Vision, Modeling, and Visualization (2023)
T. Grosch and M. Guthe (Eds.)

Visually Analyzing Topic Change Points in Temporal Text Collections

Cedric Krause[1], Jonas Rieger[2], Jonathan Flossdorf[2], Carsten Jentsch[2†], and Fabian Beck[1‡]

[1]University of Bamberg, Germany
[2]TU Dortmund University, Germany

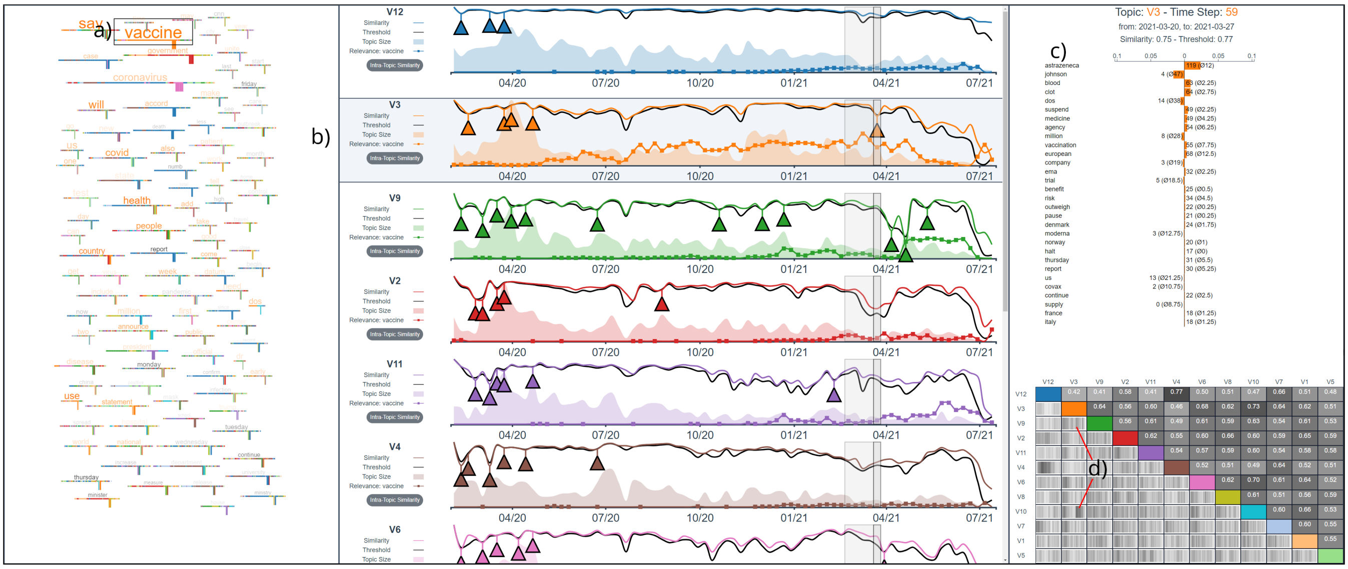

Figure 1: *Our visualization system displays a COVID-19-related news dataset. The selected word* 'vaccine' *(a) is important in the selected topic V3 (b). A triangle marks a detected change point in V3 around March 2021, when the word* 'astrazeneca' *(c) gained substantial importance and the topic possessed a comparatively high similarity with topics V9 and V10 (d).*

Abstract

Texts are collected over time and reflect temporal changes in the themes that they cover. While some changes might slowly evolve, other changes abruptly surface as explicit change points. In an application study for a change point extraction method based on a rolling Latent Dirichlet Allocation (LDA), we have developed a visualization approach that allows exploring such change points and related change patterns. Our visualization not only provides an overview of topics, but supports the detailed exploration of temporal developments. The interplay of general topic contents, development, and similarities with detected change points reveals rich insights into different kinds of change patterns. The approach comprises a combination of views including topic timeline representations with detected change points, comparative word clouds, and temporal similarity matrices. In an interactive exploration, these views adapt to selected topics, words, or points in time. We demonstrate the use cases of our approach in an in-depth application example involving statisticians.

CCS Concepts

• ***Human-centered computing*** → ***Visual analytics;*** • ***Mathematics of computing*** → ***Time series analysis;***

† e-mail: jentsch@statistik.tu-dortmund.de

‡ e-mail: fabian.beck@uni-bamberg.de

DOI: 10.2312/vmv.20231231

1. Introduction

In large text corpora, like news collections, diverse themes are covered. Topic extraction methods can automatically identify topics by grouping related terms. This is helpful in various scenarios, from the identification of articles of interest for a person, to the support for a general overview of the text collection. With new texts being added over time, the coverage of the derived topics changes. Analyzing current media coverage incrementally, algorithmic methods employ statistical indicators to detect change points at which a topic is sufficiently different from its previous state. However, these methods often appear as black-boxes as they do not explain these changes. In order to understand the result, it is necessary to analyze all output artifacts of the models. Inspecting the contents of the derived topics, their similarities with each other, and change points within the topics that occurred over time can reveal rich insights about the underlying data.

As a team of statisticians and visualization researchers, we co-developed an approach to enhance the visual analysis of such models' results. We have focused on a specific dynamic topic extraction method and used a rolling version of Latent Dirichlet Allocation (LDA) [RJR21] on a CNN news collection about COVID-19 from 2020 and 2021. We applied change point detection [RLFJ22] to the derived topics to statistically identify abrupt changes in the topics' vocabularies. We have identified relevant analysis questions to reason about the quality of the topic extraction and change detection. Based on the analysis questions, we designed a conceptual approach to address these questions, implemented a corresponding visualization system, and classified common change patterns that we identified while using our system.

Our visualization system (Figure 1) shows a word cloud that adapts word sizes and other encodings to the selection topics and time steps. Each topic is individually represented in a timeline visualization, including detected change points within the vocabulary of the topic. A butterfly bar chart in the change detail view displays which words in the dynamic vocabulary changed most noticeably and caused the detection of a change point. Moreover, we use similarity matrices to visualize intra-topic similarity, as well as pairwise similarity between different topics over time.

