Characterizing narrative time in books through fluctuations in power and danger arcs

Mikaela Irene Fudolig,∗ Thayer Alshaabi, Kathryn Cramer, Christopher M. Danforth, and Peter Sheridan Dodds

1Computational Story Lab, Vermont Complex Systems Center, MassMutual Center of Excellence for Complex Systems and Data Science, Vermont Advanced Computing Core, University of Vermont, Burlington, VT, USA
2Advanced Biomaging Center, UC Berkeley, Berkeley, CA, USA
3Department of Mathematics & Statistics, University of Vermont, Burlington, VT, USA
4Department of Computer Science, University of Vermont, Burlington, VT, USA

While recent studies have focused on quantifying word usage to find the overall shapes of narrative emotional arcs, certain features of narratives within narratives remain to be explored. Here, we characterize the narrative time scale of sub-narratives by finding the length of text at which fluctuations in word usage begin to be relevant. We represent more than 30,000 Project Gutenberg books as time series using osiometrics, a power-danger framework for essential meaning, itself a reinterpretation of the valence-arousal-dominance framework derived from semantic differentials. We decompose each book’s power and danger time series using empirical mode decomposition into a sum of constituent oscillatory modes and a non-oscillatory trend. By comparing the decomposition of the original power and danger time series with those derived from shuffled text, we find that shorter books exhibit only a general trend, while longer books have fluctuations in addition to the general trend, similar to how subplots have arcs within an overall narrative arc. These fluctuations typically have a period of a few thousand words regardless of the book length or library classification.

I. INTRODUCTION

Narrative dynamics have long been studied by scholars [1–6], mostly from a qualitative perspective, requiring human inputs to interpret text. While the world’s literary experts are far from being supplanted by computational analyses—and arguably never will be—their attentions cannot be scaled to study very large corpora. Recent advances in natural language processing, however, can predict the financial success of movies and TV shows [7] and new deep learning models [8].

Quantitative studies of narrative dynamics have also very recently gained ground [9–11]. Inspired by Vonnegut [2, 12, 13], Reagan et al. [10] examined how smoothed, “distantly measured” [14] happiness scores [15] vary across the length of a book. Reagan et al. find that these happiness time series across fiction books reveal six major emotional arcs, identified as: Rags-to-riches (rise), tragedy (fall), Icarus (rise-fall), Vonnegut’s man-in-a-hole (fall-rise), Cinderella (rise-fall-rise), and Oedipus (fall-rise-fall). More recently, Boyd et al. [16] showed that the different parts of Freytag’s dramatic arc [1] can be differentiated by the dominant word usage. Articles and prepositions are heavily used in the exposition stage, where narrators establish the setting. Once the reader has an understanding of the context of the story, plot progression can proceed with more pronouns and function words. As the story reaches a climax, more cognitive process words are used as the characters and narrator work through the conflict [16]. While nonfiction works such as TED talks, US Supreme Court decisions and The New York Times articles follow similar patterns in plot progression as works of fiction, they differ in patterns of staging and cognitive processes.

Schmidt [17] put forward the idea of “plot arcs”, where a text is seen as a path through topics that represent a multidimensional space. Similarly, Toubia et al. [18] studied narratives as paths in high dimensional word embedding space. Dividing a text into segments represented by corresponding average word embedding vectors, the narrative is then described as the path in word embedding space obtained by moving along consecutive segments. They identify three properties of these paths, namely the speed, volume and circuitousness in word embedding space, and examine how well these properties can predict the financial success of movies and TV shows as well as the number of citations in academic papers.

Though most studies examine changes in word usage, they focus on the general shape of narrative progression from beginning to end. Narrative is often characterized as a time series, with the general arc represented by the rise and fall of certain markers (such as happiness scores [10]) throughout the plot in terms of the fraction of the book covered. This normalization by the text length allows for a direct comparison of texts of different lengths.

∗ mikaela.fudolig@uvm.edu
such as a short story and a novel, and the corresponding shapes of their respective arcs. However, this approach does not allow us to compare sections of a text with other sections of comparable length, for example, a chapter of a novel vs. a short story of similar length.

To capture the dynamics over the course of a narrative, we need to use textual time, which we define as the number, rather than the fraction, of words covered in a text. This is similar to narrative time as defined in [19], the “amount of time passing within a discourse” (“time-of-telling”), in contrast to narrative time, the “amount of time passing within the story” (“time-of-what-is-told”). However, research on narrative dynamics in textual time is scant. While there are studies on demarcating narratives into sub-narratives [20], to our knowledge, there has been no quantitative study that examines narrative dynamics at resolutions finer than that of the general arc in the context of textual time. If word usage changes across the length of a text as shown by the general arc, does it also change across smaller sections of the text? And if so, how do we characterize these fluctuations?

While there are a number of ways to quantify word usage, we are particularly interested in how essential meaning changes over textual time. Since the 1950s, there have been efforts to distill the meaning of words into a few numbers. The valence-arousal-dominance (VAD) framework [21] has been one of the most widely used systems in quantifying meaning through semantic differentials. This framework was developed using factor analysis of word scores provided by human annotators for a small set of words, which resulted in the three dimensions of valence, arousal, and dominance. However, efforts to create large VAD lexicons using human annotators [22, 23] have shown that valence, arousal, and dominance are linearly correlated when a larger set of words is used. By transforming the NRC VAD lexicon—the largest VAD lexicon published with more than 20,000 curated words [23]—via singular value decomposition, Dodds et al. [24] showed that the VAD framework can be collapsed onto two linearly independent dimensions which align with the concepts of “power” and “danger”, with the third linearly independent dimension (“structure”) being less relevant in real-world corpora than the former two. Fig. 1 shows representative words in power-danger space for Terry Pratchett’s Discworld series. Real-world corpora also exhibit a bias towards low-danger words, which parallels the positivity bias in language [9]. Valence scores are highly correlated with happiness scores [24], which have been used to quantify narrative structure and uncover emotional arcs [10]. Thus, we expect power-danger scores, both having a linear dependence on valence (see Supplementary Information), to also be suitable markers in quantifying narrative structure and to give more information than using happiness scores alone, especially since the power-danger lexicon contains more words and would yield higher word coverage.

