ContextWing: Pair-wise Visual Comparison for Evolving Sequential Patterns of Contexts in Social Media Data Streams

YUHENG ZHAO, Fudan University, China
XINYU WANG, Fudan University, China
CHEN GUO, James Madison University, USA
MIN LU, Shenzhen University, China
SIMING CHEN*, School of Data Science, Fudan University, China

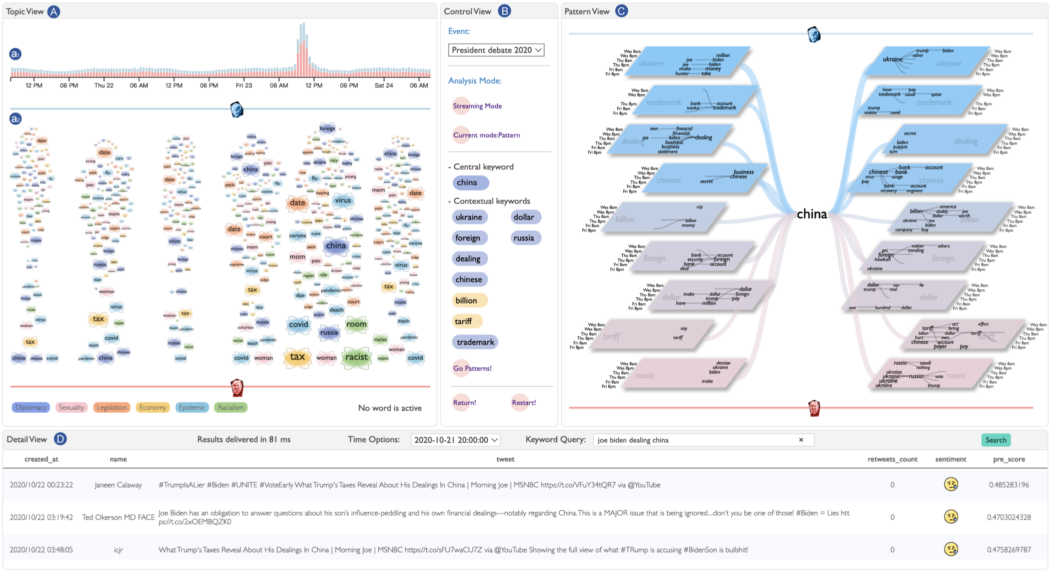

Fig. 1. The interface of ContextWing. (A) The topic view shows the number of pair-wise tweets and dynamic wordles, which allows users to explore multi-granularity topics and streaming topics. (B) The control view enables users to change the analysis mode and observe the selected topics and contextual words to iteratively generate sequential patterns. (C) The pattern view visualizes sequential patterns of contexts using a novel wing-metaphor design. (D) The detail view provides detailed information of original tweets.

*Siming Chen is the corresponding author.

Authors' addresses: Yuheng Zhao, yuhengzhao_cn@163.com, Fudan University, Shanghai, China; Xinyu Wang, chongmingbao@126.com, Fudan University, Shanghai, China; Chen Guo, guo4cx@jmu.edu, James Madison University, USA; Min Lu, lumin.vis@gmail.com, Shenzhen University, Shenzhen, China; Siming Chen, simingchen@fudan.edu.cn, School of Data Science, Fudan University, Shanghai, China.

2573-0142/2023/4-ART40 $15.00
https://doi.org/10.1145/3579473

Understanding and comparing the evolution of public opinions on a social media event is important. However, such a task requires summarizing rich semantic information and an in-depth comparison of semantics and dynamics at the same time, which is difficult for the analysis. To tackle these challenges, we propose ContextWing, an interactive visual analytics system to support pair-wise comparison for evolving sequential patterns of contexts between two data streams. The computational model of ContextWing generates dynamic topics and sequential patterns, and characterizes public attention and pair-wise correlations. A novel multi-layer bilateral wing metaphor is designed to intuitively visualizes sequential patterns merged by different contexts to reveal the similarities and differences in both temporal and semantic aspects between two streams. Interactive tools support the selection of a central keyword and its contexts to iteratively generate patterns for a focused exploration. The system supports analysis on both static and streaming settings that enables a wider range of application scenarios. We verify the effectiveness and usability of ContextWing from multiple facets, including three case studies, two expert interviews, and a user study.

CCS Concepts: • **Human-centered computing → Visualization**.