Our application study, hence, suggests a tailored visualization solution for a specific statistical approach, targeting experts as users. This limits the immediate applicability of our approach. However, our contributions span broader, as they also include results transferable to related problems and interfaces: (i) analysis questions to analyze temporal text collections along topics and detected change points, (ii) reusable concepts for the visual analysis of dynamic text collections from overview, over topic evolution, to detailed inspection of change points, (iii) a classification of observed change patterns based on the characteristics of change points, and (iv) exemplary insights from an extensive application example with real-world data.

The implementation of the prototype is publicly available under an open-source license (`https://github.com/vis-uni-bamberg/topic-change-vis`), and the supplemental material of this paper includes a video demonstration of the tool.

2. Related Work

We are not the first to visualize changing topics in document collections; approaches are numerous and constitute separate sections in literature surveys on visual text analysis [DL16] and visualization of scientific literature topics [ZLZ17]. For instance, the well-known *ThemeRiver* [HHWN02] approach visualizes changing topic sizes as a variant of stacked area charts on a timeline. Other approaches adapt this metaphor and, for instance, extend it to branching and merging streams [CLT*11; LYW*16; XWW*13; CLWW14]. But also small multiple representations are common, where different topics are shown as different rows on the same timeline [DWS*12; LYK*12; KBK11]. We decided to use such small multiples for the topics because our scenario did not allow computing quantifiable exchange of content between topics, and stacking streams would have generally limited the options to encode additional information within the topic representations.

Many approaches have already considered specific events that mark changes in the dynamic topics. Within branching and merging streams, the branch and merge points constitute a form of discrete event. These are detected, for instance, by incrementally updating topic models and applying defined branch and merge criteria [CLT*11; GJG*15] or evolutionary hierarchical topic trees combined with a dynamic tree cut method [CLWW14; LYW*16]. Some change events are specialized to certain applications, for instance, whether a topic changes from cooperation to competition within a debate [SWL*14]. Other techniques consider events within consistent topics. These can be the real-world events derived from keywords that were used within a topic's documents. The emergence of coverage of that event can also be considered a change point in the topic. *EventRiver* [LYK*12] visualizes these events as drops that encode the number of articles that are closely related content-wise and time-wise. Lu et al. [LWLM18] follow a similar procedure, but allow the user to handpick events and annotate the topic in a more user-guided analysis. *LeadLine* [DWS*12], *TopTom* [GBM*19], and *TwitInfo* [MBB*11] focus on change points within consistent topics. Using statistical methods, they identify peaks and valleys in the number of articles covering the topic. In contrast, we focus on structural breaks in the statistical analysis of the vocabulary used in the topic. We are not aware of any approach with a similar focus.

While we employ and adapt various standard visualization techniques in our approach, our use of word clouds might be the most specialized. For our topic- and time-aware, interactive word cloud, we drew inspiration from several previous works. Time-aware word clouds, like *PyramidTags* [KKE21], have been proposed to help analysts explore dynamic document collections. Together with topic extraction, they have been utilized to either display relevant words on the topic streams themselves [LZP*12] or can be accessed through an on-demand lens [XWW*13]. The word cloud visualization we use is also adaptive to the topics and time steps, but keeps a global layout, to enable comparison of the words across time and topic boundaries. *RadCloud* [BLB*14] uses stacked bar charts below the words to encode their relevance in multiple categories. We adopt this idea and use scarfplots in our word cloud to attribute words to topics at a glance.

3. Analysis Questions

Agnostic of the concrete methods used for topic extraction and change point detection, we formulate analysis questions in order to reason about the quality of such models. Based on our collaboration between visualization researchers and statisticians, we have tried to capture those questions statisticians would explicitly or implicitly want to answer when working with a visualization approach to analyze their methods and explore their results. While initial questions were formulated early and guided the design of the system, we have refined and consolidated the questions iteratively throughout the process. In the paper, we use them as a reference to connect the concepts and results discussed in different sections.

We start from a collection of documents and extract a set of *topics* at different points in time. Each derived topic may reflect multiple real-world *concerns* (e.g., a person, an entity, a concept), and it is crucial to understand what the topics capture exactly. The general *importance* of a topic may vary over time, as well as the concerns that are connected. *Similarity* between topics is expressed by the partial overlap or inherent semantic connections of their concerns. Understanding this similarity and also identifying if concerns shift between the topics provides important context to make sense of the temporal evolution of the topics.

> ***AQ 1** – Topic Evolution*
>
> **AQ 1.1** What are the importance and concerns of a topic (over time)?
> **AQ 1.2** How similar are topics to each other (over time)?
> **AQ 1.3** Do concerns move from one topic to another?

The temporal development of the extracted topics can be accompanied by a number of *change points* at which its deviation from previous states is declared as substantial. Causes for such deviations are typically new developments and events that lead to adaptions in the concerns covered by the topic, and they can potentially impact multiple topics. Each change point is further characterized by its context within and beyond the topic (e.g., the topic was not important at the time of the change).

> ***AQ 2** – Change Points*
>
> **AQ 2.1** When do change points occur and what are their characteristics?
> **AQ 2.2** What shifts in the concerns caused a change point?
> **AQ 2.3** Do change points share characteristics and causes?

The grouping into *Topic Evolution* and *Change Points* structures the questions according to the two abstraction steps made (*documents* → *topics* and *topics* → *change points*), but does not divide the analysis. The questions may contain connections across the boundaries of the grouping. Observing a peculiar instance in either might initiate a new investigation in the other block. For instance, an analysis of specific change points can lead to questions of the general importance of a topic and finding related topics.

