncom.2017.00102/full.

Julian Brooke, Adam Hammond, and Graeme Hirst. Gutentag: an nlp-driven tool for digital humanities research in the project gutenberg corpus. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 42–47, 2015.

Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. Affective computing and sentiment analysis. In *A practical guide to sentiment analysis*, pages 1–10. Springer, 2017.

Tony F Chan, Gene H Golub, and Randall J LeVeque. Updating formulae and a pairwise algorithm for computing sample variances. In *COMPSTAT 1982 5th Symposium held at Toulouse 1982*, pages 30–41. Springer, 1982.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

Zhi Chen, Kun Hu, Pedro Carpena, Pedro Bernaola-Galvan, H. Eugene Stanley, and Plamen Ch. Ivanov. Effect of nonlinear filters on detrended fluctuation analysis. *Phys. Rev. E*, 71(1):011104, January 2005. doi: 10.1103/PhysRevE.71.011104. URL https://link.aps.org/doi/10.1103/PhysRevE.71.011104.

Jonathan Cheng. Fleshing out models of gender in english-language novels (1850–2000). *Journal of Cultural Analytics*, 5(1):11652, 2020.

João Cordeiro, Pedro R. M. Inácio, and Diogo A. B. Fernandes. Fractal beauty in text. In Francisco Pereira, Penousal Machado, Ernesto Costa, and Amílcar Cardoso, editors, *Progress in Artificial Intelligence*, Lecture Notes in Computer Science, pages 796–802. Springer International Publishing, 2015. ISBN 978-3-319-23485-4. doi: 10.1007/978-3-319-23485-4_80.

Timothy M. Cox. Long range dependence: A review. In *Iowa State University*. Press, 1984.

Irina-Ana Drobot. Affective narratology. the emotional structure of stories. *Philologica Jassyensia*, 9(2):338, 2013.

Alex Duval. A Nobel calling: 100 years of controversy. *The Independent*, October 2005. URL https://www.independent.co.uk/news/world/europe/a-nobel-calling-100-years-of-controversy-319509.html. Section: News.

A. Eke, P. Herman, L. Kocsis, and L. R. Kozak. Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, 23(1):R1, 2002. URL http://stacks.iop.org/0967-3334/23/i=1/a=201.

Joseph Epstein. The Nobel Prize for Political Literature. *Wall Street Journal*, October 2012. ISSN 0099-9660. URL http://online.wsj.com/article/SB10000872396390444799904578054821709524326.html.

Gerardo Febres and Klaus Jaffe. Quantifying literature quality using complexity criteria. *Journal of Quantitative Linguistics*, 24(1):16–53, January 2017. ISSN 0929-6174, 1744-5035. doi: 10.1080/09296174.2016.1169847. URL http://arxiv.org/abs/1401.7077. arXiv:1401.7077 [cs].

Jianbo Gao, Jing Hu, and Wen-wen Tung. Facilitating Joint Chaos and Fractal Analysis of Biosignals through Nonlinear Adaptive Filtering. *PLoS ONE*, 6(9):e24331, September 2011a. ISSN 1932-6203. doi: 10.1371/journal.pone.0024331.

Jianbo Gao, Jing Hu, and Wen-wen Tung. Facilitating Joint Chaos and Fractal Analysis of Biosignals through Nonlinear Adaptive Filtering. *PLoS ONE*, 6(9):e24331, September 2011b. ISSN 1932-6203. doi: 10.1371/journal.pone.0024331. URL http://dx.plos.org/10.1371/journal.pone.0024331.

Jianbo Gao, Matthew L Jockers, John Laudun, and Timothy Tangherlini. A multiscale theory for the dynamical evolution of sentiment in novels. In *2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC)*, pages 1–4. IEEE, 2016.

John Guillory. *Cultural Capital: The Problem of Literary Canon Formation*. University of Chicago Press, Chicago, IL, March 1995. ISBN 978-0-226-31044-2. URL https://press.uchicago.edu/ucp/books/book/chicago/C/bo3634644.html.