Additional Key Words and Phrases: Social Media; Visual Analytics; Visual Comparison; Pair-wise Analysis

ACM Reference Format:
Yuheng Zhao, Xinyu Wang, Chen Guo, Min Lu, and Siming Chen. 2023. ContextWing: Pair-wise Visual Comparison for Evolving Sequential Patterns of Contexts in Social Media Data Streams. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 40 (April 2023), 31 pages. https://doi.org/10.1145/3579473

1 INTRODUCTION

With the rapid growth of social media, many people enjoy posting messages to express their opinions and concepts and spread big news, which emerge as data streams. A collection of tweets containing the same keyword forms a social media data stream. To facilitate quick understanding of the massive social media data for social science researchers and public opinion analysts, it is important to provide a summary of opinions embedded in social media messages. A visual summary of these tweets may enable users quickly understand these textual data. Word cloud is a common approach to provide visual summaries of textual data [36, 65]. However, word cloud provides limited contextual information and does not provide connections between keywords to convey a sentence's meaning. Therefore, we extract a sequence of keywords that appear sequentially in a sentence as the summary for a tweet. Meanwhile, since many tweets include the same sequence, we define such sequences as a *pattern*. For example, during a US presidential debate period, there were many discussions on the debate time, such as *"The Final debate is scheduled to start at 9pm ET on Thursday"* and *"The U.S final presidential debate will start on Thursday"*, etc. People have different expressions, but they all mention the same sequences: *"final-debate-start-Thursday"*, which is a pattern. Such patterns are very diverse, and we need to compare the similarities and differences between them to understand the public opinions. Moreover, as these patterns belong to different time periods, we also need to compare the patterns from both temporal and semantic aspects. In addition, to help analyze public attitudes, comparing relationships of patterns and different data streams is needed. To deal with these complex analysis, visualization techniques can be used to support the comparison.

The visual comparison of text is a broad research topic [14]. First, it is difficult to combine the comparison of semantics and dynamics in sequences analysis. SentenTree [35] uses a tree-style structure that tackles the challenge of sequences comparison, helping people quickly understand essential concepts and ideas. However, such approaches are limited to static text sequence data and fail to support temporal comparison. SparkClouds [43] supports the comparison of temporal trends between multiple tag clouds, but it cannot support sequence comparison due to the lack of connection between keywords. Therefore, it is hard to visualize the temporal and semantic comparison of sequences at the same time. Second, it is challenging to compare semantics and dynamics in different data streams. Co-Bridges [12] addresses the challenge of pair-wise visual comparison of multi-items between two data streams, but still, it cannot be applied to sequences to show more contexts and connections to aid understanding. Third, in addition to historical social media data, the analysis for real-time data is more challenging but important for the real world. The difficulty is that it requires fast modeling methods and dynamic visualizations to reveal the characteristics in a short memory. Whisper [9] visualizes the information diffusion process on both static and streaming settings, but it was focused on geolocation data and does not support pair-wise comparison for sequences. Overall, there is a lack of a visual technique that supports the pair-wise comparison of temporal and semantic sequence patterns in two data streams at the same time and lack of generalization on streaming setting.

To address the above challenges, we propose a system named ContextWing (Figure 1) that combines a pattern generation model and a novel visual design (Figure 2), a multi-bilateral wing structure that connects patterns with the same central words and merges patterns with same contextual keywords for more clearly show its semantic differences and similarities. Patterns are arranged vertically corresponding to the time index on each layer. Keywords in a pattern are connected in a syntactic order horizontally on either side of the central keyword. The visual design enables users to pair-wise visual comparisons of evolving patterns of contexts at the same time, which overcomes the word clouds and word trees' limitations.

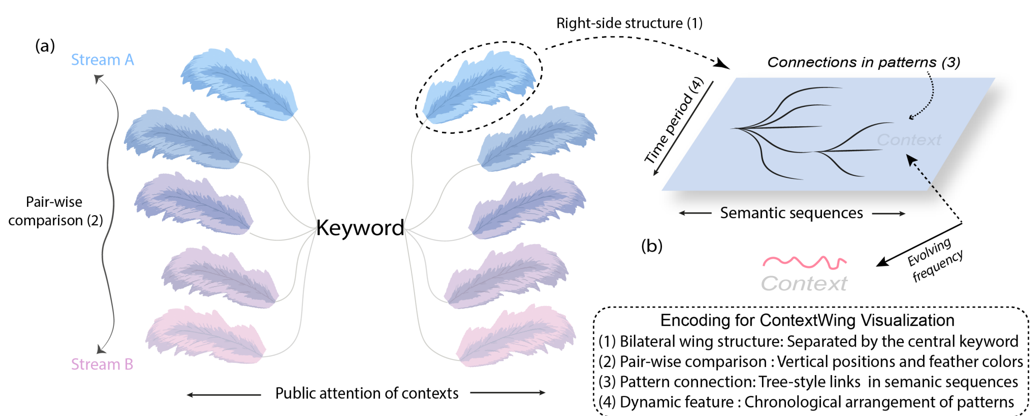

Fig. 2. Design rationale of ContextWing. (a) A multi-bilateral wing structure that connects a central keyword and its contextual feathers. (b) On each pair of the context feathers, patterns are aligned from the left to the right of the central keyword and connected by tree-style links.

