Term Bursts and Battle Lines

Trends in Public Discourse During and After the English Civil War
(1641 - 1661)

Jacqueline Basu

11 December 2016

1 Introduction

The two decades between 1641 and 1661 were a time of great tumult for Great Britain. Political struggles between Parliament and the monarchy were exacerbated by – indeed, often expressed in the language of – ongoing religious strife. Nine of these years (1642 - 1651) were marked by civil war. This war, however, was not confined to the battleground. It was equally a war of words, in which political theory, religious doctrine, and civil discourse played a central role. When the armed conflict ended, this ongoing discourse continued to reshape the institutions of British power, as "both the practical and the philosophical bases of the British monarchy were being challenged" [15]. Luckily, this public debate over the philosophy and theology of politics was captured by the popular press, which was unusually active at this time.

Of the almost 30,000 documents produced during this period, about 22,000 have been preserved in a unique corpus known as the "Thomason Tracts." This corpus contains all manner of public speech – orations, manifestos, popular chatter, sermons, letters, and news. In such a large and varied corpus, traditional historical methods are not well-designed to assess broad trends, strands of debate, and evolution in this debate over time. To this end, nearly half of this corpus has been made digitally available, although it has not yet been the subject of large-scale computational analysis.

This project aims to begin the work of characterizing such trends by observing variations in word usage frequency over time. Specifically, it intends to a) identify when individual words in the corpus experience periods of intensified usage, or "bursts"; b) observe which terms burst simultaneously and determine how these simultaneous bursts aggregate into broader conceptual 'trends'; c) identify "landmark years" which mark major shifts in the prevailing trends of language use; and finally d) relate these shifts in language use to their sociopolitical context, to observe how linguistic trends reflect and respond to temporal events.

1.1 Data Collection and Preparation

The data consists of 9180 texts which span the years 1641-1661 (cf. Table 1). These documents were obtained from Early English Books Online (EEBO), a database which serves as the primary digital repository of early English texts [15, 16]. Access to the full digital corpus is restricted to use for academic purposes, and was obtained through Stanford's Social Science Data and Software group [3]. Metadata that was collected along with the text of each document includes: a) author; b) document title; c) publication year; d) city of publication; e) unique bibliographic identifier.

Because early modern English does not have standardized spelling conventions, it was necessary to mitigate some of the wide variation in spelling present in the original documents. Dr. Alistair Baron from the University of Lancaster has developed a software tool to normalize early modern English spelling [1]; this tool, known as VARD 2, was used by the Distant Reading Early Modernity (DREaM) group at McGill University to generate a dictionary of 80,676 spelling normalizations [14]. I used this dictionary to normalize my own texts, checking the normalized output against VARD 2 to ensure that the process worked on my corpus as well. I then stemmed and lemmatized these normalized texts using the Porter Stemmer [2]. Finally, I converted the data into a document-term matrix (DTM), which contains the 1000 most prevalent terms in the corpus, as well as the number of occurrences of each of those terms within each document.

1641	**1645**	**1649**	**1653**	**1657**	**1661**
god	god	god	god	god	god
church	church	christ	christ	christ	king
hath	christ	lord	man	thing	thing
lord	lord	man	men	man	church
time	men	men	hath	hath	thou

Table 2: Top 5 Terms by Rate of Appearance, 1641-1661

Table 2 indicates that the most frequently used terms in the corpus remain remarkably static over time: from the table, we see that political ('king', 'law'); social ('man', 'men'); and religious ('god', 'christ', 'church') themes are dominant, but the sheer prevalence of these primary terms masks meaningful variation that would help to make sense of how these themes evolve and interact. To capture this variation in the document stream, this analysis intends to a) implement a method

of identifying "bursts" in the usage of individual terms; and b) characterize how the co-occurrence of those bursts lends insights into broader currents of sociopolitical change.

2 Literature Review

2.1 Identifying Bursts

Kleinberg's conceptual definition of 'bursts' is a simple one: "a 'burst of activity' " is one in which "certain features rise sharply in frequency." [11,2] Many models that measure such intensifications of activity do so by analyzing user *behavior*: they model a static underlying network of users (blogs, news outlets, Twitter users, etc.); and they observe how social behavior propagates the burst of a term, news item, hashtag, or other viral piece of text [10, 9, 8]. Other models characterize the text stream directly, without relying on the underlying authorial or user structure [13, 11]. These models simply track the frequency with which term appear in the corpus, identifying "bursts" as those periods in which particular words or groups of words appear at a significantly elevated rate.