Chloe Harrison and Louise Nuttall. Re-reading in stylistics. *Language and Literature*, 27(3):176–195, August 2018. ISSN 0963-9470. doi: 10.1177/0963947018792719. URL https://doi.org/10.1177/0963947018792719. Publisher: SAGE Publications Ltd.

Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

Jing Hu, Jianbo Gao, and Xingsong Wang. Multifractal analysis of sunspot time series: the effects of the 11-year cycle and Fourier truncation. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(02):P02066, February 2009. ISSN 1742-5468. doi: 10.1088/1742-5468/2009/02/P02066. URL http://stacks.iop.org/1742-5468/2009/i=02/a=P02066?key=crossref.879d2c42ec8804831202df82da8d7a1a.

Kun Hu, Plamen Ch. Ivanov, Zhi Chen, Pedro Carpena, and H. Eugene Stanley. Effect of trends on detrended fluctuation analysis. *Physical Review E*, 64(1), June 2001. ISSN 1063-651X, 1095-3787. doi: 10.1103/PhysRevE.64.011114. URL https://link.aps.org/doi/10.1103/PhysRevE.64.011114.

Qiyue Hu, Bin Liu, Mads Rosendahl Thomsen, Jianbo Gao, and Kristoffer L Nielbo. Dynamic evolution of sentiments in never let me go: Insights from multifractal theory and its implications for literary analysis. *Digital Scholarship in the Humanities*, 36(2):322–332, 2021.

Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

Swapnil Jain, Shrikant Malviya, Rohit Mishra, and Uma Shanker Tiwary. Sentiment analysis: An empirical comparative study of various machine learning approaches. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*, pages 112–121, Kolkata, India, December 2017. NLP Association of India. URL https://aclanthology.org/W17-7515.

Syeda Jannatus Saba, Biddut Sarker Bijoy, Henry Gorelick, Sabir Ismail, Md Saiful Islam, and Mohammad Ruhul Amin. A Study on Using Semantic Word Associations to Predict the Success of a Novel. In *Proceedings of *SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics*, pages 38–51, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.starsem-1.4. URL https://aclanthology.org/2021.starsem-1.4.

Bahram Jannesar. Goodreads Book Datasets With User Rating 2M, 2018. URL https://www.kaggle.com/datasets/b2dde9353c9d10c36e4d6b593a74c109dbaca6393a1ca0f2c7abafeba7633641.

Jianbo Gao, H. Sultan, Jing Hu, and Wen-Wen Tung. Denoising Nonlinear Time Series by Adaptive Filtering and Wavelet Shrinkage: A Comparison. *IEEE Signal Processing Letters*, 17(3):237–240, March 2010. ISSN 1070-9908, 1558-2361. doi: 10.1109/LSP.2009.2037773. URL http://ieeexplore.ieee.org/document/5345722/.

Matthew Jockers. Syuzhet: Extracts sentiment and sentiment-derived plot arcs from text (version 1.0. 1), 2017.

Jan W Kantelhardt, Eva Koscielny-Bunde, Henio H. A Rego, Shlomo Havlin, and Armin Bunde. Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 295(3):441–454, June 2001. ISSN 0378-4371. doi: 10.1016/S0378-4371(01)00144-3. URL https://www.sciencedirect.com/science/article/pii/S0378437101001443.

Jan W. Kantelhardt, Stephan A. Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H. Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.

Evgeny Kim and Roman Klinger. A survey on sentiment and emotion analysis for computational literary studies. 2018.

Evgeny Kim, Sebastian Padó, and Roman Klinger. Investigating the Relationship between Literary Genres and Emotional Plot Development. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 17–26, Vancouver, Canada, August 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-2203. URL https://aclanthology.org/W17-2203.