4. Application Scenario

We focus our application study on a specific scenario. On the one hand, this is necessary to operationalize, for instance, how *concerns* are described that reflect in a topic or how *similarities* of topics can be computed. On the other hand, we want to provide in-depth insights for statisticians that consider the specifics of a method. Hence, our approach can be considered a *white-box analysis*.

Data As input we expect a document collection as a set of texts timestamped by publication date. We split the considered time frame into discrete time steps so that each time step is covered by a substantial set of new documents. While our approach works with any such set of texts, we use a collection of news by CNN on COVID-19-related articles as an application example throughout the whole paper. The texts were scraped based on the script by Pasquali et al. [PCR*21]. The dataset consists of 27 432 articles, having a median length of 82 words (142 before preprocessing), and a total number of 35 544 distinct words (44 605 before lemmatization). The articles are all written in English and were published from early 2020 to mid of 2021. Divided into weekly bins, this resulted in 79 time frames, however, with lower coverage towards the end; to avoid time steps without any documents, we considered only the first 76 out of 79 time frames. We have selected this example because of its broad relevance and understandability, and specific insights can easily be checked with other sources.

Topic Extraction and Similarity We first conduct common preprocessing steps to format characters to lowercase, remove numbers and punctuation, apply a list of trusted stopwords and a lemmatization dictionary, and remove words with fewer than two characters. We then use LDA [BNJ03] to derive 12 latent topics in the data. It assigns each occurrence of a word to one of the topics based on its prevalence of occurring together with other words. Words from the same document can be assigned to different topics, since documents may touch upon multiple concerns (Figure 2, left). For the dynamic collection, we use a rolling version of LDA [RJR21], which, in comparison to other dynamic versions of LDA, allows keeping topics consistent over time, as words in newly added documents are allocated based on the topics in previous time steps as well. For our scenario, we defined a reference period of up to four weeks (or shorter if a change point was detected within that time) and weighted its vocabulary with 15%, in contrast to the 85% of the new vocabulary of the time step. The parameter selection is rather conservative and leads to comparatively few change points. Figure 2 (right) illustrates that, using this method, we obtain a temporal sequence of changing vocabularies that describe the extracted topics (**AQ 1.1**). The *pairwise cosine similarity* detects similarities between topics, even if the sizes of the topics are different (*inter-topic similarity*). It can be computed globally across all time steps or locally regarding a specific time step—the similarity may change as a pair of topics can be similar for a given time frame but deviate again later (**AQ 1.2**). To learn how a single topic changes (**AQ 1.1**), the similarity measure can be applied internally by comparing different points in time (*intra-topic similarity*).

Change Point Detection and Characteristics To detect abrupt topic changes (**AQ 2.1**), we employ a recently published change

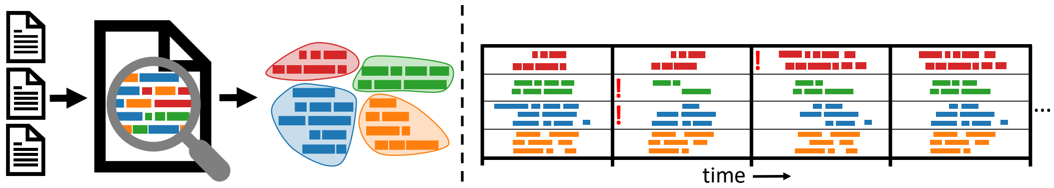

Figure 2: *LDA allocates the words in each document to a topic, then topics are represented by sets of words (left). Dynamic LDA results in dynamic vocabularies where, if the vocabulary of a topic is sufficiently different from previous states, a change point is detected (right).*

point detection method [RLFJ22], which computes similarities of the topics regarding a predefined reference period of preceding time steps. It has two parameters that mainly control the sensitivity of the algorithm: the maximum length of the reference period with which a topic at a new time step is compared and a weighting parameter, which regulates the minimum intensity for a change to be detected. Based on this predefined reference period and weighting, the cosine similarity of the vocabularies indicates whether the documents from the new time increment have a relevant impact on the topic. A change is detected if the similarity falls below a dynamically adjusted threshold that accounts for varying topic sizes over time (i.e., with a smaller sample size—the number of words—a topic is expected to vary more). When a topic at a given time step is less similar to its previous states than expected, we can draw conclusions from it. First, the difference between similarity and threshold indicates how different the vocabularies were in the time step compared to its predecessors. Secondly, the exact values of the similarity and the threshold can contain additional information on whether the topic was stable and we expected a high similarity, or whether it fluctuated, and we already expected a low similarity that was yet undergone. Together with general properties of the topic at this point, this provides important characteristics to interpret the change and the circumstances of its detection (**AQ 2.1**).

Change Cause For each time step, and especially for each change point, we can investigate the impact of changed frequencies for certain words, by comparing the similarity of the topic with its previous state when leaving the word out. The words that had the biggest impact and how their frequency increased/decreased can give a thorough impression on why a topic actually changed in the given time step (**AQ 2.2**). Furthermore, shifts of concerns from one topic to another can be detected through such analysis (**AQ 1.3**).

Interlinked Topic Changes We consider the aforementioned aspects to reveal common or related changes within the topics (**AQ 2.3**). For example, a topic that changed through the emergence of a few new words may experience another relevant change if said words disappear from the topic again just as abruptly. Some changes in the documents of a text collection might be large or cross-cutting enough such that they have a relevant impact on multiple topics. Either the emergence of a new word is so present that it starts to occur very frequently in all topics at the same time, or the same new word might impact different topics at different times. Reasoning about change causes may support identifying the words that caused relevant changes not only in different topics, but also at different times.

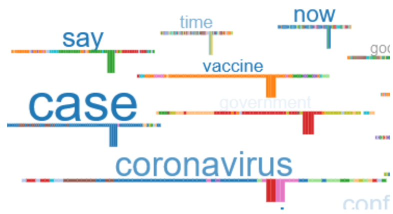

Figure 3: *The opacity of word changes based on their relevance to the topic at a selected time step (**AQ 1.1**). While* ‘coronavirus’ *is important in the topic over all, the enlarged segments of the scarfplot show that it was more relevant in other topics at the time. On the other hand,* ‘case’ *is most relevant to the selected topic.*

5. Visualization Approach

We propose a visualization approach that enables users to explore the output artifacts of the topic extraction and change detection methods. The user interface shown in Figure 1 visually links related information and adapts to interactive selection of the color-coded topics, time steps, and words.