To further support the exploration process, we provide interactive techniques (Figure 1B) to support the selection of keywords from the topic view (Figure 1A) to generate wings in the pattern wing view iteratively (Figure 1C). The exploration and comparison of wing structures with the detailed text information in (Figure 1D) contribute to the summary of opinions and concepts of a keyword. Through the iterative analysis of multiple wings, users can comprehend a whole event.

The research contributions of this work include:

- **A novel visual metaphor for pair-wise comparison.** The wing-metaphor design visualizes evolving sequential patterns of contexts, which supports in-depth pair-wise comparison of sequential patterns of contexts in both temporal and semantic aspects.
- **A visual analytics system to explore patterns of contexts both static and streaming settings.** To the best of our knowledge, our visual analytics system is the first one addressing

pair-wise visual comparison for patterns of contexts from social media data streams. It supports flexible analysis of topics, contexts and patterns and detailed exploration.

- **An evaluation of ContextWing.** We verified our system through two case studies, two interviews with two domain experts from the journalism and political science domain, and a formal user study with both subjective and objective evaluation. The evaluation study confirmed the effectiveness of the system.

2 RELATED WORK

We review and summarize the previous work related to social media text visualization, including static text visualization, dynamic text visualization, and visual comparison analysis.

2.1 Time-invariant Text Visualization

As a kind of unstructured data, texts contain massive amounts of information. In analyzing static visualization, word clouds, word trees, tag clouds are the most widely used methods to visualize keywords in popular discussions that help people quickly know the text content [8, 31, 62]. However, these methods lack visualization for semantic contexts of keywords. Some scholars have researched such semantic analysis problems and found that semantically clustered word clouds and word trees can improve the understanding of large document collections [55, 61]. Hearst et al. [33]also evaluated the effectiveness of semantically grouped word cloud designs. However, a semantically aggregated word cloud is not enough to understand a large number of short social media texts due to the lack of semantic associations between words in tweets. SentenTree [35] tackles this problem by using a word tree to display frequent sentence patterns of social media texts. The context semantic sequence allows an intuitive understanding of the key concepts and opinions in an extensive social media text collection. However, word trees are unsuitable for analyzing time-varying text data [43]. Since word trees, similar to word clouds, do not explicitly represent trends or support comparisons, there is a high cognitive load for people to perceive trends in these multiple structures. Our design adopts a wing design with an interactive timeline, which is positioned in a more layout-effective manner to visualize the temporal sequences.

2.2 Offline Visual Analytics for Time-varying Data

There have been various designs to visualize temporal text data [18, 21, 23, 24, 39]. Many scholars use the metaphor of rivers to visualize evolving topics [17, 19, 47, 48, 63]. For example, ThemeRiver [49] is the first system to automatically create a stacked stream graph layout that effectively visualizes topics in multiple time series. RoseRiver [20] adopted a tree-cut algorithm to explore hierarchical evolutionary topics. Some of them focused on dynamic relationships between different topics [64]. For example, Xu et al. [65] showed the dynamic competition to gain public attention among events and the roles leaders play in the process. However, these methods are limited in visualizing detail semantics. Nevertheless, Parallel tag cloud proves that the details-in-context display is helpful for understanding time-varying text data [15]. However, the words in each period are arranged alphabetically, which does not support comparison on dynamic contextual sequences. PyramidTags [37] proposed a design that supported the analysis of temporal evolution and semantic relations of keywords. However, there are no borders between unrelated tags and no connections between related tags, so it is difficult to capture the exact order of words and relations between different sequences. To address these challenges, we propose a wing structure that can visualize topics with more detailed contexts and connections to promote understanding.

2.3 Online Visual Analytics for Time-varying Data

The visual analysis of streaming text data is challenging since the data needs to be processed with effective and fast methods to support incremental updates and help semantic understanding, and meanwhile dynamic visualizations need to be able to visually reveal the characteristics of the data in the user's short-term memory.