Corina Koolen, Karina van Dalen-Oskam, Andreas van Cranenburgh, and Erica Nagelhout. Literary quality in the eye of the dutch reader: The national reader survey. *Poetics*, 79:101439, 2020.

Kayvan Kousha, Mike Thelwall, and Mahshid Abdoli. Goodreads reviews to assess the wider impacts of books. 68(8):2004–2016, 2017. ISSN 2330-1643. doi: 10.1002/asi.23805. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23805. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23805.

Nikita Kuznetsov, Scott Bonnette, Jianbo Gao, and Michael A. Riley. Adaptive Fractal Analysis Reveals Limits to Fractal Scaling in Center of Pressure Trajectories. *Annals of Biomedical Engineering*, 41(8):1646–1660, August 2013. ISSN 0090-6964, 1573-9686. doi: 10.1007/s10439-012-0646-9. URL http://link.springer.com/10.1007/s10439-012-0646-9.

Ida Marie Schytt Lassen, Yuri Bizzoni, Telma Peura, Mads Rosendahl Thomsen, and Kristoffer Laigaard Nielbo. Reviewer Preferences and Gender Disparities in Aesthetic Judgments. In *CEUR Workshop Proceedings*, pages 280–290, Antwerp, Belgium, 2022. URL https://ceur-ws.org/Vol-3290/short_paper1885.pdf.

Bernth Lindfors. Africa and the Nobel Prize. *World Literature Today*, 62(2):222–224, 1988. ISSN 0196-3570. doi: 10.2307/40143532. URL https://www.jstor.org/stable/40143532. Publisher: Board of Regents of the University of Oklahoma.

Shiyu Liu, Ming Lun Ong, Kar Kin Mun, Jia Yao, and Mehul Motani. Early prediction of sepsis via smote upsampling and mutual information based downsampling. In *2019 Computing in Cardiology (CinC)*, pages

Page–1. IEEE, 2019.
Hoyt Long and Teddy Roland. Us novel corpus. Technical report, Textual Optic Labs, University of Chicago, 2016. URL http://icame.uib.no/brown/bcm.html.
Suraj Maharjan, John Arevalo, Manuel Montes, Fabio A. González, and Thamar Solorio. A multi-task approach to predict likability of books. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1217–1227, Valencia, Spain, April 2017. Association for Computational Linguistics. URL https://aclanthology.org/E17-1114.
Suraj Maharjan, Sudipta Kar, Manuel Montes, Fabio A. González, and 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: Volume 2, Short Papers*, pages 259–265, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-2042. URL https://aclanthology.org/N18-2042.
Benoit Mandelbrot. *The Fractal Geometry of Nature*. Times Books, San Francisco, updated ed. edition edition, 1982. ISBN 978-0-7167-1186-5.
Benoit B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E*. Springer, New York, 1997 edition edition, September 1997. ISBN 978-0-387-98363-9.
Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437, 1968. ISSN 00361445. URL http://www.jstor.org/stable/2027184.
Inderjeet Mani and I Zhang. knn approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of workshop on learning from imbalanced datasets*, volume 126, pages 1–7. ICML, 2003.
John McDonough and Andrzej Herczyński. Fractal patterns in music. *Chaos, Solitons & Fractals*, 170: 113315, May 2023. ISSN 09600779. doi: 10.1016/j.chaos.2023.113315. URL https://linkinghub.elsevier.com/retrieve/pii/S0960077923002163.
Saif Mohammad. From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114, Portland, OR, USA, June 2011. Association for Computational Linguistics. URL https://aclanthology.org/W11-1514.
Mahdi Mohseni, Volker Gast, and Christoph Redies. Fractality and variability in canonical and non-canonical english fiction and in non-fictional texts. 12, 2021. ISSN 1664-1078. URL https://www.frontiersin.org/article/10.3389/fpsyg.2021.599063.
Mahdi Mohseni, Christoph Redies, and Volker Gast. Approximate entropy in canonical and non-canonical fiction. *Entropy*, 24(2):278, February 2022. ISSN 1099-4300. doi: 10.3390/e24020278. URL https://www.mdpi.com/1099-4300/24/2/278.
Franco Moretti. *Distant reading*. Verso Books, 2013.
Lisa Nakamura. "Words with friends": Socially networked reading on Goodreads. *PMLA*, 128(1):238–243, 2013. doi: 10.1632/pmla.2013.128.1.238.
J.D. Porter. *Stanford Literary Lab Pamphlet 17: Popularity/Prestige*. Stanford Literary Lab, 2018. URL https://litlab.stanford.edu/LiteraryLabPamphlet17.pdf.
Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. The emotional arcs of stories are dominated by six basic shapes. 5(1):1–12, 2016. ISSN 2193-1127. Publisher: SpringerOpen.
Allen Riddell and Karina van Dalen-Oskam. Readers and their roles: Evidence from readers of contemporary fiction in the Netherlands. *PLOS ONE*, 13(7):e0201157, July 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201157. URL https://dx.plos.org/10.1371/journal.pone.0201157.
Michael A. Riley, Scott Bonnette, Nikita Kuznetsov, Sebastian Wallot, and Jianbo Gao. A tutorial introduction to adaptive fractal analysis. *Frontiers in Physiology*, 3, 2012. ISSN 1664-042X. doi: 10.3389/fphys.2012.00371. URL http://journal.frontiersin.org/article/10.3389/fphys.2012.00371/abstract.
Rishabh Rustogi and Ayush Prasad. Swift imbalance data classification using smote and extreme learning machine. In *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, pages 1–6. IEEE, 2019.
Philip H Swain and Hans Hauska. The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, 15(3):142–147, 1977.
Wen-wen Tung, Jianbo Gao, Jing Hu, and Lei Yang. Detecting chaos in heavy-noise environments. *Physical Review E*, 83(4), April 2011. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.83.046210.
Ted Underwood. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019. ISBN 978-0-226-61297-3. doi: 10.7208/9780226612973. URL https://www.degruyter.com/