The simplest way of determining what constitutes a 'significantly elevated rate' is to set a constant-valued threshold T that applies to all terms: T, then, represents the usage rate above which a term can be deemed "bursty." Ratkiewicz et al. adopt this definition in their analysis of political astroturf [10]; however, this method is prone to three central problems: a) *overrepresentation* of high-frequency terms which appear at a uniformly high rate throughout the corpus; b) lack of basis for *ranking* the intensity of different terms' burst phenomena; c) *unsmoothed* or ragged bursts, caused by outlier/unrepresentative time steps which spuriously induce or terminate burst phenomena. By contrast with the threshold approach, other models offer algorithmic methods of identifying bursts which overcome all three of these problems. [13, 11]. Kleinberg provides one such algorithm:

To track 'bursty' behavior for a particular term a, Kleinberg provides a method for modelling the text stream as a finite-state automaton $\mathcal{A}$. In a simple binary model, if term a is used at time t with a frequency higher than some threshold quantity, $\mathcal{A}$ undergoes a phase change from baseline state q_0 to some elevated state q_1. This phase change signals the beginning of a period of "bursty" behavior for term a. A cost is incurred to change from the lower state q_0 to the bursting state q_1. This cost introduces some inertia in the model, so that $\mathcal{A}$ does not register a new phase change for each time step in which term a is disproportionately represented (thus overcoming problem 3, above). Because each term's burstiness is measured relative to its own baseline frequency within the dataset, high-frequency terms are not overly represented in the weighting of burst phenomena (overcoming problem 1). Finally, Kleinberg provides a method of using the cost function for the optimal state sequence to rank bursts by their intensity (overcoming problem 2).

2.2 Characterizing Burst Context

After providing methods to identify bursts, many papers also seek to situate bursts within their broader textual or temporal context. Some such methods are rooted in the features of contemporary social media: Mathioudakis and Koudis, studying Twitter, use tweet metadata and named entity parsers to locate trends of bursting terms within the universe of Twitter users and Twitter content. [13]. Other methods rely less heavily on the infrastructure and vocabulary of contemporary social media, and are thus more applicable to historical data. Kleinberg and Mathioudakis/Koudas describe two similar such approaches. They both begin by co-locating bursting terms in time: at each time step n, K_n is the set of terms which are in a bursting phase. They then seek to find, within time step n, some set L_n of 'landmark texts', within which a set of bursting terms $l_n \subset K_n$ appears in elevated numbers. For a word w which bursts during time n, the sets l_n and L_n in which w is involved provide different types of context for its bursting behavior. By looking at l_n, we can observe which other words experience bursts of elevated activity along with term w, forming what Mathioudakis/Koudas call a conceptual 'trend.' Meanwhile, the set L_n contains the full document text which houses that trend; though the full document is not bursting during time n, its topical content can lend insight to w's bursting behavior [13, 11]. Kulkarni et al. bolster this idea of using a term's textual context (in tandem with the term's usage frequency) to characterize periods of linguistic change. [17]

2.3 Political Theory and Computational Text Analysis

Since about 2013, political science has slowly been incorporating computational text analysis into its methodological tool kit [7]. However, it has not yet gained much use within the more insulated subfield of political theory/political philosophy. Blaydes, McQueen, and Grimmer recently made one of the first steps in this direction, producing a paper which conducted a theoretical analysis of medieval advice texts in Middle Eastern and Christian polities using computational topic modeling [12].

3 Methods

3.1 Identifying Bursts

The Thomason Tracts are segmented by year into twenty discrete batches, so cannot be modelled as a continuous text stream. The finite-automaton version of Kleinberg's algorithm was designed for batched data of this kind. Therefore, I will identify bursts by adapting his framework to model my data as a finite automaton $\mathcal{B}_s^2$ with two states q_0 and q_1.

For each word w in the corpus vocabulary (or a subset comprised of the most frequent terms in this vocabulary), for t in years $t \in [0, 21]$ (because the corpus spans the two decades between 1641 and 1661), the t^{th} batch is the aggregate of all words in year t. Then d_t is the total number of words uttered this year, and r_t is the number of occurrences of w. Let $R = \sum_{t=0}^{21} r_t$ and $D = \sum_{t=0}^{21} d_t$. If $\mathcal{B}$ is in the base state q_0 in a given time step t, then p_0 – the expected number of occurrences of w relative to the total number of words in the time step – is simply equal to R/D. In q_1, the 'burst' phase, p_1 increases: $p_1 = (R/D) * s$ where s is some scaling factor $s > 1$.