Many researchers have previously investigated how to construct real-time visual analysis systems for social media [14]. One representative approach is combining the stream graph to visualize the topics dynamically. For example, Dork et al. introduced one of the first online visual analysis systems that combined a stream graph to capture the dynamic topics [22]. Liu et al. [48] proposed a sedimentation-based stream visualization to help understand hierarchical topic evolution in high-volume text streams. Twitcident [1] is developed for filtering, searching and analyzing facet information for streaming data such as entity recognition (NER), the classified content type, etc. The map-like visualizations are also proposed to display grouped data. STREAMIT [4] dynamically projects documents on 2D domain based on vector similarity with the force-directed layout, which enables users to explore streaming documents. TwitterScope [26] uses with a map metaphor to visualize clustered documents into "countries" with keyword summaries. Knittel et al. [38] propose a bag-of-words-based dynamic spherical k-Means++ algorithm that implements dynamic clustering of topics and a visualization interface that visualize topics and tweets, which can inspire us in terms of computational methods. However, we not only provide streaming information, but also explore and design new methods to explore and compare the temporal and semantic contexts of streaming data, which has not been solved before.

In addition, supporting static and streaming analysis is more challenging and few works lead to it. Whisper [9] verifies the usefulness of supporting both modes, but it is proposed to visualize information diffusion of geolocalized tweets. However, it is difficult to collect massive tweets with geolocation, so we focus more on the analysis of tweet content to offer topics analysis, contextual patterns and pair-wise comparison with two modes for more generalized social media analysis scenarios.

2.4 Visual Comparison of Text Data

Comparison is one of the most important analytical tasks in visualization techniques and data analysis processes. Gleicher et al. [29] suggested three primitive information visualization for comparison: juxtaposition (or separation), superposition (or overlay), and an explicit representation of the relationships (i.e., explicit encoding). Sequence Surveyor [2] used juxtaposition to compare aligned genomic sequences. Each row represents the sequence of genes of an organism. Sometimes these three approaches are hybridized. For example, Topographic BGPlay [16] used explicit encoding and superposition to visualize ISP prefix data. In terms of the data source type of the comparison tasks, previous researchers conducted a series of studies on comparison tasks on social media data. Specifically, they adopted various comparative methods to analyze multiple data streams. (1) Geographical and temporal information [46]: TravelDiff [41] proposed a visual analytics system to investigate travel trajectories from microblog messages. E-Map [13] used a map-style graphic design to analyze multi-faceted data on a social media event. (2) Textual information [15]: Juxtaposition visualization is primarily used to visualize and compare text data. For example, Embedding Comparator [6] presented a method of embedding spaces. BarcodeTree [45] visualized comparisons of topological structures and node attribute values of multiple trees. Bremm et al. [7] introduced a visual design for comparing multiple hierarchical structures. A tool named Word Storms [10], visualized corpora of documents with multiple word clouds, which made similar documents easier to compare.

These previous methods facilitate the comparative analysis. However, most previous work focuses on multi-object comparison tasks [3, 27]. These methods are not effective for us to specifically analyze pair-wise relationships in social media data streams. Recently, several visualization techniques were proposed to visualize pair-wise comparisons. For example, Duet [42] is a visual analysis system designed to facilitate novices to understand pair-wise comparisons. Co-bridges proposed a pair-wise comparison workflow to support visual analytics on multi-stream data [12]. However, they could not support the comparison of dynamic contextual sequences.

Different from previous work, we propose a new pair-wise comparison design using a wing metaphor, which can intuitively demonstrate the dynamic contextual sequences in two streams at the same time, and supports the analysis on both static and streaming settings.

3 OVERVIEW

In this section, we introduce the research challenges of social media data analysis through expert interviews and discuss the analytical tasks.

3.1 Expert Interview

Social media data analysis is a multidisciplinary research method that can be combined with traditional methods to address many research problems in social science [25]. Therefore, we aim to design a general-purpose visual analysis tool that benefits users from different research fields with demands in social media analysis.

Participants. To better characterize the application problems, we conducted semi-structured interviews with three experts (EA, EB, EC) from different social science research fields. In particular, EA and EB are university professors who majored in public policy and political science. Both of them have over ten years of professional experience. EA and EB are concerned about the combination of social media and their research, but without effective tools to support in-depth exploration in the research process, and it can only be used to understand social media through texts. The last time EA analyzed social media events was three months ago to learn about how people commented on pandemic policies. EB conducted a research project about the political participation of citizens on social media six months ago. EC is a senior researcher in media and communication with more than five years of research experience, and also has work experience in news editing in an authoritative news media organization. EC also has several social media data analysis experiences. The most recent analysis experience was six months ago, to analyze the distribution of topics posted by Korean users on Twitter, with the main purpose of text classification and user interest portrait.