An illustration of how a word usage marker, such as the danger score, changes over the course of a text is given in Figure 2A–D. Each time series was constructed by splitting the text into windows of size $N_w$ words that skip every $N_s$ words; in previous studies, time series for each book was constructed using the same value for $N_w$, which range from a few hundred to a few thousand, with either $N_s$ or the the number of windows fixed. Each window corresponds to a point in the time series and is characterized by the mean score of the words it contains. Using a larger window size smooths out the time series, sometimes revealing the general shape, such as the steady increase in danger scores in “The Strange Case of Dr. Jekyll and Mr. Hyde” by Robert Louis Stevenson, or an up-down pattern as in “The Winter’s Tale” by William Shakespeare. Longer books, however, tend to retain fluctuations at the same window sizes. If books of different lengths are compared by the fraction of the book covered, which we define as the normalized textual time (Figure 2E), it would seem that the word usage changes more steadily for shorter books than longer books. However, if we compare the time series in (raw) textual time, so that the time series are of different lengths (Figure 2F), the shorter time series appear comparable to sections of

![Figure 1. An ‘ousiogram’ [24] displaying power and danger scores for a subset of 14,499 unique words appearing in Terry Pratchett’s 41-book Discworld series. The example words overlaid map the indicated eight major directions of the compass of essential meaning. The full set of 20,006 words with power and danger scores is derived from the NRC VAD lexicon [23]. These 14,499 unique words collectively account for 1,140,027 total words (types versus tokens). The histogram’s color map indicates frequency of usage. The histogram, along with the marginal distributions for power and danger (right and top) present the same safety bias observed across disparate corpora [24].]
the longer time series. Thus, changes in word usage may
be related to the length of a text, with longer texts being
a concatenation of shorter texts. In the context of narra-
tives, this implies that long narratives may be composed
of shorter narratives, each with its own arc, that function
as the basic unit of story. Some of the fluctuations found
in the time series may not be unwanted noise, but rather
a measure of the lengths of these basic units. Our aim is
to characterize these fluctuations for various texts, and
to relate them to properties of the texts themselves.

With this objective, we cannot use large window sizes,
since they run the risk of smoothing out potentially rel-
levant fluctuations. To properly characterize these basic
units of narrative, we need to use smaller window sizes,
but also isolate the fluctuations that arise from noise.
Since we are interested in how word usage changes over
the course of a text, we can compare the time series to
that obtained from a shuffled version of the text, which
contains the same words but orders them randomly. Us-
ing this as a reference isolates the effect of word order,
and also avoids making assumptions on the nature of con-
taminating noise. As expected, time series obtained from
shuffled texts are flatter, with smaller fluctuations than
those found in the original text (Figure 2A–D).

To extract and characterize fluctuations in the time
series at different scales, we use empirical mode decom-
position (EMD), a technique that factors a signal into a
sum of internal mode functions (IMF), each of which is
a mean-zero oscillatory time series with frequency and
amplitude modulations, and a non-oscillating trend [25].
While it is similar to wavelet decomposition in terms of
its objective, EMD is data-adaptive, requiring almost no
critical input from the user other than the raw time series
itself, and is also well-suited for both nonstationarity and
nonlinearity in time series. A more detailed explanation
of EMD is given in the Supplementary Information.

Figure 3 shows the result of performing ensemble em-
pirical decomposition (EEMD), an EMD variant that
is more robust to noise [26], on a time series obtained
from “The Iliad”. While the original time series, derived
from small, non-overlapping windows, contains signifi-
cant noise, EEMD is able to separate the signal into com-
ponents of different characteristic frequencies. Further,
we see that a partial reconstruction of the time series,
obtained by summing the low-frequency IMFs, can repli-
cate the time series obtained with larger window sizes,
producing a denoised version.

We mentioned earlier that the time series for shuffled
text are flatter than that of the original text (Figure 2A–
D), indicating that the IMFs of original and shuffled text
may differ in their variances. An illustration of how the
variance changes as the IMF order increases for both the
original and shuffled texts is given in Figure 4. The vari-
 ance of the IMFs in the shuffled texts generally decreases
as the IMF order increases, similar to the observation for
fractional Gaussian noise [27, 28]. However, this is not
always true for the original texts. For example, “The
Iliad” shows a clear example of a book that differs in

II. RESULTS

We analyzed more than 30,000 books from Project
Gutenberg, an online repository of books that are now
in the public domain. The selected books are almost
exclusively in English, with at least 60% of the unique
words in each book included in the lexicon. Around 60%
of the books are in the “Language and Literature” Li-
brary of Congress Classification (LCC) code (class label
“P”), while the remaining are spread out among the vari-
ous LCC class labels, with “World History” (class label
“D”) as the next largest category of books in the dataset
(around 8% of the books). While we performed the analy-

sis for both danger and power scores, the results are
similar for both, and we only discuss the danger scores
in the main text. The results for power scores are in-
cluded in the Supplementary Information.
Figure 2. (a–d) These plots illustrate the effect of window sizes, shuffling, and book length on how danger scores change over the course of a book. Using larger window sizes (solid vs. dotted) results in smoother curves, although longer books tend to retain fluctuations in the time series for the same window size (a, b vs. c, d). Shuffling the text results in time series that are flatter than the original (gray). (e) Comparing books using normalized textual time shows more fluctuations for longer books than shorter ones, but (f) plotting the time series in raw textual time shows similarities between shorter books and sections of longer books. A skip size of $N_s = 200$ words was used for all plots shown above.

A. Cutoff IMF orders

Figure 5A shows histograms of the number of words found in books that have a cutoff IMF order below the trend, and those that do not. While the choice of the rescaling factor influences the stringency of the cutoff criteria, they also offer an insight into the robustness of the results.

Rescaling to the 1st percentile of the first IMF requires a higher threshold to differentiate the IMFs, and thus more books are classified as trend-only. For those classified as having fluctuations on top of the trend, the cutoff periods obtained may be higher. On the other hand, no rescaling makes it more likely to see differences in the lower IMF orders, resulting in lower predicted cutoff periods. We note that when no rescaling is performed (i.e., the rescaling factor is 1), the cutoff IMF order for many of the books was the first IMF, which does not make much sense given that the window size of 50 words is relatively small and that the first IMF only corresponds to a period of around 100 words. However, we know from the IMF decomposition of fractional Gaussian noise [27, 29] that the first IMF has a different behavior compared to the higher-order IMFs. Thus, for purposes of comparison, we disregard the first IMF in obtaining the cutoff IMF order when no rescaling is applied.