Given these basic definitions and the document-term matrix with the counts of each word w per year, we simply select an optimal state sequence $Q = (q_i 1 ... q_i t)$, where i is the state at time t, which minimizes the cost function:

$$c(Q) = \left(\sum_{t=0}^{n-1} \tau(i_t, i_{t+1})\right) + \left(\sum_{t=0}^{n} -ln\left[d_t r_t p_i^{r_t}(1-p_i)^{d_t - r_t}\right]\right)$$

The first term in the cost function smooths the burst sequence by penalizing transitions from the base state q_0 to the burst state q_1, with a cost determined by the smoothing parameter τ; transitions in the opposite direction incur no cost. The second term in the cost function rewards a close fit between the model and observed data. Thus, the optimal state sequence is one that is parsimonious in its transition into burst phases, but also maps well to the observed data.

Ranking Bursts by Weight: Finally, Kleinberg offers a simple method to rank the intensity of burst phenomena, by determining the improvement in the cost function attained by transitioning into state q_1 for the duration of the burst $[t_1, t_2]$. He calls this score the burst's "weight": higher-weighted bursts are assumed to be more significant. He also uses this weight measure to assign a cumulative validity score over the whole dataset, by taking the sum of burst weights over all terms in the corpus. Thus, variations of the algorithm (e.g. with different parameters s and τ) can be compared, as they will produce different cumulative weight scores over the same dataset: higher weight scores imply stronger burst phenomena.

3.2 Selecting parameters s and τ

Scaling Parameter s: The scaling parameter s determines the level of intensified term usage required to trigger a phase change in the automaton. Thus, increasing values of s increase the intensity of term usage required to trigger the change to a "bursty" state. Kleinberg notes that, while he uses a value of $s = 2$ for the shorter time-scale data (email, conference proceedings) he increases this to $s = 8$ or even $s = 16$ for the longer time-scale data (e.g. the centuries-long state of the Union address records). Because my data occurs over the fairly long time scale of two decades, I ran the algorithm over a few values of $s > 2$, although they appeared to generate far fewer bursts, with weaker overall weight scores (cf. Table 1). I ultimately settled on using $s = 2$ for the rest of my analysis.

During this phase of experimentation, I also varied s to try and identify "anti-bursts", in which the resolution parameter $s \in (0, 1)$. Kleinberg only models *intensifications* of term use over time. However, I can also imagine that it might be useful to track the decline in use of terms, relative to the uniform baseline. I did not use these in my later analysis, but it seems like another potentially useful way of approaching text data, especially highly-political text data in which abrupt declines in term use might signal bouts of regime-directed censorship or suppression of civil society.

s **value**	**# Bursts**	**Tot. Weight**
4	141	116636
3	261	169569
2	572	263217
0.5	704	209885
0.25	258	88736

Table 1: Varying s values

Smoothing Parameter τ: The constant τ is responsible for smoothing bursts, so that erratic bursting behavior is penalized: a value of τ that is close to 1 imposes a small penalty for erratic bursting, while large values of τ creates larger penalties. Kleinberg uses $ln(n)$ for his value of τ, where n is the number of time steps. However, Kleinberg also models the document stream as continuous, so that each document arrives within a unique time step: there is a one-to-one correlation between the number of documents and the number of time steps, so that the value of τ is also representative of the overall number

of documents being analyzed. By contrast, my data is highly batched: a very large number of documents arrives over a very small number of time steps (9260 and 21, respectively), so that the ratio between documents and time is very large. Thus the small value of $ln(n)$ does not seem to adequately represent the overall volume of text being analyzed. Indeed, $ln(n) = ln(21)$ is approximately 3, while the average burst weight is 263; since burst smoothing is one of the desirable features of this algorithm, I tried tuning the value for τ so that more smoothing would occur.

By increasing the value of τ, we see fewer "orphaned" bursts, or distinct burst periods separated by non-burst periods (see Table 2). By the same token, we also see the average longest burst increase, as the automaton is discouraged from leaving a burst state when two burst periods are separated by a period of lower term usage. Finally, there is some small loss in the total weight, because we are abstracting away from perfect congruence with the data; however, this does not appear significant enough to outweigh the smoothing effect of increasing parameter τ. For the rest of my analysis, I used $\tau = 56$.

τ	# Bursts	Orphans	Longest	Weight
3	571	2.15	1.91	263217
12	543	2.07	2.00	262630
28	501	1.99	2.11	260799
56	424	1.87	2.28	255119
100	350	1.72	2.48	244865

Table 2: Varying τ values

3.3 Evaluating Burst Validity

Assessing Significance of Findings Using the idea of burst "weight", Kleinberg offers the "permutation test" as a way to determine whether the burst phenomena identified in a textstream are valid or spurious. The mechanism for this test is simple: the identification of burst phenomena depends on the order in which texts appear in the stream. Therefore, we test whether we have found valid bursts by randomly shuffling the order of documents in the stream, assessing the "bursty" behavior of its terms, and then comparing the sum of the burst weights for the original textstream with that of the randomly permuted one. To make this test more robust, we compute the average sum of burst weights over a number of randomly permuted corpora; we should find the average of these cumulative weights to be lower than the cumulative weight of our original, unshuffled corpus. If so, we can safely conclude that our identification of burst phenomena in the original unshuffled corpus was valid (Kleinberg 20).