Interview Procedure. Each interview lasted about an hour, with open-ended questions from three aspects. First, to identify their expertise in social media analysis, we asked experts about their previous experience, such as *"Have you analyzed social media data in your research?*; *"When was the last time, and what was the study about?"*. Second, experts' proficiency with visual analytics tools for social media may create different visualization needs. Thus, we asked them *"Have you used social media analysis systems or visualization tools?"*; *"What are the weaknesses in the visualization tools you use?"*. Third, to explore the important dimensions of social media and domain research together, we asked them *"What aspects of social media data are you most concerned about?"*; *"What are your suggestions for integrating social media visual analysis tools into your research?"* and so on. We also asked the experts follow-up questions according to their answers. For example, if the expert has limited relevant experience, we will ask *"Have ever you paid attention to related social media analysis work in your field?"*. We took notes and recorded videos throughout the interviews.

Through analyzing the notes and interview transcripts, we summarized their main challenges which were validated and refined for several rounds as follows.

C1: Summarizing the opinions in context. Due to the complexity and large amount of data, it is difficult to gather massive social media data and analyze its core topic in detail. EA said that *"there are a large number of users on social media, especially the highly educated and politically active users. Thus, the combination of sample survey and interactive analysis would be a better choice"*. EC believed that *"with increasing tweets on social media, a combination of quantitative and qualitative semantic analysis is needed for news' topic selection and news reporting in journalism research"*. Although they tried some tools, such as an online word cloud generator, to get some summary, the contexts and topic information are very limited. Therefore, they are looking forward to using more powerful visual analysis tools.

C2: Understanding the time-varying opinions. Given the changing nature of social media opinions over time, experts naturally point out the need to analyze the evolution of topics. EC mentioned that *"the temporal analysis could lead us to know which topic dominates an event and changes in the context of a topic may reveal the development of the event."* Since there are complicated and abundant topics in an event, and the time-varying opinions of the different topics are not identical and evolving as well, which brought many difficulties to their research. According to the traditional survey method, they can only sample for a small amount of coverage or social media tweets to summarize, which is prone to sampling bias and not enough to analyze the evolution.

C3: Finding out the public tendency in two data streams. An essential challenge for studying social media data is to get a picture of the public tendencies and thus assess whether the event/policy is justified or not. EB studies the political selections and he mentioned that *"for many political events there is a game between two parties, such as 'Trump' and 'Biden' in the presidential debate"*. Comparing the change of public tendencies of them contributes to the study of *"how social media is changing the pattern of election communication and dissemination in the U.S"*. EA mentioned a related research direction [54] she is interested in, *"comparing public attitudes on social media to reformulate the policy or highlight certain aspects of the policy"*. EC also believed that *"comparative analysis methods are important in opinion analysis, but such comparative relationships needed to be quantified"*, and *"it would be better if sentiment analysis assisted"*.

C4: Analyzing real-time data. In addition to studying historical data, real-time analysis is also important to help track a new event. EA commented that *"traditional questionnaires can obtain the attitudes at the time, but cannot effectively reflect changes in public opinion in real-time"*. And being aware of updates is difficult without the help of a visual analytics system. EC pointed out that *"for journalists, the streaming topic analysis is strongly required for news topic selection and public opinion survey, especially for the timely reporting of controversial topics, so as to avoid the partial intensification of public opinion"*. This view is also confirmed by Kate et al. [59]. However, dynamic analysis involves complex computational methods and visual design, which is challenging to implement the detailed analysis.

The feedback from three experts suggests that a visual analytics system of social media text data is necessary to empower analysts in social sciences-related research to summarize discussions of an event with multi-facet information quickly. To design this system, we combined their suggestions with a user-centered design process [53] and invited them to evaluate the system.

To elaborate on the problem, we define the following terms to describe our target data.

- **Keyword** refers to the single word in each tweet, tagged with a time moment and a syntactic position
- **Topic** is a cluster of similar words, and each topic can be represented by single or multiple keywords.
- **Data stream** is filtered by an entity, denoted by the keyword.

- **Key entity** is the focus of the event (e.g., in the presidential debate, "Trump" and "Biden" are two key entities, and in the Brexit event, the two attitudes "Leave" and "Remain" are key entities).
- **Contexts** are keywords in the same tweet and are contexts for each other, also called contextual keywords.
- **Pattern** refers to a sequence of keywords retaining the words' order in a tweet and multiple tweets usually shared.
- **Public attention** refers to the influence of tweets containing different keywords measured by counting the number of retweets.
- **Semantics** is a generic expression that contains four levels of semantic information, including keywords, topics, contexts and patterns.