The predicted cutoff IMF order, including the case when it does not exist, differs for majority of the books examined. However, we find some general results that hold regardless of the rescaling factor used. For instance, we find that longer books tend to have relevant fluctuations on top of the trend, while shorter books do not (Figure 5A), especially for books with less than 3000 words or greater than 100,000 words. Books with word counts in between these values may or may not have relevant fluctuations on top of the trend and the choice of the rescaling factor may influence the prediction for the cut-
Figure 3. The decomposition of the danger time series of “The Iliad” using EEMD. The raw time series was obtained using non-overlapping windows with \( N_w = 50, N_s = 50 \) (top panel), and the IMFs and the trend obtained from EEMD are shown. The lowermost panel shows the sum of IMFs 5-7 and the trend (“reconstructed”), superimposed with the danger time series obtained using larger overlapping windows \( (N_w = 5000, N_s = 200) \). Note that the partial reconstruction from time series obtained using smaller windows is very similar to the raw time series obtained using larger, overlapping windows.

We find similar observations from our analysis on both fiction and nonfiction, we also analyze the different book classifications in the next subsection and include more details in the Supplementary Information.

Similarly, the raw variances of the cutoff IMFs also do not exhibit any relationship with the book length (Figure 5E–G). The range of the variance values is also wider in comparison to that observed for the first IMF (Figure S1E–F), indicating that the books generally start to differ from each other at higher IMF orders, while the first IMF likely corresponds to noise due to the windowing technique used for all of the books. Further, while the number of IMFs found by EEMD for the original time series increases with the book length (Figure 5H–J), the cutoff IMF order itself shows no such correlation (Figure 5K–M).

We find similar observations from our analysis on
Figure 4. A comparison of the variances of the IMFs for the original text and 100 different realizations of shuffled text for three books in the dataset. Lower-order IMFs are more likely to correspond to noise, while higher-order IMFs are more likely to contain relevant information. Note that while technically the trend is not an IMF, for comparison purposes, the trend is included in this plot as the highest IMF order found for a given time series.

power time series. Shorter books are more likely to be trend-only, while longer books have relevant fluctuations above the general trend. While the cutoff periods are not identical to those found in danger scores, they are of the same order of magnitude (see Supplementary Information, Figure S3, Table S1).

B. Relationship to book content

The Gutenberg corpus assigns books to their Library of Congress Classification (LCC) subclass labels (e.g., PS for American Literature). While it is possible for a book to have more than one LCC subclass label, around 95% of the books we examined only had 1 label. The top 5 subclass labels with the most books in the dataset are American literature (PS), English literature (PR), Fiction and juvenile belles lettres (PZ), Periodicals (AP), and French/Italian/Spanish/Portuguese literature (PQ). We also looked at the class labels rather than the subclass labels (i.e., the first letter of the subclass labels). Around 60% of the class labels are for Language and Literature (P), while World History and History of Europe, Asia, Africa, Australia, New Zealand, etc. (D); Philosophy, Psychology and Religion (B); General Works (A); and History of America (E) round up the top 5 class labels. We also found that the class labels rather than the subclass labels (i.e., the first letter of the subclass labels) were more discriminative.

While there are differences in the medians for the cutoff periods and variances across subclass and class labels, the spread of both the cutoff IMF period and variance are comparable across the top 5 class and subclass labels. This indicates that the breadth of the LCC system at the class or subclass levels obscures the differences that may arise from smaller groups within these categories (Figure 6).

On the other hand, we find that using very specific filters on the title of the book yields more insightful results (Figure 7). For books with a word beginning with “poem” in the title, those with a cutoff IMF order below the trend exhibit shorter cutoff IMF periods and markedly higher cutoff IMF variances compared to the rest of the dataset, which is consistent with poems being typically short, emotional and compact. On the other hand, books with titles containing a word beginning with “manual” exhibit a lower median cutoff IMF variance. This may be because books in this category (“A manual of clinical diagnosis”, “The ladies’ book of etiquette, and manual of politeness: a complete hand book for the use of the lady in polite society”, “The skilful cook: a practical manual of modern experience,” etc.) tend to be instructional and uniform in terms of topic and mood. Books with words beginning with “play” have a higher median cutoff IMF period and lower cutoff IMF variance than books without. Results for keywords such as “collection”, “short stor” (includes both “short story” and “short stories”), “report” and “essay” are more sensitive to the choice of rescaling factor.

For power time series, while the values for the cutoff IMF periods and variances obtained are different, we also find similar overlap across different LCC class and subclass labels. We also find differences across different books depending on words in their titles (see Supplementary Information, Figures S4 and S5).

III. DISCUSSION

A. Summary of results

We examine the fluctuations in the danger and power dimensions, a reinterpretation of the valence-arousal-
Figure 5. Characterizing relevant fluctuations in the danger time series of books in the corpus. (a) These histograms show the number of books that have fluctuations on top of the trend (green) and those that do not (purple). The different line widths and line styles correspond to the different rescaling factors: solid thick lines for using the median of the first IMF, solid thin lines for using the 1st percentile of the first IMF, and dotted lines for no rescaling. (b–d). These are heatmaps that show the relationship between the period of the cutoff IMFs (as the number of words) and the book length for the various rescaling factors (shown above each plot). We can see that most of the points fall below the 45 degree line (red dashed line), indicating that the cutoff IMF period is less than the book length for the vast majority of books. The histograms for the cutoff IMF period are shown on the right side of each plot, with the 25th, 50th, and 75th percentiles shown in dotted black, solid red, and dashed black lines, respectively. The rest of the figures are similar to (b–d), but for different quantities: variance (e–g), number of IMFs (h–j; excludes the trend), and cutoff IMF order, with the first IMF counted as 1 (k–m).
Figure 6. Periods and variances of the cutoff IMF in the danger time series for the top 5 LCC subclass and class labels with the most number of books in the dataset examined. (a-f) These are the boxplots for the periods (see x-axis at the bottom for scale) and the variances (see x-axis at the top for scale) of the cutoff IMF for books in the entire dataset (“all”) as well as that for each of the top 5 LCC subclass and class labels by size. The black line inside each box is the median, the box ranges from the 25th to 75th percentiles, and the whiskers extend from the 9th percentile to the 91st percentile. (g-h) show the number of books analyzed in the dataset (“all”) and the top 5 subclass and class labels, including trend-only books.
Figure 7. Periods and variances in the danger time series for books with a word in the title if they have a cutoff IMF order below the trend. Note that since the number of books for each word is much less than the total number of books examined, the boxplots for books not containing a given word in the title will be almost identical to the boxplots for the entire dataset (“all”). (a–c) show the boxplots for the cutoff IMF period, while (d–f) show the boxplots for the cutoff IMF variance for different rescaling factors. The black line inside each box is the median, the box ranges from the 25th to 75th percentiles, and the whiskers extend from the 9th percentile to the 91st percentile. (d) shows the number of books with the given keywords in the title, including trend-only books.

dominance framework, for more than 30,000 English books in Project Gutenberg. Changes in word usage across text are analyzed by segmenting the text into windows and converting it to a time series. While window size has conventionally been preset independent of the book length, we find that large window sizes remove oscillations in word usage in shorter books but retain them in longer books. A comparison of the time series in textual time, which we define as the number of words seen up to the present moment in a book, reveals that the shorter time series are similar to subsections of longer time series. As books generally contain narratives, this
indicates that longer texts may consist of shorter texts that correspond to a basic unit of narrative, similar to subplots in a novel.