As such, I constructed ten random document corpora, in which I kept the same number of documents per year by randomly shuffling the dates through the corpus. I then determined their average cumulative burst weights, as well as their average number of bursts, and longest burst. Table 3 represents these findings (std deviations in parentheses). The average cumulative burst weight for the randomized data is less than half that of the true corpus, and the average number of bursts and longest burst length are both significantly lower.

Run Type	Bursts	Longest	Weight
True Data	424	2.28	255119
Random	303 (12)	1.44 (0.04)	107500 (5869)

Table 3: Burst Output vs. Randomized Output

Presence of Outlier Texts Looking through the output that arose from running the algorithm on the randomized corpora, I noticed that a number of terms appeared to burst in every randomized document. Furthermore, some of these seemed like idiosyncratic terms; for instance, 'ut', 'pro', 'quod', and 'ad', all of which seem to signal a text written in Latin. I hypothesized that there might be some idiosyncratic texts in the corpus, in which such terms are so highly concentrated as to force a burst when they appear, regardless of other texts in the time step. This would then inflate the burst weight for the randomized texts. To test this, I did a quick heuristic test: for each term which bursted in all of the randomized texts, I simply removed the document with the most occurrences of that term from the corpus. I then re-ran the algorithm on the true corpus (minus these texts). I also generated a randomized corpus and ran the algorithm on both the full and the truncated versions of this randomized corpus.

Run Type	Bursts	Weight
True Data (Full)	424	255119
True Data (Trunc)	400	198038
Random (Full)	285	104310
Random (Trunc)	222	51623

Table 4: Removing High Term-Frequency Texts

The cumulative burst weight of the random text decreases by 50% (compared to the real data's 22%), while the number of bursts in the random texts decreases by 22% (compared to the real data's 5%. The weight and number of bursts in the true corpus are significantly less affected by the loss of those high-frequency texts than the randomized data, ostensibly because the real data captures collective behavior (which persists despite the loss of individual texts).

4 Results

4.1 Temporal Clustering of Bursts

When we plot the temporal location of bursts, a clear clustering effect is visible. Figure 1 is a heatmap of the temporal location for all bursts that last longer than one time step. The bursts are sorted by length of burst and then by weight before being plotted. This heatmap appears to delineate a sharp division between the years 1648 and 1649. High-weight/long-length bursts predominate in the years leading up to 1848, then fall away in 1649 to be replaced by a scattering of different terms.

Strengthening this image of a bifurcated timeline, Figure 2 plots the correlation in burst patterns between years (again, for bursts of length longer than 1). Again, the timeline appears to divide rather cleanly into two lobes: one consisting of the years between 1641 and 1648, the other beginning in about 1650 and continuing through to 1661.

This observation has strong historical face validity: 1649 was a cataclysmic year in British history [4]. In this year, the monarchy was abolished, King Charles I was executed, Parliament was purged of most of its members, and the Commonwealth regime was instituted under Cromwell. Political and religious institutions were utterly refashioned. It makes sense that bursts would cluster into two primary groups, to distinguish civil discourse "before 1649" from that "after 1649". However, the question then arises, what particularly did these linguistic changes consist of? How did they reflect or respond to these sociopolitical developments? To answer these questions, I characterize the temporal context within which the bursts occurred.