5.1. Adaptive Word Cloud

On the left, the word cloud visualization provides an overview of the top 100 most frequent words in the entire dataset. Size encodes a word's frequency on a square-root scale with a minimum and maximum value for legibility. In addition, the frequency is redundantly encoded in the position, with the words placed greedily towards the top left corner and smaller words filling the gaps between larger ones. Reflecting the temporal changes in which topics a word was used, we introduce a scarfplot below each word. It consists of segments equal to the number of time steps in the dataset. Each segment adopts the color of the topic in which it possesses the highest relevance at the respective time step. Relevance is defined as the share of the word's occurrences relative to the number of total words at the time step (in the topic). This encoding displays whether a word moved from the context of one topic to another over time (**AQ 1.3**).

The word cloud is topic-aware and adjusts the size of the words to the frequency of the words within a selected topic (**AQ 1.1**).

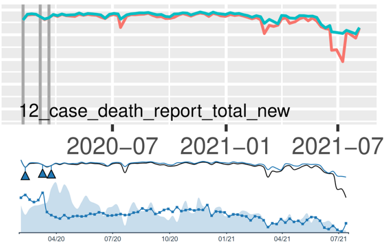

Figure 4: *To the similarity and the threshold in the previous visualization (top) [RLFJ22], our new visualization (bottom) adds the size of the topic as well as the relevance of a selected word over time (here: the word* ‘case’ *in topic V12) (* ***AQ 1.1****).*

Words that surpass a certain frequency threshold within the topic are also colored according to the selected topic to emphasize their relevance. To keep a stable mental map, the layout remains unchanged. To avoid overlaps between words, this implies that the size of the words cannot grow based on the topic selection and, instead, must be strictly less than before. The relevance of a word in a topic relative to the overall collection can be perceived by the degree to which the word shrinks. The length of the scarfplot always keeps the length of the word's original size to allow for an easier comparison. If the user selects a time step, the word cloud adapts in two ways as well (Figure 3). First, the opacity of each word changes based on its relevance at the given time step. Words that were used frequently overall, but rarely at the selected time step, will fade out and leave the words describing the time step more accurately clearly visible (**AQ 1.1**). The scarfplot also highlights the selected time step and its reference period by enlarging the respective segments.

5.2. Topic Timelines

At the center of our interface is the list of aligned topic timelines, one per topic and sorted with the largest topics at the top. We decided against a flow metaphor for representing the topics, such as used by Gad et al. [GJG*15], as our topic detection incrementally updates a fixed number of topics. Within each timeline, we visualize four time series as explained in the legend on the left. The timeline is an extension of the visualizations in the original publication of the statistical methods [RLFJ22] (Figure 4). The similarity and dynamic threshold—already included in the old version—are calculated based on the reference period in a rolling time window. In our visualization, the similarity is the solid line in the color of the topic, whereas the threshold, as the baseline, is always black. The detected change points are marked by triangles at points where the similarity fell below the threshold (**AQ 2.1**). An additional area chart encodes the size of the topic (**AQ 1.1**), and makes clear how the value of the threshold is partly influenced by the topic size. Moreover, we added functionality to display the relevance of an interactively selected word in the topic over time. The relevance is plotted in a line chart with rectangular glyphs for each time step in which the word occurred in the topic. Comparing the relevance over multiple topic timelines provides a more detailed picture compared to the scarfplot in the word cloud, where each segment is assigned based on a *winner-takes-all* selection (**AQ 2.3**). Through the timelines, the user can select both, a topic and a time step, which is propagated to the word cloud and change detail view as well.

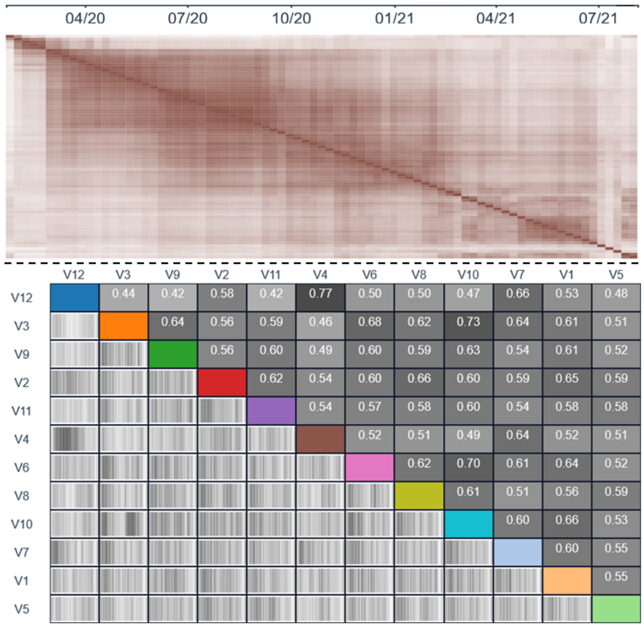

Figure 5: *The intra-topic similarity matrix (top) shows the pairwise similarity between time steps within the same topic (* ***AQ 1.1****). The inter-topic similarity matrix (bottom) shows the pairwise similarity between topics at the same time step over time in the bottom half and the 0.95-percentile value in the top half (* ***AQ 1.2****).*

5.3. Temporal Similarity Matrices

On demand, the intra-topic similarity across the whole time span can be accessed (Figure 5, top). On the same horizontal axis as the timeline, the pairwise cosine similarity of the vocabulary between time steps in the same topic is visualized in a matrix. The intra-topic similarity matrix reveals time spans during which the topic kept a similar vocabulary beyond the reference period, and also time spans where the vocabulary changed more drastically (**AQ 1.1**).