We extract the different scales of fluctuations in the time series obtained from text using empirical mode decomposition (EMD). EMD decomposes a time series into a non-oscillatory trend and a sequence of oscillatory intrinsic mode functions (IMFs) that differ in their characteristic frequencies. It allows for both nonlinearity and nonstationarity, and gives an intuitive and data-driven understanding of the underlying fluctuations in a given time series.

For each book, we derive the danger and power time series for both the original text as well as an ensemble of shuffled versions. By comparing the variances of the resulting IMFs of the time series obtained from the original text to those of the shuffled texts, we find that shorter books tend to exhibit only a non-oscillatory trend while longer books tend to exhibit relevant fluctuations with periods around the order of a few thousand words, shorter than the length of the book. The period and variance of the relevant fluctuations do not depend on the book length, despite the number of IMFs increasing with the book length. Segregation of books by the Library of Congress Classification class or subclass codes does not result in well-defined groups that show variation in the periods and variances of the relevant fluctuations. However, we observed differences when we applied very specific filters, such as words in a title. We infer that what characterizes the scale of a book’s relevant fluctuations is neither its general topic nor length, but rather more specific aspects such as its structure and content (e.g., poems vs. manuals).

The impact of our study is mainly on two fronts: providing (1) a quantitative basis for the length of a basic unit of narrative, and (2) a method for denoising time series obtained from texts without resorting to using arbitrarily large window sizes. We discuss these in the following paragraphs.

B. Word counts and cutoff periods

The shape of the general narrative arc has been the focus of a number of studies [1, 2, 10, 11, 16, 17]. Though the existence of subplots within these general narrative arcs has long been acknowledged in both the literature and computational domain [10, 20], to our knowledge, our study is the first to quantitatively analyze, in textual time, these basic units of narratives on which longer narratives are built.

While it is theoretically possible to write a basic narrative in a few words (e.g., “For sale, baby shoes, never worn.”) or a hundred thousand, balancing the flexibility of longer texts with the constraints imposed by reader engagement and publication costs has given writers and editors rules of thumb for word counts. Short stories and novellas are characterized by a focus on a single central conflict [30] or a single chain of events [31]. In contrast, the longer word count of novels (>40,000 words [32, 33]) allows for a fuller development of its characters and themes [31] through the use of chapters, each of which is similar to a short story. This is consistent with our results: trend-only books begin to decrease in number above a word count of 10,000, while the number of books with relevant fluctuations above the trend continues to increase, reaching its peak for books with 50,000 to 100,000 words (Figure 5). Further, the cutoff IMF periods are in the order of a few thousand words, comparable to the length of chapters, which are typically 1,500–5,000 words long [34, 35]. We also note that our dataset is biased towards successful books that have stood the test of time, so it is unlikely to find books that do not satisfy general guidelines set by writers and editors.

C. Fiction and nonfiction works

While plots are generally associated with fiction writing, narratives can be found in nonfiction works as well. Academic papers, for example, also show changes in word usage, although with a different signature than that found in fiction pieces [16, 36]. While 60% of the books we examined are in the Language and Literature Library of Congress category, other categories also exhibit similar general patterns for the word counts of books that are trend-only and those that are not, as well as the range of the periods and variances of the relevant fluctuations (see Supplementary Information for a comparison between literature and non-literature books, Table S1, Figures S6–S9). Our results indicate that fluctuations within the general narrative arc may be present in both fiction and nonfiction works. As nonfiction works also aim to communicate ideas, it is expected that they are also bound by rules of word usage, which is again consistent with our findings.

D. Data-adaptive denoising of text-derived time series

On the quantitative side, the method we used in this paper allows us to smooth out fluctuations in time series derived from text of various lengths. While using large window sizes reduces noise, it is unclear if it also inadvertently smooths out relevant information, such as fluctuations associated with subplots. By performing partial reconstruction using the relevant IMF orders, we have a data-driven denoising approach that works for both short and long texts.

As our method relies heavily on empirical mode decomposition, it also carries the same limitations. Although EMD will ideally construct different modes for fluctuations of sufficiently different frequencies, mode mixing may occur due to signal intermittency. Using ensemble EMD (EEMD) mitigates this problem, but does not en-
sure that it will remove mode mixing in all cases. We also note that we compare the original and shuffled texts only in the variance of their IMFs. While we verified that the pairwise comparison of IMFs results in comparable IMF periods for the original and shuffled text up to the cutoff IMF order, we only defined a difference in IMFs in terms of their variance. While this method has been proposed for finding the appropriate cutoff IMF order in using EMD for denoising [27, 29], it will miss any other difference that is not associated with the variance. We also considered using the probability density function to compare IMFs; however, this method failed to produce accurate results in synthetic data, while the variance comparison method performed well in all the tests we performed. While our general observations are robust to the choice of parameters in the variance comparison method, different results may be obtained for a particular book depending on the parameters used. Thus, while our method can extract general trends on relevant fluctuations across a corpus, sensitivity analysis must be performed when results for a particular book are of interest.

E. Future work

We have studied the scale of fluctuations within texts in their danger and power time series. While danger and power are orthogonal to each other and produce similar temporal results, we did not look at how they work together as a book progresses. Similar to the work by Toubia et al. [18], we hope to do a spatial analysis in power-danger space, specifically on the path taken by the narrative. Other possible avenues for future work include expanding our corpus to include various texts, such as screenplays and movies, as well as comparing different versions of a book, such as the first draft and the final published version.

IV. MATERIALS AND METHODS

We downloaded more than 45,000 books from Project Gutenberg, an online repository of books in public domain. The Gutenberg headers were removed using code from the Standardized Project Gutenberg Corpus [37]. Contractions, when unambiguous, were replaced with their expanded versions (e.g., “n’t” to “not”); if ambiguous, they were deleted, similar to what was done in [38]. The remaining text was then converted to lowercase and tokenized using whitespace as separators, disregarding words that contain non-word characters and digits, and ignoring punctuation marks. This converts the text into a sequence of words.