3.2 Analytical Tasks

By discussing with the experts for several rounds, we compiled a list of analytical tasks that our system should support.

T1: Semantics extraction. Investigating the main discussions of an event requires a wealth of semantics. Thus, different levels of semantic information, including keywords, topics, contexts and patterns should be extracted from a large corpus for quickly understanding an event. (C1)

T2: Temporal analysis of semantics. The most important dimension of social media events is time, and it is necessary to understand their development rules to form a complete cognition of events. Thus, analyzing the temporal distribution of topics and patterns is strongly required. (C2)

T3: Comparative analysis of semantics. Comparing the similarities or differences between the extracted semantics can yield richer insights. The derived semantics can be conducted with qualitative comparison (common or distinct) and quantitative comparison (frequencies, public attention, etc.). For example, patterns that belong to different contexts may be similar or different, which indicates if people are discussing similar content. (C1, C2)

T4: Pair-wise comparison. To pair-wisely comparing extracted semantics with different key entities at different times can lead us to understand the relations between public and key entities and their evolution, which is fundamental for many scenarios. This complex analysis goal requires quantifying the relationship of semantics with the key entities. The sentiment analysis can be applied to understand public attitudes further. (C1, C2, C3)

T5: Analysis on the streaming setting. The analysis of historical data helps summarize the rules and experience for research. However, in many cases, the real-time updating events need to be analyzed according to the above tasks. Therefore, we need to adjust the method of extracting semantics and the corresponding visual design. (C4)

4 MODELING

In this section, we will describe ContextWing's computational methods and how to apply them in both static and Streaming modes. The results of the measurements enable us to create and visualize a comprehensive view of the evolving wordles and patterns.

4.1 Data Pre-process

We first tokenized the original tweets, lemmatized the words and deleted the stopwords of the tweets collected from Twitter. For each tweet, we applied a BERT-based sentiment model [11] which was trained by social media data to predict the sentiment of the tweet. The prediction result will be positive or negative with a sentiment score. In the following, each topic, contextual keyword and pattern with two different emotions are counted separately to aid in understanding semantic information at different granularity.

4.2 Analysis on Static Setting

This section will introduce how to model data in static mode, including topic extraction, pair-wise correlation, public attention, pattern generation, etc. These methods can be easily applied to streaming settings.

Topic Extraction. In the static setting, since the data of social media events are topic specific, we use Word2Vec [52] to directly cluster the topic keywords by calculating the similarity. As a result, we can obtain high dimension vector for each word. Words with high spatial similarity indicate that they are likely to be related to the same topic. Then we apply K-Means[40] to cluster word vectors to distinguish different topics. As it is historical data, we can combine prior knowledge to select topics and make the clustering effect more aligned with the experts' expectations. Since each cluster often has a large number of words, we retain top-N words as keywords for visualization. For extraction of contextual words of each keyword, we obtain its top-N words with high co-occurrence frequency to support contextual exploration. Considering the clear topic view layout in visualization, N is generally chosen from 20 to 30.

Pair-wise Correlation. In each event, there must be key entities that are the focus of discussions and greatly influence the trend of public opinions. We quantify the correlation of keywords between the two data streams based on their co-occurrence, i.e., "correlation degree", denoted as C_i^t and $C_i^t \in [0, 1]$.

$$C_i^t = \frac{Rank\left(\beta_{iA}^t - \beta_{iB}^t\right)}{N^t}, \quad \forall i \in W^t \tag{1}$$

In equation 1, β_{ij}^t indicates the co-occurrence frequency of word i and j at time t, A and B respectively represents data stream A and stream B; $rank$ is a function that returns the rank of word i's co-occurrence frequency; N^t represents the total number of words at time t, W^t represents the collection of all the words at time t. If $C_i^t \rightarrow 1$, the word i at time t is more related to stream A.

Public Attention. The most straightforward way to quantify the relationships between a central keyword and its patterns of contexts is to calculate its co-occurrence frequency in social texts. However, there are some limitations to the information presented based on co-occurrence frequency simply. Therefore, based on the co-occurrence frequency, we propose public attention to characterize such distance (Equation 2). The basic idea of public attention is derived from the popularity that it gets retweeted, i.e., retweets number.

$$\mathscr{A}(c,k) = Log\left[\frac{\sum_{i=1}^{n} u_i(c,k)\eta_i \cdot \sum_{i=1}^{n} u_i(c,k)r_i \;/ \sum_{i=1}^{n} u_i(c,k)}{\sum_{i=1}^{n} u_i(c,-k)\eta_i \cdot \sum_{i=1}^{n} u_i(c,-k)r_i \;/ \sum_{i=1}^{n} u_i(c,-k)}\right] \tag{2}$$