To analyze the similarity between different topics over time, we also introduced the inter-topic similarity matrix (Figure 5, bottom). Every topic has a row and a column. Cells where different topics intersect in the top half contain the 95-percentile of similarities between the two topics. Using the percentile instead of the average points out strong relationships more clearly. However, the exact choice of the percentile depends on the dataset (e.g., the number of time steps). The bottom half of the matrix shows the course of similarities between the two topics over time and reveals time spans where two topics were particularly similar or dissimilar (**AQ 1.2**).

5.4. Change Detail View

Selecting a topic and a time step, relevant changes in word frequencies are shown in detail, even if no change point was detected. Since

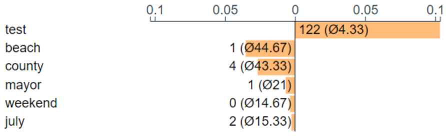

Figure 6: *The words with the highest impact on similarity that led to the detection of the change point in step 25, topic V1 (* **AQ 2.2***).*

change points are triggered by gapping cosine similarities, seeing how individual words impacted this gap helps understand how the topic has changed. The top 30 words regarding similarity change—based on *leave-one-out* calculations—are listed in the top right of the system (**AQ 2.2**). For each word in the list, we contrast the frequency of the word in the selected time step with the average frequency in the reference period. The length of bars in a butterfly chart shows the word's negative or positive impact on the cosine similarity in the time step (Figure 6).

6. Change Patterns

We have identified a set of generalizable change patterns within the topics. Their identification typically requires looking at different visualizations and combining answers of several analysis questions (Table 1). In the following, we describe the patterns along their data characteristics and classify them by the number of change points.

No Change Point Our approach enables users to also explore the topics at time steps where no change point was detected.

- **Stable behavior:** A stable topic is typically indicated in the timeline by both, high absolute values for similarity and threshold, as well as few dips in these time series. It can additionally be confirmed by investigating the intra-topic similarity.
- **Volatile behavior:** Some topics do not always have sharp boundaries. They can either be inherently volatile (e.g., political coverage concerning daily events) or can be a loose composition of words not fitting well the other topics. Topics with such a behavior commonly display in the timeline comparatively low absolute values for similarity and threshold, with frequent swings in them.
- **Slow topic drift:** The dynamic change detection method we use is specifically designed to identify abrupt changes, but slow and steady changes might go unnoticed by the method. However, the intra-topic similarity matrix can reveal such drifts by showing a gradient getting lighter between detected change points. It is further interesting to investigate in the inter-topic similarity matrices whether other topics take up parts of the drifting topic.

Single Change Point A change point always indicates a difference that is caused by substantial changes in the topic, but circumstances might be different. Mainly, our detail view can reveal the impact of individual words and what exactly caused the change.

- **Single-concern dominance:** Some change points are caused mostly by the changed frequency of a single word. Closely related, some change points might be caused by multiple words that belong to the same concern (e.g., *'mask'* and *'mandate'*).

Table 1: *Mapping of topic change patterns to visualization views and analysis questions;* ● clearly relevant, ○ partly relevant.

Change pattern	*Cloud*	*Timeline*	*Matrices*	*Detail*	AQ 1.1	AQ 1.2	AQ 1.3	AQ 2.1	AQ 2.2	AQ 2.3
No Change Point										
Stable behavior		●	○		●					
Volatile behavior		●			●					
Slow topic drift	○	○	●		●	○	○			
Single Change Point										
Single-concern dominance		○		●	○				●	
Multi-concern change		○		●	○				●	
Multiple Change Points										
Toggle		●	●	●	○	○		●	●	●
Multi-step progression		●		●	○			●		○
Synchronous change	○	●	○	○	○		●	●	○	●

- **Multi-concern change:** In contrast, change points can also be caused by the impact of changed frequencies of multiple, not directly related words. These words seem to originate from different concerns but change frequencies together.

Multiple Change Points Sometimes, multiple change points are connected. This information is lost when investigating each change point only individually, but the aligned topic timelines of our approach reveal co-occurrences and temporal progression.

- **Toggle:** A new concern that constitutes a change point might disappear again as abruptly (either quickly or after multiple time steps), so it causes another change point with a similar set of words as the initial change point. Aside to the timeline, the intra-topic similarity matrix and detail view help reveal such patterns.
- **Multi-step progression:** Some concerns evolve across several steps. Their importance to the topic is so large that the changes in the words' frequencies are sufficient to trigger multiple change points. This pattern is characterized by successive change points describing a similar development (e.g., first the closing of restaurants, then the closing of schools, and then the ban of airplane travel). The major words as shown in the detail view are, typically, different but semantically connected.
- **Synchronous change:** The same term can be used in different contexts and are then is likely assigned to multiple topics in the same time step. If the appearance of these words is abrupt in several topics, it is possible for change points in different topics to be traced back to the same source. Another manifestation of this pattern is different words causing change points in different topics, although they stem from the same concern.

Table 1 details the mapping of change patterns to the visualizations and analysis questions. The timeline visualization contains the change point markers and is naturally relevant across the board. The change detail view, as the source of information on explaining change points, is just as straightforwardly tied to the change patterns that contain at least one change point. On the other hand, the word cloud visualization and analysis question **AQ 1.2** are not directly reflected in the patterns. The word cloud displays an overview of the whole time span or can focus on a snapshot of the concerns at a particular time step, but its only capabilities of representing change is in the scarfplots. Similarly, **AQ 1.2** is not targeted at specific temporal changes and is more cross-cutting.

7. Results

Analyzing the data with our new visualization approach, first, we can get an overview of the major concerns in the text collection by looking at the word cloud visualization in its initial state. With the biggest and most frequent words in the collection towards the top left, we quickly observe many COVID-19-related words like *'coronavirus'*, *'case'*, and *'vaccine'*. Looking further, we also identify words that give more context, like *'us'* (United States), or *'test'*. Considering all words, we can infer that the text collection comprises different perspectives, be it geographical or others.