We then examine the word coverage of the power-danger lexicon. While the original NRC-VAD lexicon [23], from which the danger and power scores were derived, contained around 20,000 words, we expanded this to include noun plurals and conjugated forms not in the lexicon. Scores of the base forms of the verbs and nouns were used as the scores of their conjugated versions, expanding the lexicon to 32,721 words. We only consider books with a 60% unique word coverage, which consist almost exclusively of English text and cover 93% of all English books in the downloaded set. The median unique word coverage of this subset of books is 73%. Filtering for word coverage, removing books with duplicate titles, as well as requiring that the time series must have at least one lexicon word and that the EEMD can be successfully computed (i.e., the EEMD decomposes up to the trend level, such that the mean of the sum of the EEMD results is within 10% of the mean of the original time series), leaves us with 31,690 books (Figure 8).

We construct the danger and power time series by segmenting the sequence of words into non-overlapping windows of size $N_w = 50$, each of which corresponding to a point in the time series. In each window, we take each word $w_i$ with score $s_i$ in the lexicon that occurs $n_i$ times in the window. If there are $m$ unique words in the window that are in the lexicon, then the score for the window is

$$s_w = \frac{\sum_{i=1}^{m} n_is_i}{\sum_{i=1}^{m} n_i}$$  

The reference time series are constructed by shuffling the tokenized version of the text, performing the windowing technique, and recomputing the average scores for each window. This assures that the time series of

![Figure 8. Word coverage of the PDS lexicon. The blue histograms correspond to the unique word coverage, while the orange histograms correspond to the word coverage allowing for word repetition. We show the histograms for all the books downloaded in our dataset (empty bars) as well as those we analyzed (shaded bars), and the dashed lines show the median word coverage of the books analyzed. The books included in our analysis do not include duplicate titles and must have (1) a unique word coverage of at least 60%, (2) at least one lexicon word in every window and (3) intrinsic mode functions (IMFs) that can be successfully computed.](image-url)
both the target (original text) and the reference (shuffled text) are of the same length.

We use ensemble empirical mode decomposition (EEMD) [26] with an ensemble size of 100 to obtain the internal mode functions (IMFs) for the time series corresponding to the original text. Each time series in the ensemble is the sum of the raw time series and white noise with a standard deviation $0.2\sigma$, where $\sigma$ is the standard deviation of the raw time series, as suggested by Wu and Huang [26]. As our reference time series, we use 100 different shuffled versions of the original text. These are generated by shuffling the tokenized version of the text, performing the windowing technique, and recomputing the average scores for each window. This assures that the time series of both the target and the reference are of the same length. For each of the reference time series, we use basic empirical mode decomposition (EMD) to obtain the IMFs. We use the Python package emd [39] to perform all EMD-related calculations.

We then divide the variances of the IMFs of the time series from original text by a rescaling factor, as suggested in [27, 29]. Three different rescaling factors were considered based on the distribution of the first IMFs of the reference time series: the median, the 1st percentile, and unity (no rescaling). The lowest IMF order at which the rescaled variance is higher than the 99th percentile of the variances for shuffled text is considered as the cutoff IMF order.

Once the cutoff IMF order is identified, the corresponding period is computed from the center of the frequency bin with the highest energy as obtained using the Hilbert-Huang transform (HHT). Since we want to count periods in the unit of number of words, we compute for the HHT setting the sampling rate at $N_{w}^{-1}$ word$^{-1}$, where $N_{w} = 50$ is the skip size in the windowing procedure. We use logarithmically spaced frequency bins, spanning from $10^{-6}$ word$^{-1}$ to 1 word$^{-1}$, resulting in a range of period values between 1 to $10^6$ words, chosen because none of the texts examined exceed $10^6$ words in length. The choice of logarithmic spacing is motivated in part by how EMD on white noise performs like a dyadic filter, with IMF frequencies decreasing by roughly a factor of 2 for every order. Further, in our preliminary analysis for select books, we find that logarithmic spacing provides an adequate representation of the spectra, especially since the IMF frequencies and periods span orders of magnitude. We emphasize that obtaining a characteristic value for the period of the IMF is independent of the method used to extract the cutoff IMF order discussed earlier, and that the bins used for the HHT are the same across all texts.

ACKNOWLEDGMENTS

The authors are grateful for the computing resources provided by the Vermont Advanced Computing Core and financial support from the Massachusetts Mutual Life Insurance Company. Computations were performed on the Vermont Advanced Computing Core supported in part by NSF award No. OAC-1827314. M.I.F. is also grateful to Miguel Fudolig, Joshua Minot, Andy Reagan and Aritra Banerjee for helpful discussions.

[1] Gustav Freytag, Freytag’s Technique of the drama: an exposition of dramatic composition and art. An authorized translation from the 6th German ed., 3rd ed. (Scott, Foresman, Chicago, 1900) open Library ID: OL7168981M.
[2] Kurt Vonnegut, Palm Sunday: An Autobiographical Collage (Random House Publishing Group, 1999).
[3] Paul Ricoeur, “Narrative Time,” Critical Inquiry 7, 169–190 (1980), publisher: The University of Chicago Press.
[4] Gérard Genette, Narrative Discourse: An Essay in Method (Cornell University Press, 1983) google-Books-ID: yEPuQg7SOxIc.
[5] James Phelan and Peter Rabinowitz, “Time, Plot, Progression,” in Narrative theory: Core concepts and critical debates (Ohio University Press, 2012) pp. 57–65.
[6] Steven Brown and Carmen Tu, “The shapes of stories: A “resonator” model of plot structure,” Frontiers of Narrative Studies 6, 259–288 (2020).
[7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space,” in Proceedings of ICLR Workshops (2013) arXiv: 1301.3781.
[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is All you Need,” in Advances in Neural Information Processing Systems (2017).
[9] Peter Sheridan Dodds, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdichian, Matthew T. McMahon, Brian F. Tivnan, and Christopher M. Danforth, “Human language reveals a universal positivity bias,” Proceedings of the National Academy of Sciences 112, 2389–2394 (2015), publisher: National Academy of Sciences Section: Social Sciences.
[10] 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,” EPJ Data Science 5, 31 (2016).
[11] Jianbo Gao, Matthew L. Jockers, John Ludlow, and Timothy Tangerlini, “A multiscale theory for the dynamical evolution of sentiment in novels,” in 2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC) (2016) pp. 1–4.
[12] Kurt Vonnegut, Jr., A Man Without a Country, stories (Seven Stories Press, New York, 2005).
[13] “Kurt Vonnegut on the shapes of stories,” (2010), https://www.youtube.com/watch?v=OP3clh8vZ2Q, ac-
cessed May 15, 2014.

[14] Franco Moretti, Distant Reading (Verso, New York, 2013).