Here k is a selected central keyword , c represents the a selected contextual keyword of the central keyword k; n represent the total number of tweets in the data set; $u_i(c,k)$ is an inclusion condition that represents whether the i-th tweet contains c and k, while $u_i(c,-k)$ is whether the i-th tweet contains c but without k; η_i represents whether the i-th tweet is retweeted, r_i represents the number of the i-th tweet being retweeted. Both the numerator and denominator reflect empirical estimates of the retweet number under the inclusion condition. This method can help characterize the distance between the central word and each selected contextual word. If $\mathscr{A}(c,k)$is positive, it means that the closer their relationship is, the higher the public attention is; if negative, it means that the less close their relationship is, the lower the public attention is.

Pattern Generation. The last part of our model is pattern generation, which is proposed to adapt our analytical tasks, shown in Algorithm 1.

We first choose a central keyword with a contextual word (denoted as *"w"*) to form two initial patterns with different orders, which are both two-tuples, like *"w-keyword"* or *"keyword-w"*. We

Algorithm 1: Pattern Generation

```
Data: a central keyword and a selected contextual keyword to find surrounding patterns,
      with all data in a period t
Result: patterns for this central keyword and selected contextual keyword
leaf sequential patterns = empty list;
Function FindPatternAsTree(pattern):
    node = pattern;
    if find iteration <= pattern length then
        next word is determined by sorting algorithm;
        if node has valid right leaf then
            update node right leaf as FindPattern (node with next keyword);
            FindPatternAsTree (node right leaf);
        end
        if node has valid left leaf then
            update node left leaf as FindPattern (next keyword with node);
            FindPatternAsTree (node left leaf);
        end
        update interation;
    end
    push pattern to patterns list if the pattern contains the central keyword
Function FindPattern(pattern):
    let lines = pop line from all data;
    set a counter dict to do the counting;
    while visible lines > 0 do
        if lines contain pattern then
            new pattern = pattern with next keyword as right counter for new pattern
              counts;
            new pattern = pattern with next keyword as left counter for new pattern counts;
        end
    end
    sort counter by counts belonged to these new patterns;
    return top k patterns;
```

then iterate around each two-tuple to search for the next new keyword to form a new pattern, which is a triplet. Since the processed texts consist of verbs, nouns, and adjectives, the general length of a pattern sequence could be four or adjusted further. In the searching process, a new keyword is determined by a sorting algorithm. We first set a skip value, which means finding the following keyword in the tweet near the parent pattern within the skip value. Since we want to make the process more self-adapting, the skip value sets as $\frac{l}{2}+1$, l is the pattern length. Then we extract all patterns (n is the number of new patterns, P_i represents the $i-th$ new pattern), calculate the frequencies α_{P_i} of them and sort them to get the top k patterns, which allow us to choose the number of patterns of each period we want for visualization. The frequency of the pattern should be a large threshold to ensure the pattern can represent more people's opinions. η_P^t means the top k patterns and their frequencies in period t (Equation 3).

$$\eta_P^t = Sort([\alpha_{P_1}, \alpha_{P_2}, ..., \alpha_{P_n}], k) \tag{3}$$

The number of keywords in the pattern was expanded by an iterative search. When the patterns length l is up to the setting value, the iteration stops. Therefore, a pattern tree is formed for each central keyword and a selected contextual keyword in each period. Since there are many selected contextual keywords, the algorithms iterate multiple times.

4.3 Analysis on Streaming Setting

Compared with the static setting, the streaming modeling is more complex, especially the modeling for topic clustering. Other methods such as correlation and pattern generation, can be migrated to streaming mode with interval configuration. Therefore, in the following, we will focus on our dynamic topic clustering method.

Our topic modeling method is a dynamic clustering approach based on BERTopic [30] and K-Means++ [5]. BERTopic is a topic modeling technique that use transformer-based model to calculate the sentence embeddings, apply UMAP [51] to reduce the high dimensional vectors and clusters them by HDBSCAN [50]. Since BERTopic does not support dynamically clustering in the streaming set, we combined it with K-Means++. K-Means++ is one of the fastest clustering algorithms suitable for stream settings. Finally, the model generates topics by class-based TF-IDF algorithm.

The reason we do not use the Word2Vec-based method on streaming settings includes two aspects. First, the generation of word vectors depends on the corpus, but the vector of the same word in each minute of data will change. Thus, the clustering centers cannot be passed to the next generation, unless we set a large sliding window and consider the whole window as a bag of words [38]. However, this method will bring a time difference with real time. Therefore, we use the transformer-based embedding method, which is based on a pre-trained model that can produce the same word vector in each minute. Second, the Word2Vec-based method we use needs initial keywords to extract words with high similarity, but we need to know the topics of an event in advance. Therefore, an automatically clustering method is required to help users know the coming topics. Thus, the dynamic BERTopic-KMeans++ method is better for streaming settings.