The area chart that encodes the topic size in each timeline shows no major differences in overall topic size, but that the topics differ substantially at specific time steps. We observe an overall trend of topics shrinking over time, which was not visible with the older visualization technique shown in Figure 4 (**AQ 1.1**). The biggest topic (V12) at the top is very stable regarding similarity and its threshold. It has three change points at the beginning, all with a minimal difference between similarity and threshold. Afterward, it displays **stable behavior**, with high absolute similarity values and no observable trend across time in terms of most impactful words (**AQ 1.1**). The intra-topic similarity matrix for V12 shows a uniform high similarity throughout the time span, and the topic-adjusted word cloud supports this. The major words are *'case'*, *'death'*, and *'report'*, which we would expect to find in a recurring topic that mostly concerns the reporting of case numbers.

The inter-topic similarity matrix reveals that the reporting topic V12 is similar to topic V4 in the first half (**AQ 1.2**). The adjusted word cloud for topic V4 shows *'case'* as an important word, too. Through the scarfplot, we can infer that, while the word was most relevant in V12, V4 utilized the word even more during a few periods of the first half. Selecting the word, we see that its importance is high in the first half and slowly decreases from around September 2020. The topic's other important word *'state'*, however, stays important through the whole topic. Interactively investigating more time steps and their highest similarity impacts in the change detail view, we also see frequently changing US state names (Figure 7). We infer that topic V4 mostly concerns US news at state-level and exemplifies a **slow topic drift** from case reporting (which constituted the similarity to V12) to more general news (**AQ 1.1**). This aligns with the intra-topic similarity matrix in Figure 5 (top), which shows decreasing similarity even though no change points were triggered for a long time span.

Two topics that have different characteristics from the aforementioned ones are V1 and V9. In general, we can see many more change points in the topic timelines (**AQ 2.1**). Investigating major words in the topic-adjusted word clouds, we found them to be political topics from the UK (V1) and US (V9) (**AQ 1.1**). The fast-paced coverage of political news causes these topics to have a lower overall intra-topic similarity, which is visible in the similarity matrices within the topics and the similarity and threshold time series. Figure 8 shows that even in times of no detected change points, these topics display the pattern of **volatile behavior**.

Topic V8 concerns many stories about events in different countries and their implications on travel. The nature of such events to occur abruptly is reflected in the change points as well. They

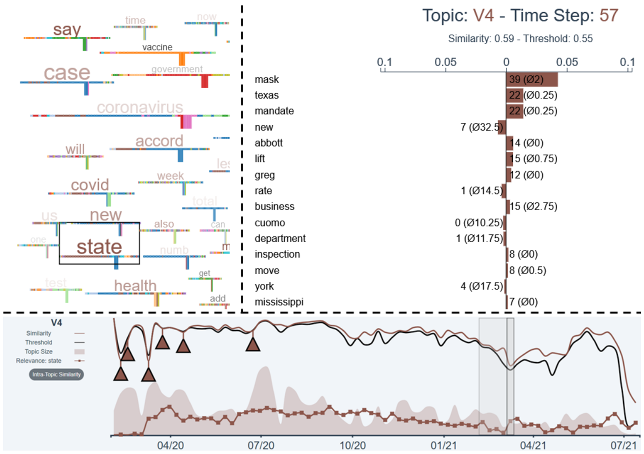

Figure 7: *While* 'case' *is relevant in topic V4, considering the whole time span, it becomes less important towards the end. The word* 'state' *remains important throughout.*

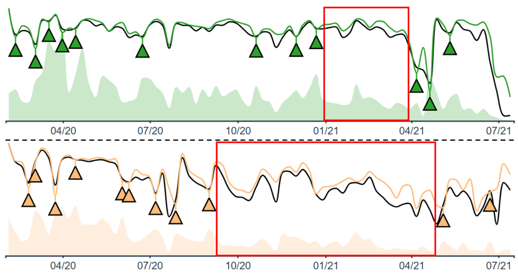

Figure 8: *The similarity and threshold line for topics V9 (top) and V1 (bottom) are fluctuating but no change point is detected, which is an indicator for the* ***volatile behavior*** *pattern.*

are distributed quite evenly over the time span, with relatively stable periods in between (**AQ 2.1**). The change points themselves are then mostly following the **single-concern dominance** pattern (**AQ 2.3**), which we can clearly see in the butterfly chart of the change detail view. Examples are the sudden shortage of oxygen supplies in Indian hospitals in early May 2021, or the identification of a new coronavirus UK variant in December 2020 (**AQ 2.2**), both in topic V8. Selecting the word *'variant'*, we observed its rise in topic V8, but also in topic V6. Similarly, we identified another time step in the topic timelines at which a single story impacted multiple topics in December 2020, when the first vaccines were available and the vaccination process was rolled-out. This story constituted **synchronous changes** with topics V9 and V10, as confirmed in the change detail view (**AQ 2.1**, **AQ 2.3**).

We noticed many **toggle** patterns. In some cases, the second change point occurred just one time step after the first, like in the case of UK politician Dominic Cummings violating restrictions with a trip to Durham in May 2020 within topic V1. A week later, the absence of the story caused the next change point (**AQ 2.2**). In other cases, there were 2–4 weeks before the second change point of the **toggle** pattern, like in the case of the words *'bill'* and *'stimu-*

Proceedings published by Eurographics - The European Association for Computer Graphics.