[15] Peter Sheridan Dodds, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth, “Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter,” PLOS ONE 6, e26752 (2011).

[16] Ryan L. Boyd, Kate G. Blackburn, and James W. Pennebaker, “The narrative arc: Revealing core narrative structures through text analysis,” Science Advances 6, eaab2196 (2020).

[17] Benjamin M. Schmidt, “Plot arcology: A vector-space model of narrative structure,” in 2015 IEEE International Conference on Big Data (Big Data) (2015) pp. 1667–1672.

[18] Olivier Toubia, Jonah Berger, and Jehoshua Eliashberg, “How quantifying the shape of stories predicts their success,” Proceedings of the National Academy of Sciences 118 (2021), 10.1073/pnas.2011695118.

[19] Andrew Piper, Richard Jean So, and David Bamman, “Narrative Theory for Computational Narrative Understanding,” in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021) pp. 298–311.

[20] Byron Wallace, “Multiple Narrative Disentanglement: Unraveling Infinite Jest,” in Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Association for Computational Linguistics, Montréal, Canada, 2012) pp. 1–10.

[21] Charles Egerton Osgood, George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning (University of Illinois Press, 1957).

[22] Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert, “Norms of valence, arousal, and dominance for 13,915 English lemmas,” Behavior Research Methods 45, 1191–1207 (2013).

[23] Saïf Mohammad, “Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Association for Computational Linguistics, Melbourne, Australia, 2018) pp. 174–184.

[24] P. S. Dodds, T. Alshaabi, M. I. Fudolig, J. W. Zimnerman, J. Lovato, S. Beaulieu, J. R. Minot, M. V. Arnold, A. J. Reagan, and C. M. Danforth, “Osiometrics and Telegnomics: The essence of meaning conforms to a two-dimensional powerful-weak and dangerous-safe framework with diverse corpora presenting a safety bias,” arXiv:2110.06847 [physics] (2021), arXiv: 2110.06847.

[25] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454, 903–995 (1998).

[26] Zhaohua Wu and Norden E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” Advances in Adaptive Data Analysis 01, 1–41 (2009).

[27] Zhaohua Wu and Norden E. Huang, “A study of the characteristics of white noise using the empirical mode decomposition method,” Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 460, 1597–1611 (2004).

[28] P. Flandrin, G. Rilling, and P. Gonçalves, “Empirical mode decomposition as a filter bank,” IEEE Signal Processing Letters 11, 112–114 (2004), conference Name: IEEE Signal Processing Letters.

[29] Patrick Flandrin, Paulo Gonçalves, and Gabriel Rilling, “Emp equivalent filter banks, from interpretation to applications,” in Hilbert-Huang Transform and Its Applications, Interdisciplinary Mathematical Sciences, Vol. 5 (World Scientific, 2005) pp. 57–74.

[30] MasterClass, “Learn the Differences Between Novelettes, Novellas, and Novels,” (2021), https://www.masterclass.com/articles/learn-the-differences-between-novelettes-novellas-and-novels.

[31] Chris Balderick, The Oxford Dictionary of Literary Terms (Oxford University Press, 2015) publication Title: The Oxford Dictionary of Literary Terms.

[32] World Science Fiction Society, “Hugo Award Categories,” (2022), https://www.thehugoawards.org/hugo-categories/.

[33] Science Fictions & Fantasy Writers of America, “Nebula Rules,” (2020), https://nebulas.sfwa.org/about-the-nebulas/nebula-rules/.

[34] Harry Bingham, “How long should a chapter be?” (2020), https://jerichowriters.com/how-long-should-a-chapter-be/.

[35] “How Long Should A Chapter Be? The Master Guide To Chapter Length.” (2017), https://blog.reedsy.com/how-long-should-a-chapter-be/.

[36] Eitan Adam Pechenick, Christopher M. Danforth, and Peter Sheridan Dodds, “Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution,” PLOS ONE 10, e0137041 (2015), publisher: Public Library of Science.

[37] Martin Gerlach and Francesc Font-Clos, “A Standardized Project Gutenberg Corpus for Statistical Analysis of Natural Language and Quantitative Linguistics,” Entropy 22, 126 (2020).

[38] Mikaela Irene Fudolig, Thayer Alshaabi, Michael V. Arnold, Christopher M. Danforth, and Peter Sheridan Dodds, “Sentiment and structure in word co-occurrence networks on Twitter,” Applied Network Science 7, 1–27 (2022).

[39] Andrew Quinn, Vitor Lopes-dos Santos, David Dupret, Anna Nobre, and Mark Woolrich, “EMD: Empirical Mode Decomposition and Hilbert-Huang Spectral Analyses in Python,” Journal of Open Source Software 6, 2977 (2021).

[40] Abdel-Ouahab Boudraa and Jean-Christophe Cexus, “EMD-Based Signal Filtering,” IEEE Transactions on Instrumentation and Measurement 56, 2196–2202 (2007).

[41] A.O. Boudraa, J.C. Cexus, S. Benramdane, and A. Beghdadi, “Noise filtering using Empirical Mode Decomposition,” in 2007 9th International Symposium on Signal Processing and Its Applications (IEEE, Sharjah, United Arab Emirates, 2007) pp. 1–4.

[42] Albert Ayenu-Prah and Nii Attoh-Okin, “A Criterion for Selecting Relevant Intrinsic Mode Functions in Empirical Mode Decomposition.” Advances in Adaptive Data Analysis 2, 1–24 (2010).
[43] Ali Komaty, Abdel-Ouahab Boudraa, Benoit Augier, and Delphine Daré-Emzivat, “EMD-Based Filtering Using Similarity Measure Between Probability Density Functions of IMFs,” IEEE Transactions on Instrumentation and Measurement 63, 27–34 (2014).
SUPPLEMENTARY INFORMATION

Power-danger framework

While the valence-arousal-dominance (VAD) framework [21] has been one of the longstanding theories for the quantification of meaning, attempts to measure VAD scores for a larger set of words reveal that these three dimensions are not necessarily orthogonal. In particular, the NRC VAD lexicon [23], which has around 20,000 words, shows moderate correlation between pairs of variables (r_{V,A} \simeq -0.27, r_{A,D} \simeq 0.30, and r_{V,D} \simeq 0.49). Performing singular value decomposition creates three new dimensions (goodness, energy and structure) that are linear combinations of valence, arousal and dominance, and a clockwise rotation of \pi/4 in the goodness-energy plane creates a framework where each dimension can be interpreted by the words that occupy its extreme values [24]. The resulting three dimensions are power, danger, and structure, and are related to valence, arousal, and dominance in the following way:

\[
\begin{bmatrix}
\text{Power} \\
\text{Danger} \\
\text{Structure}
\end{bmatrix} = \begin{bmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
+0.86 \\
-0.16 \\
0.48
\end{bmatrix} \begin{bmatrix}
\text{Valence}' \\
\text{Arousal}' \\
\text{Dominance}'
\end{bmatrix}
\]