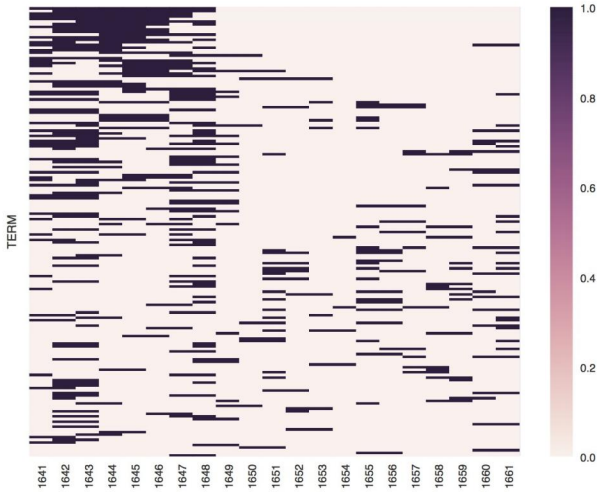

Figure 1: Terms by Burst Length, then Weight

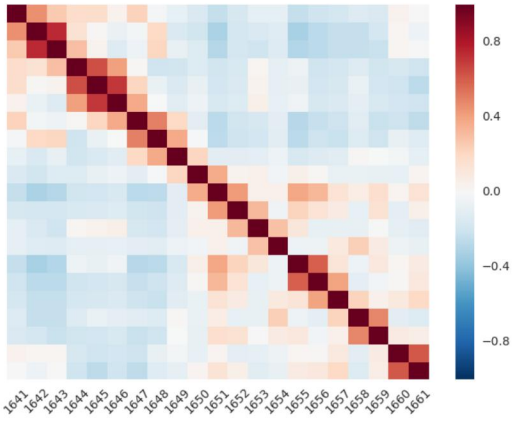

Figure 2: Year-Year Burst Correlation

4.2 Defining Burst "Context"

Kleinberg observes that his algorithm has the benefit of being able to easily identify "landmark texts", which are documents in which significant bursts (or groups of bursts) begin, or are sustained over time. The content of these landmark texts, then, serves to contextualize the terms which burst within them. I had hoped to find similar such "landmark texts" in my data. However, I did not fully appreciate, before I began, how the shape of the data Kleinberg used made it possible for him to do this (and very difficult for me to do it). As mentioned before, Kleinberg's data is structured so that the ratio of time steps to the number of documents is nearly one-to-one. Indeed, in both the infinite-state version of the automaton and in the historical State of the Union data, each document inhabits a unique time step, creating a perfect cor-

respondence between text and time. Because of this correspondence, it is easy to attribute a burst to its initial author, and to observe which authors continue to propagate it in future time steps.

By contrast, there are hundreds of texts per time step in my data, nearly all of which are highly-active burst participants. Of the 9256 texts in the corpus, only 284 do not participate in any of the bursts occurring during the year they were published. (I say that a document "participates" in the burst for term t if it demonstrates a frequency of use for t that exceeds the baseline probability p_0). Indeed, most documents participate in a large fraction of the bursts that occur during their publication year: on average, each document participates in 25.9 % of the year's bursts (std dev = 14.1). While I knew at the outset of this project that the large number of texts per year might make it difficult to parse out "landmark documents", I was simply not prepared to see such widespread activity. In this chorus of voices, it is difficult to isolate single agents from the crowd.

4.3 Method: Characterizing Trends

This does not mean, however, that I could not characterize "landmarks" or "context" in the corpus. Indeed, while his paper does not speak directly to the question of widespread burst participation, Kleinberg does observe that bursts might be a product of "collective activity" just as much as they might be "that of just a single user." (Kleinberg 20). Since the interesting bursts in my dataset seem largely to fall into the former category, I simply needed to change how I identified "landmarks" in the data.

When I initially devised the project, I defined the set of "landmark texts" as follows: for each time step n, K_n is the set of actively-bursting terms. Given this set K_n, L_n is the set of 'landmark texts', each of which contains one bursting term or a set of bursting terms. Near-universal participation in burst activity simply indicates that the set L_n of "landmark texts" for a period ought to consist of all the texts published in that period, and that the important differences to investigate within the dataset are those that occur *between* years rather than *within* years. Instead of inferring burst context from a small set of entrepreneurial documents, I will seek out "landmark years" to try to identify how the characteristics of language in each distinct time period distinguish it from other time periods. In what years, and in what ways, do major topics of discourse change?

To do this, I follow the following process: First, I use k-means clustering to generate groups of years based on the similarity of their term bursts (i.e. the output of the Kleinberg algorithm generated in the previous section). Second, I update the original document-term matrix to reflect the cluster membership of each document: I then aggregate all of the terms used in each year-cluster, and apply distinguishing-features measures (standardized mean difference and standardized log odds ratio[5, 6]) to determine which terms are most and least characteristic of each year-cluster. Finally, I reflect on the historical circumstances encompassed by each year-cluster, in conjunction with its list of most-characteristic terms.

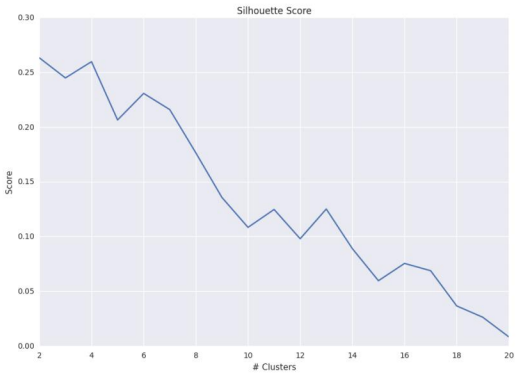

Figure 3: Silhouette Scores

4.4 Landmark Years: K-Means Clustering

I determined the number of clusters to use in the k-means analysis by selecting the number with the highest silhouette score. As Figure 3 demonstrates, a model with two clusters has the highest silhouette score, followed closely by a model with four clusters. I ran both 2- and 4- cluster models, each for 10 iterations. Figure 4 shows both the 2- and 4- cluster models plotted on a timeline.