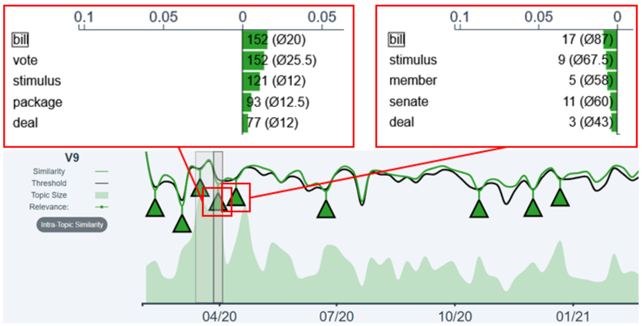

Figure 9: *Detailed inspection reveals the* ***toggle*** *pattern, as the same words cause change points with their appearance and disappearance in quick succession in the same topic. Both change points individually also represent the* ***singe-story dominance*** *pattern.*

lus' in March/April 2020 in V9 (Figure 9) or *'beach'* and *'county'* in July 2020 in V1. The second change point of the latter case also exemplifies the **multi-story change** with the word *'test'* seemingly replacing *'beach'* and *'county'*, which disappeared, as is visible in the change detail view (**AQ 1.1**, **AQ 2.2**) (Figure 6). Within all **multi-story changes**, we only identified those where one story left the topic and another entered. We attribute it to LDA that two new topics never enter the same topic at the same time step.

The multiple change points (**AQ 2.1**) at the beginning of topic V11 are related to each other as they describe developments on restrictions like closings of schools and restaurants (**AQ 2.2**) in a **multi-step progression** of the topic. The stories belong to a real-world concern that is not visible in the topic extraction. Hence, the user needs to bring domain knowledge for the confirmation of that pattern because the shared characteristics are only implicit (**AQ 2.3**). A multi-step progression of the same words rising consecutively is not possible in our data, since the selection of cosine similarity makes unlikely high-relevance words triggering change points by becoming even more frequent.

8. Discussion

Aside discussing results and limitations, we highlight implications of our approach supporting experts to understand their methods and how other users could benefit as well.

Analysis Questions and Change Patterns We gave examples that our analysis questions unveiled in-depth insights into the detected change points, although a larger user evaluation and tests with different datasets are still lacking to validate to which extent these analysis questions can be answered through our approach. We have covered all analysis questions except for **AQ 1.3**, which we could not observe. Our statisticians attributed this to the intricacies of LDA as a topic extraction technique. We believe that, if we have moving concerns between topics, it would be visible in our visualization system as a **synchronous change** pattern. Our focus on the analysis of detected change points (**AQ 2**), in general, revealed rich opportunities for analysis and even led us to the categorization of method-agnostic change patterns. The fact that these patterns are latent in the data, but the algorithm is not designed to explicitly detect these, informs the need for further developments in this direction. Once a method for the automated identification of these patterns is available, its results could be well integrated into our approach. The current point-in-time interpretation of changes could be extended to changes that span a time interval.

Generalizability While our implementation is somewhat tied to the specific outputs of a (rolling) LDA topic extraction model, many of our contributions are more broadly applicable to other methods as well. The high-level analysis questions and change patterns we have identified are independent of the used method for change point detection. Furthermore, the comparative word clouds and timeline visualizations can be ported to other methods. The temporal correlation matrices within and between topics, as well as the detailed inspection of detected changes, could be adjusted to display the same or similar information for other models with a few tweaks. The list of the most impacting words at a given time step can be extended to intervals of time to account for approaches that detect gradual changes rather than abrupt ones.

Usage Scenarios Change point detection is typically done online (i.e., on data streaming at the time of the analysis), whereas we investigated an offline scenario to allow experts an in-depth retrospective analysis of their method and the data. Nonetheless, our results are also relevant to users interested in a text collection itself (e.g., literary scholars). Instead of understanding in detail the topic extraction and change detection method, we would then assume a sufficient quality of the outputs and reduce the complexity of the approach. We could focus on domain-specific aspects like the inclusion of multiple text sources to compare the topics across those boundaries. Another important step would be integrating the original text documents, selecting representative documents for specific topics and detected changes. Complementing this, we could also make the topic and change point detection steerable to users, similar as described by El-Assady et al. [ESD*19].

9. Conclusion

We proposed a visualization approach to support the comprehension of change points in temporal text collections. We defined the abstract problem space along analysis questions to help examine the role of change points in dynamic topic extraction, and discussed our concrete application scenario in this context. We developed a visualization approach to analyze and connect output artifacts of the analysis methods, and demonstrated the usefulness of the approach through an in-depth analysis of a temporal text collection on COVID-19-related news data. Our analysis generated insights on the derived topics and their contents over time. We classified eight high-level change patterns. Future opportunities lie in the automated detection of such high-level patterns and their more explicit visual presentation to a wider base of users.

Acknowledgements

This work is partly funded by *Mercator Research Center Ruhr* (*MERCUR*; project: "Vergleichende Analyse dynamischer Netzwerkstrukturen im Zusammenspiel statistischer und visueller Methoden").

References

[BLB*14] BURCH, MICHAEL, LOHMANN, STEFFEN, BECK, FABIAN, et al. "RadCloud: Visualizing multiple texts with merged word clouds". *Proceedings of the International Conference on Information Visualisation* (2014), 108–113. DOI: 10.1109/IV.2014.72 2.

[BNJ03] BLEI, DAVID M, NG, ANDREW Y, and JORDAN, MICHAEL I. "Latent Dirichlet Allocation". *Journal of Machine Learning Research* 3 (2003), 993–1022. URL: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf 3.

[CLT*11] CUI, WEIWEI, LIU, SHIXIA, TAN, LI, et al. "TextFlow: Towards Better Understanding of Evolving Topics in Text". *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), 2412–2421. DOI: 10.1109/tvcg.2011.239 2.

[CLWW14] CUI, WEIWEI, LIU, SHIXIA, WU, ZHUOFENG, and WEI, HAO. "How Hierarchical Topics Evolve in Large Text Corpora". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), 2281–2290. DOI: 10.1109/tvcg.2014.2346433 2.

[DL16] DOU, WENWEN and LIU, SHIXIA. "Topic- and Time-Oriented Visual Text Analysis". *IEEE Computer Graphics and Applications* 36.4 (2016), 8–13. DOI: 10.1109/mcg.2016.73 2.