(S1)

where the ‘ symbol indicates that the valence, arousal, and dominance values are rescaled to lie in the range \([-\frac{1}{2}, \frac{1}{2}]\]. In real-world corpora, power and danger are the more relevant dimensions. We refer the reader to [24] for an in-depth discussion of the power-danger-structure framework.
Comparison between first and cutoff IMFs

Figure S1. Histograms of the median ratios (measured across the different shuffling realizations) of the period of the IMF corresponding to the original text to that of the shuffled text for the danger time series, as well as the period and variance corresponding to the first IMF. (a–d) The unshaded histograms refer to ratios for the first IMF, while the shaded histograms refer to the ratios for the cutoff IMF. In (a), we include all books, whether a cutoff IMF order below the trend was detected (“fluc”) or not (“trend”). In (b)–(d), we only include the ratios for the books where a cutoff IMF order below the trend was detected using the variance comparison method when the IMFs are rescaled by the rescaling factor (50th percentile for (b), 1st percentile for (c), no rescaling for (d)). (e–f) These heatmaps summarize the period and variance corresponding to the first IMF of the books in the dataset as a function of the book length. The right side of each heatmap shows the corresponding histogram and the 25th (black, dotted line), 50th (red, solid line) and 75th (black, dashed line) percentiles.
Figure S2. Histograms of the median ratios (measured across the different shuffling realizations) of the period of the IMF corresponding to the original text to that of the shuffled text for the power time series, as well as the period and variance corresponding to the first IMF. (a–d) The unshaded histograms refer to ratios for the first IMF, while the shaded histograms refer to the ratios for the cutoff IMF. In (a), we include all books, whether a cutoff IMF order below the trend was detected (“fluc”) or not (“trend”). In (b)–(d), we only include the ratios for the books where a cutoff IMF order below the trend was detected using the variance comparison method when the IMFs are rescaled by the rescaling factor (50th percentile for (b), 1st percentile for (c), no rescaling for (d)). (e–f) These heatmaps summarize the period and variance corresponding to the first IMF of the books in the dataset as a function of the book length. The right side of each heatmap shows the corresponding histogram and the 25th (black, dotted line), 50th (red, solid line) and 75th (black, dashed line) percentiles.
Figure S3. Characterizing relevant fluctuations in the power time series of books in the corpus (c.f. Figure 5). (a) These histograms show the number of books that have fluctuations on top of the trend (green) and those that do not (purple). The different line widths and line styles correspond to the different rescaling factors: solid thick lines for using the median of the first IMF, solid thin lines for using the 1st percentile of the first IMF, and dotted lines for no rescaling. (b–d). These are heatmaps that show the relationship between the period of the cutoff IMFs (as the number of words) and the book length for the various rescaling factors (shown above each plot). We can see that most of the points fall below the 45 degree line (red dashed line), indicating that the cutoff IMF period is less than the book length for the vast majority of books. The histograms for the cutoff IMF period are shown on the right side of each plot, with the 25th, 50th, and 75th percentiles shown in dotted black, solid red, and dashed black lines, respectively. The rest of the figures are similar to (b–d), but for different quantities: variance (e–g), number of IMFs (h–j), and cutoff IMF order, with the first IMF counted as 1 (k–m).
Figure S4. Periods and variances of the cutoff IMF in the power time series for the top 5 LCC subclass and class labels with the most number of books in the dataset examined. (a-f) These are the boxplots for the periods (see x-axis at the bottom for scale) and the variances (see x-axis at the top for scale) of the cutoff IMF for books in the entire dataset (“all”) as well as that for each of the top 5 LCC subclass labels by size. The black line inside each box is the median, the box ranges from the 25th to 75th percentiles, and the whiskers extend from the 9th percentile to the 91st percentile. (g-h) shows the number of books analyzed in the dataset (“all”) and the top 5 subclass and class labels, including trend-only books.
Figure S5. Periods and variances in the power time series for books with a word in the title if they have a cutoff IMF order below the trend. Note that since the number of books for each word is much less than the total number of books examined, the boxplots for books not containing a given word in the title will be almost identical to the boxplots for the entire dataset (“all”). (a–c) show the boxplots for the cutoff IMF period, while (d–f) show the boxplots for the cutoff IMF variance for different rescaling factors. The black line inside each box is the median, the box ranges from the 25th to 75th percentiles, and the whiskers extend from the 9th percentile to the 91st percentile. (d) shows the number of books with the given keywords in the title, including trend-only books.
Comparison between literature and non-literature books

| Rescaling factor | Percentile 25th | Percentile 50th | Percentile 75th | 1st percentile 25th | 1st percentile 50th | 1st percentile 75th | Without rescaling 25th | Without rescaling 50th | Without rescaling 75th |
|------------------|-----------------|-----------------|-----------------|---------------------|---------------------|---------------------|------------------------|------------------------|------------------------|
| **Danger**       |                 |                 |                 |                     |                     |                     |                        |                        |                        |
| All              | 1000            | 1400            | 3200            | 1200                | 2800                | 6400                | 500                    | 600                    | 1400                   |
| Literature       | 1000            | 2100            | 3700            | 1400                | 2800                | 6400                | 400                    | 900                    | 2100                   |
| Non-literature    | 900             | 1200            | 2800            | 1200                | 2800                | 5600                | 400                    | 500                    | 1200                   |
| **Power**        |                 |                 |                 |                     |                     |                     |                        |                        |                        |
| All              | 1000            | 2400            | 4200            | 1400                | 3200                | 7400                | 400                    | 500                    | 1200                   |
| Literature       | 1000            | 2400            | 5600            | 2400                | 3200                | 7400                | 400                    | 500                    | 1200                   |
| Non-literature    | 1000            | 1400            | 3200            | 1200                | 2800                | 6400                | 400                    | 500                    | 1100                   |

Table S1. Cutoff IMF period in number of words, rounded up to the nearest hundred, for danger and power time series for books with a cutoff IMF order below the trend.
Figure S6. Characterizing relevant fluctuations in the *danger* time series of books in the corpus for books under the Language and Literature Library of Congress Classification code “P” (c.f. Figure 5). (a) These histograms show the number of books that have fluctuations on top of the trend (green) and those that do not (purple). The different line widths and line styles correspond to the different rescaling factors: solid thick lines for using the median of the first IMF, solid thin lines for using the 1st percentile of the first IMF, and dotted lines for no rescaling. (b–d). These are heatmaps that show the relationship between the period of the cutoff IMFs (as the number of words) and the book length for the various rescaling factors (shown above each plot). We can see that most of the points fall below the 45 degree line (red dashed line), indicating that the cutoff IMF period is less than than the book length for the vast majority of books. The histograms for the cutoff IMF period are shown on the right side of each plot, with the 25th, 50th, and 75th percentiles shown in dotted black, solid red, and dashed black lines, respectively. The rest of the figures are similar to (b–d), but for different quantities: variance (e–g), number of IMFs (h–j), and cutoff IMF order, with the first IMF counted as 1 (k–m).
Figure S7. Characterizing relevant fluctuations in the *danger* time series of books in the corpus for books in codes other than the Language and Literature Library of Congress Classification code “P” (c.f. Figure 5). 