2-Cluster Output:
Group 1 (1641-48), Group 2 (1649-61)
As the burst heatmap and correlation plot (Figures 1 and 2) led us to hypothesize, the 2-cluster kmeans model divides the years into a "before 1649" cluster and an "after 1649" cluster. All 10 iterations of kmeans arrived at this conclusion. As mentioned before, 1649 was a seminal year in British history, as it was the year that the monarchy was abolished (disrupting status quo church and parliamentary institutions as well). In its place, the Cromwellian Commonwealth was instituted: an unprecedented reestablishment of England as a republic, but also a regime with a strictly Puritan social and religious

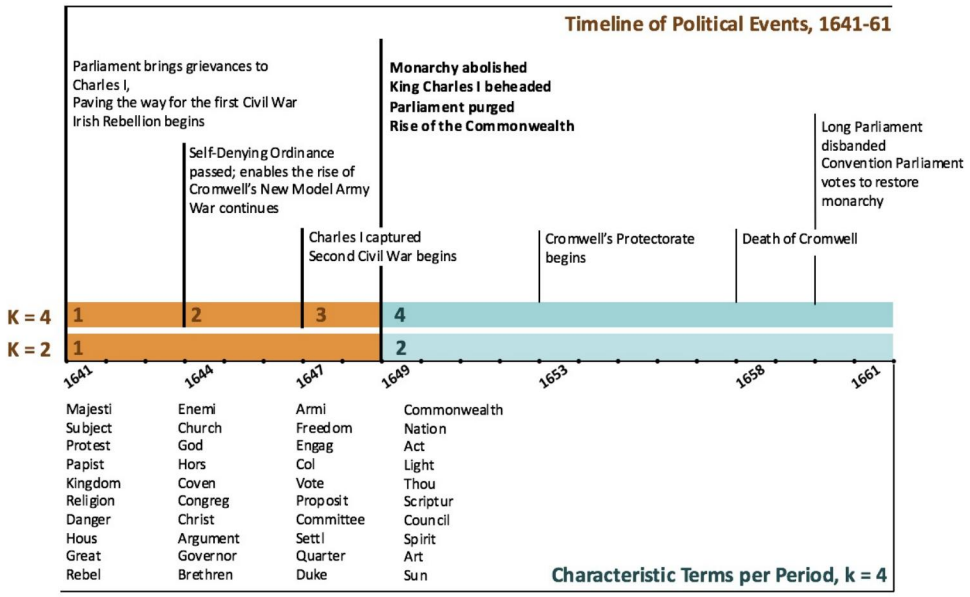

Figure 4: Burst Timeline and Discriminating Words

sensibility.

Both of the distinguishing-features measures clearly reflect the regime change which divide the two clusters. Figures 5 and 6 plot the output of the standardized mean difference and standardized log odds measures, respectively. In both cases, the most-distinguishing terms are those that pertain to the regime in power during the period: "kingdom" and "majesti" for the pre-1649 (or 'monarchic') cluster; "commonwealth" and "nation" for the post-1649 ('republican') cluster. Further investigation of the monarchic cluster reveals the preponderance of terms related to institutions which fell silent during the republican period: 'house', 'common', 'assembl', 'religion'. Though both Parliament and the Church of England did persist through the Cromwellian period, both were disempowered. Linguistically, both were replaced by a less-worldly Puritan lexicon: 'scripture', 'spirit', 'father', 'soul', 'rightous'.

4-Cluster Output:
Group 1 (1641-43), Group 2 (1644-46)
Group 3 (1647-48), Group 2 (1649-61)
In the 10 runs of 4-cluster kmeans, 7 runs came up with this grouping of years (this is the grouping that I used for the subsequent determination of distinguishing features for each cluster). In two runs, the year 1647 moved from Group 3 into Group 2, leaving 1648 in an isolated group by itself. And in 1 run, the year 1649 moved from group 4 into group 3. However, the fundamental divide between the monarchic and republican periods remains throughout. Figure 4 shows the timeline of the period, along with the terms most characteristic of each cluster.

The distinguishing features for these clusters refer more particularly than the 2-cluster model to specific military, religious, and political conflicts that occurred during the republican period. For instance, Cluster 1 contains the term 'rebel', signalling the beginning of the Irish Rebellion. Clusters 2 and 3 both contain military language 'armi', 'hors', 'enemi', 'quarter', reflecting the battles ongoing during the period. The terms in Cluster 4, which encompasses the whole republican period, are largely unchanged from the 2-cluster model.