document/doi/10.7208/9780226612973/html. Publication Title: Distant Horizons.

Ted Underwood, David Bamman, and Sabrina Lee. The transformation of gender in english-language fiction. *Journal of Cultural Analytics*, 3(2):11035, 2018.

Andreas van Cranenburgh and Rens Bod. A data-oriented model of literary language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1228–1238, Valencia, Spain, April 2017. Association for Computational Linguistics. URL https://aclanthology.org/E17-1115.

Andreas van Cranenburgh and Corina Koolen. Results of a single blind literary taste test with short anonymized novel fragments. *arXiv preprint arXiv:2011.01624*, 2020.

Willie van Peer. Ideology or aesthetic quality? In Willie van Peer, editor, *The quality of literature: linguistic studies in literary evaluation*, pages 17–29. John Benjamins Publishing, Amsterdam ; Philadelphia, 2008.

Robert von Hallberg. Editor's Introduction. *Critical Inquiry*, 10(1):iii–vi, 1983. ISSN 0093-1896. URL https://www.jstor.org/stable/1343403. Publisher: The University of Chicago Press.

Melanie Walsh and Maria Antoniak. The goodreads 'classics': A computational study of readers, amazon, and crowdsourced amateur criticism. *Journal of Cultural Analytics*, 4:243–287, 2021a.

Melanie Walsh and Maria Antoniak. The Goodreads "Classics": A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism. *Post45: Peer Reviewed*, April 2021b. URL https://post45.org/2021/04/the-goodreads-classics-a-computational-study-of-readers-amazon-and-crowdsourced-amate

Xindi Wang, Burcu Yucesoy, Onur Varol, Tina Eliassi-Rad, and Albert-László Barabási. Success in books: Predicting book sales before publication. *EPJ Data Science*, 8(1):31, December 2019. ISSN 2193-1127. doi: 10.1140/epjds/s13688-019-0208-6.