(a) These histograms show the number of books that have fluctuations on top of the trend (green) and those that do not (purple). The different line widths and line styles correspond to the different rescaling factors: solid thick lines for using the median of the first IMF, solid thin lines for using the 1st percentile of the first IMF, and dotted lines for no rescaling. 

(b–d). These are heatmaps that show the relationship between the period of the cutoff IMFs (as the number of words) and the book length for the various rescaling factors (shown above each plot). We can see that most of the points fall below the 45 degree line (red dashed line), indicating that the cutoff IMF period is less than the book length for the vast majority of books. The histograms for the cutoff IMF period are shown on the right side of each plot, with the 25th, 50th, and 75th percentiles shown in dotted black, solid red, and dashed black lines, respectively. 

The rest of the figures are similar to (b–d), but for different quantities: variance (e–g), number of IMFs (h–j), and cutoff IMF order, with the first IMF counted as 1 (k–m).
Figure S8. Characterizing relevant fluctuations in the power time series of books in the corpus for books under the Language and Literature Library of Congress Classification code “P” (c.f. Figure 5). (a) These histograms show the number of books that have fluctuations on top of the trend (green) and those that do not (purple). The different line widths and line styles correspond to the different rescaling factors: solid thick lines for using the median of the first IMF, solid thin lines for using the 1st percentile of the first IMF, and dotted lines for no rescaling. (b–d). These are heatmaps that show the relationship between the period of the cutoff IMFs (as the number of words) and the book length for the various rescaling factors (shown above each plot). We can see that most of the points fall below the 45 degree line (red dashed line), indicating that the cutoff IMF period is less than the book length for the vast majority of books. The histograms for the cutoff IMF period are shown on the right side of each plot, with the 25th, 50th, and 75th percentiles shown in dotted black, solid red, and dashed black lines, respectively. The rest of the figures are similar to (b–d), but for different quantities: variance (e–g), number of IMFs (h–j), and cutoff IMF order, with the first IMF counted as 1 (k–m).
Figure S9. Characterizing relevant fluctuations in the power time series of books in the corpus for books in codes other than the Language and Literature Library of Congress Classification code “P” (c.f. Figure 5). (a) These histograms show the number of books that have fluctuations on top of the trend (green) and those that do not (purple). The different line widths and line styles correspond to the different rescaling factors: solid thick lines for using the median of the first IMF, solid thin lines for using the 1st percentile of the first IMF, and dotted lines for no rescaling. (b–d). These are heatmaps that show the relationship between the period of the cutoff IMFs (as the number of words) and the book length for the various rescaling factors (shown above each plot). We can see that most of the points fall below the 45 degree line (red dashed line), indicating that the cutoff IMF period is less than the book length for the vast majority of books. The histograms for the cutoff IMF period are shown on the right side of each plot, with the 25th, 50th, and 75th percentiles shown in dotted black, solid red, and dashed black lines, respectively. The rest of the figures are similar to (b–d), but for different quantities: variance (e–g), number of IMFs (h–j), and cutoff IMF order, with the first IMF counted as 1 (k–m).
Empirical mode decomposition

Given a time series \( x(t) \), EMD takes the mean \( m_1 \) of the upper envelope (connecting the local maxima by a cubic spline) and the lower envelope (connecting the local minima, also by a cubic spline). The difference \( h_{1,1} = x(t) - m_{1,1} \) is then checked for two conditions: (a) the number of extrema and the number of zero crossings must either equal or differ at most by one, and (b) the mean of \( h_1 \) must be zero. If either condition is not met, this process is repeated \( k \) times until both conditions are met, resulting in:

\[
h_{1,k} = h_{1,k-1} - m_{1,k}
\]  

This process is called sifting, and the function \( c_1 := h_{1,k} \) is defined as the first IMF. The difference \( h_{2,1} = x(t) - c_1 \) is obtained and the same procedure is repeated, with each subsequent IMF corresponding to decreasing frequencies as the iteration progresses. This iterative process continues \( M \) times until \( x(t) - c_M \) can no longer satisfy the conditions for an IMF; this is called the residual, \( r_M \). Thus, the original signal can be decomposed as the sum of the IMFs and the residual

\[
x(t) = \sum_{j=1}^{M} c_j + r_M.
\]  

Each IMF can then be characterized by its frequency spectrum, which is obtained using the Hilbert (Hilbert-Huang) transform, and the characteristic frequency can be taken to be the center of the frequency bin with the highest energy. The sum of the lower-frequency IMFs can then be used to denoise a time series. While similar to using a low-pass filter, using EMD also separates the fluctuations corresponding to each characteristic frequency, allowing comparison of IMFs across different temporal scales.

A commonly used variant is the ensemble empirical decomposition method (EEMD), which was developed to reduce the effect of noise on the IMFs. Several realizations of white noise are added to the raw signal, and EMD is performed on each. The EEMD then takes the average of the IMFs obtained across all these augmented signals. Because of the ensemble averaging step, the resulting IMFs are more robust to the effect of noise.

While EMD can extract oscillatory modes of different frequencies from a time series, the method does not dictate which of the IMFs are due to noise and which are relevant to the true signal. Methods developed include looking at the variances of the IMFs to set a cutoff IMF order [40, 41], comparing the correlation coefficients of each IMF and the raw time series [42], and comparing the probability density functions of the IMFs to that of the raw time series [43]. However, preliminary analysis showed that methods comparing the IMFs to the raw time series do not work sufficiently on our data, likely due to the choice of window size resulting in a low signal-to-noise ratio, so that the assumption that the relevant IMFs are closer to the original time series than the noise IMFs does not hold. We thus adapt a method that compares the time series to a null reference [27, 28], but use shuffled text instead of Gaussian noise as the null model.