5 Conclusions and Future Work

Burst detection and characterization provides a promising method of grappling with and analyzing a large corpus of disparate documents, identifying major shifts in language use patterns, and imputing these shifts to the broader temporal context in which they appear.

The analysis was, however, limited by the coarseness of the time steps: it was difficult to parse out the nuanced variation that occurred within each year. For the texts in this dataset, the month of publication is generally referenced within in the text, but not in the easily-accessed metadata. If a method could be devised to extract the month of publication as well as the year,

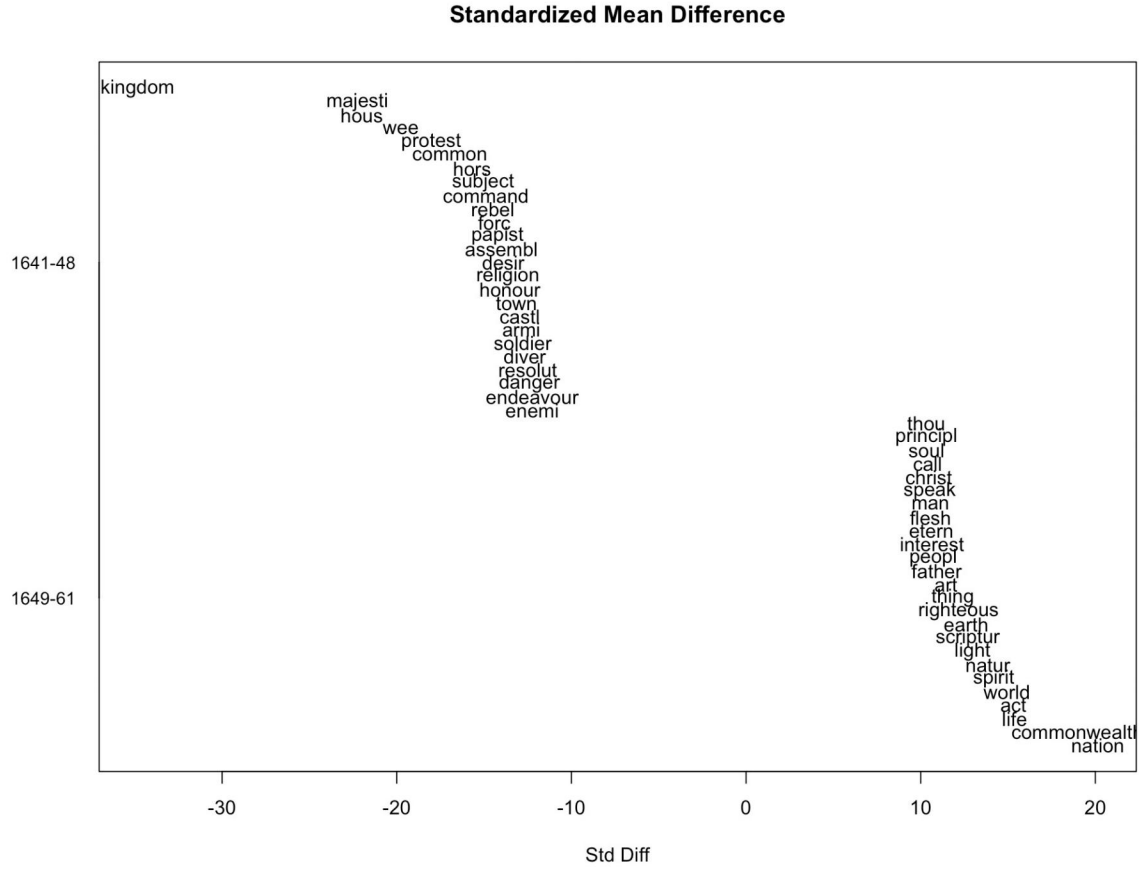

Figure 5: 2-Clusters, Distinguishing Terms

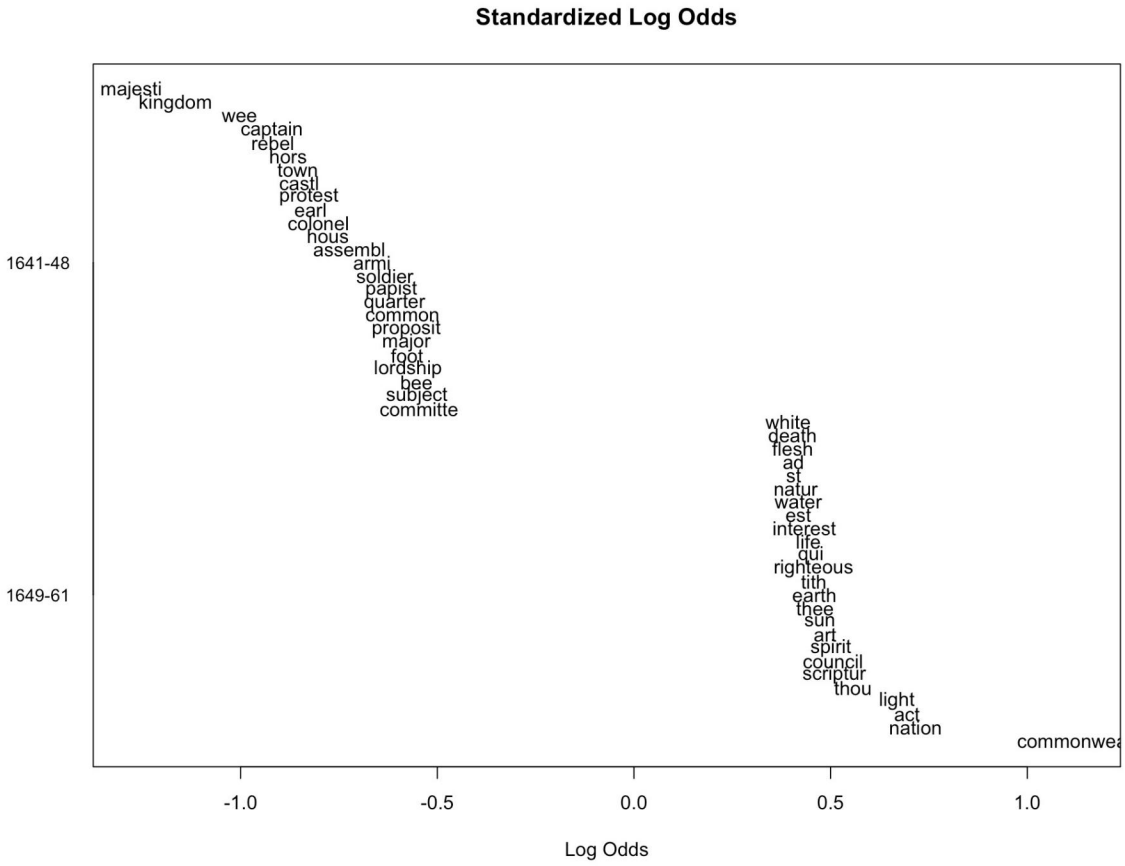

Figure 6: 2-Clusters, Distinguishing Terms

some of the problems that arose from the coarseness of the time steps might be overcome, allowing for more nuanced, particular analysis.

References

[1] About VARD 2. URL http://ucrel.lancs.ac.uk/vard/about/.

[2] The Porter Stemming Algorithm. URL https://tartarus.org/martin/PorterStemmer/.

[3] Stanford Social Science Data and Software (SSDS). URL https://ssds.stanford.edu/.

[4] Juliet Gardiner and Neil Wenborn, editors. *The Columbia Compaion to British History*. Columbia University Press, 1997.

[5] Justin Grimmer, . Class Notes, Lecture 5, PoliSci 452 (Fall 2014).

[6] Justin Grimmer, . Class Assignment, Problem Set 3, PoliSci 452 (Fall 2014).