[DWS*12] DOU, WENWEN, WANG, XIAOYU, SKAU, DREW, et al. "LeadLine: Interactive Visual Analysis of Text Data through Event Identification and Exploration". *IEEE Conference on Visual Analytics Science and Technology (VAST)*. 2012. DOI: 10.1109/vast.2012.6400485 2.

[ESD*19] EL-ASSADY, MENNATALLAH, SPERRLE, FABIAN, DEUSSEN, OLIVER, et al. "Visual Analytics for Topic Model Optimization based on User-Steerable Speculative Execution". *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), 374–384. DOI: 10.1109/tvcg.2018.2864769 8.

[GBM*19] GOBBO, B., BALSAMO, D., MAURI, M., et al. "Topic Tomographies (TopTom): a Visual Approach to Distill Information from Media Streams". *Computer Graphics Forum* 38.3 (2019), 609–621. DOI: 10.1111/cgf.13714 2.

[GJG*15] GAD, SAMAH, JAVED, WAQAS, GHANI, SOHAIB, et al. "ThemeDelta: Dynamic Segmentations over Temporal Topic Models". *IEEE Transactions on Visualization and Computer Graphics* 21.5 (2015), 672–685. DOI: 10.1109/tvcg.2014.2388208 2, 5.

[HHWN02] HAVRE, S., HETZLER, E., WHITNEY, P., and NOWELL, L. "ThemeRiver: Visualizing Thematic Changes in Large Document Collections". *IEEE Transactions on Visualization and Computer Graphics* 8.1 (2002), 9–20. DOI: 10.1109/2945.981848 2.

[KBK11] KRSTAJIC, M., BERTINI, E., and KEIM, D. "CloudLines: Compact Display of Event Episodes in Multiple Time-Series". *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), 2432–2439. DOI: 10.1109/tvcg.2011.179 2.

[KKE21] KNITTEL, JOHANNES, KOCH, STEFFEN, and ERTL, THOMAS. "PyramidTags: Context-, Time- and Word Order-Aware Tag Maps to Explore Large Document Collections". *IEEE Transactions on Visualization and Computer Graphics* 27.12 (2021), 4455–4468. DOI: 10.1109/tvcg.2020.3010095 2.

[LWLM18] LU, YAFENG, WANG, HONG, LANDIS, STEVEN, and MACIEJEWSKI, ROSS. "A Visual Analytics Framework for Identifying Topic Drivers in Media Events". *IEEE Transactions on Visualization and Computer Graphics* 24.9 (2018), 2501–2515. DOI: 10.1109/tvcg.2017.2752166 2.

[LYK*12] LUO, DONGNING, YANG, JING, KRSTAJIC, MILOS, et al. "EventRiver: Visually Exploring Text Collections with Temporal References". *IEEE Transactions on Visualization and Computer Graphics* 18 (2012), 93–105. DOI: 10.1109/TVCG.2010.225 2.

[LYW*16] LIU, SHIXIA, YIN, JIALUN, WANG, XITING, et al. "Online Visual Analytics of Text Streams". *IEEE Transactions on Visualization and Computer Graphics* 22.11 (2016), 2451–2466. DOI: 10.1109/tvcg.2015.2509990 2.

[LZP*12] LIU, SHIXIA, ZHOU, MICHELLE X., PAN, SHIMEI, et al. "TIARA: Interactive, Topic-Based Visual Text Summarization and Analysis". *ACM Transactions on Intelligent Systems and Technology* 3.2 (2012), 1–28. DOI: 10.1145/2089094.2089101 2.

[MBB*11] MARCUS, ADAM, BERNSTEIN, MICHAEL S, BADAR, OSAMA, et al. "TwitInfo: Aggregating and visualizing microblogs for event exploration". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, 227–236. DOI: 10.1145/1978942.1978975 2.

[PCR*21] PASQUALI, ARIAN, CAMPOS, RICARDO, RIBEIRO, ALEXANDRE, et al. "TLS-Covid19: a new annotated corpus for timeline summarization". *European Conference on Information Retrieval*. Springer. 2021, 497–512. DOI: 10.1007/978-3-030-72113-8_33 3.

[RJR21] RIEGER, JONAS, JENTSCH, CARSTEN, and RAHNENFÜHRER, JÖRG. "RollingLDA: An Update Algorithm of Latent Dirichlet Allocation to Construct Consistent Time Series from Textual Data". *Findings of the Association for Computational Linguistics: EMNLP 2021*. 2021, 2337–2347. DOI: 10.18653/v1/2021.findings-emnlp.201 2, 3.

[RLFJ22] RIEGER, JONAS, LANGE, KAI-ROBIN, FLOSSDORF, JONATHAN, and JENTSCH, CARSTEN. "Dynamic change detection in topics based on rolling LDAs". *Proceedings of the Text2Story'22 Workshop*. Vol. 3117. CEUR-WS. 2022, 5–13. URL: http://ceur-ws.org/Vol-3117/ 2, 4, 5.

[SWL*14] SUN, GUODAO, WU, YINGCAI, LIU, SHIXIA, et al. "EvoRiver: Visual Analysis of Topic Coopetition on Social Media". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), 1753–1762. DOI: 10.1109/tvcg.2014.2346919 2.

[XWW*13] XU, PANPAN, WU, YINGCAI, WEI, ENXUN, et al. "Visual analysis of topic competition on social media". *IEEE Transactions on Visualization and Computer Graphics* 19 (2013), 2012–2021. DOI: 10.1109/TVCG.2013.221 2.

[ZLZ17] ZHANG, CHANGHONG, LI, ZEYU, and ZHANG, JIAWAN. "A Survey on Visualization for Scientific Literature Topics". *Journal of Visualization* 21.2 (2017), 321–335. DOI: 10.1007/s12650-017-0462-2 2.