[7] B. Stewart J. Grimmer. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 2013.

[8] C. Faloutsos J. Lescovec, M. McGlohon. Cascading behavior in large blog graphs: Patterns and a model. *Proceedings of the 2007 SIAM International Conference on Data Mining*, 2007.

[9] J. Kleinberg J. Lescovec, L. Backstrom. Memetracking and the dynamics of the news cycle. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009.

[10] M. Meiss-et al. J. Ratkiewicz, M. Conover. Detecting and tracking the spread of astroturf memes in microblog streams. *Proceedings of the 20th International Conference Companion on World Wide Web*, 2011.

[11] Jon Kleinberg. Bursty and hierarchical structure in streams. *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, 2002.

[12] A. McQueen L. Blaydes, J. Grimmer. Mirrors for princes and sultans: Advice on the art of governance in the medieval christian and islamic worlds. forthcoming.

[13] M. Mathioudakis and N. Koudas. Twittermonitor: Trend detection over the twitter stream. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2010.

[14] Introducing DREaM (Distant Reading Early Modernity). URL http://earlymodernconversions.com/introducing-dream/.

[15] Early English Books Online. About the thomason tracts. URL http://eebo.chadwyck.com/about/about.htmtract. Accessed: 2016-11-16.

[16] Text Creation Partnership. URL http://www.textcreationpartnership.org/.

[17] R. Al-Rfou S. Skiena V. Kulkarni, B. Perozzi. Statistically significant detection of linguistic change. *Proceedings of the 24th international conference on World Wide Web (WWW '15)*, 2015.