The principle of dynamic KMeans is to initialize the centroids by the last clustering result. We use k-means++ to initialize the centroids when data arrive in the first minute (C_t). After the first clustering is completed, we pass the cluster centers to the next minute C_{t+1} to maintain the information from the last step and improve the clustering efficiency. Each clustering will generate up to 6 topics with at most 20 keywords under each topic, considering the user's mental map is limited for real-time changing information.

In order to get more coherent topics, after clustering in each minute, we will merge each cluster ($C_{t+1}^k, k = 0, ..., 5$) into the clusters (C_t^k) in the previous minute, if the 25% keywords in C_{t+1}^k exist in C_t^k. The top 2 keywords will be taken as the topic name of C_{t+1}^k. The merged keywords list is used for merging with C_{t+2}^k. If we do not keep previous keywords, the previous clustering results will be lost, and it will be difficult to maintain the association if they appear again later. However, we also save the before-merged top n keywords of the current generation C_{t+1}^k for searching contextual words to generate patterns. Moreover, to improve the coherence of the topic over a longer period, we set up a configurable sliding window (e.g., size = 20 mins) and cluster tweets every minute. After finishing the dynamic clustering in one sliding window, we cluster all tweets in the window to get the centroids of the clustering result and then pass it into the initial minute of the next window.

As for the clustering performance, we evaluated Bertopic-kMeans++ with BoW-kMeans++ [38] on a standard data set (20 Newsgroups data set), which comprises 18,846 posts. We tested the NMI (Normalized Mutual Information) to judge the quality of the resulting clustering according to the class labels. We ran each method five times to calculate the median value and found that the median NMI for Bertopic-kMeans++ is $0.61[0.60-0.62]$, that for BoW-kMeans++ is $0.42[0.39-0.43]$, which shows that Bertopic-kMeans++ is better than BoW methods in terms of the clustering result. As for concerns about its computational efficiency, we tested our dynamic Bertopic-kMeans++ method on the data sets in our case studies, which includes an average of 80 tweets per minute. We found that dynamic Bertopic-kMeans++ can process 1-minute's data within 6-7s. Therefore, for many social media event data sets within around 800 tweets per minute, our method is feasible regarding both the clustering effect and time efficiency.

5 VISUALIZATION DESIGN

In this section, we discuss the design rationales, the visual encoding and the construction process for the wing-metaphor design in pattern view (Figure 1C).

5.1 Design Rationales

Thinking through the analytical challenges and tasks we summarized in Section 3, it is required to propose a new design that can be used for visualizing the evolving sequential patterns of contexts and allow interactions to compare them.

Previous work such as ThemeRiver [49], cannot see the details-in-contexts and the keywords' connection that are important for users to get an in-depth understanding of a topic. Using interactions might add word clouds to the topic flow to show some contexts [65], but it's hard to get a rich overview at once. In addition, the basis of understanding the text is to read the text according to the word order in the text. Sententree [35] visualizes such sequences but does not support dynamic analysis. Co-Bridges [12] supports pair-wise comparison but lacks context information. Therefore, we need to present a new visual design to address the analytical challenges.

In ContextWing, the principle metaphor is the wing and feathers, as shown in Figure 2.

Wing Metaphor: Wing represents hierarchical connections of patterns for a central keyword. Horizontally, the wing is divided into left and right structures. The left side represents the words appearing ahead of the central keyword, and the right side represents the words behind it.

Feather Metaphor: Each horizontally symmetric pair of feathers aggregates patterns with the identical selected contextual keyword. We termed such a pair of feathers as a layer. Feather' color and vertical position represent the pair-wise correlation between two key entities. And the horizontal position represents public attention.

5.2 ContextWing Construction Process

Based on the design rationale, we implement ContextWing to support dynamic patterns comparison by the following processes. To more conveniently understand the application of contexts in design, we use contextual keywords instead of contexts in this section.

Construct layers for selected contextual keywords. We assign each selected contextual keyword to a layer with the same length, and the width can be automatically fine-tuned according to the number of selected words. The vertical positions and the color encoding of the layers indicate the pair-wise comparison. We can use the method discussed in Section 4.2 to calculate the pair-wise correlation, use the color of the layers to encode the absolute pair-wise correlation, and use the vertical position to indicate relative closeness. As Figure 1 shows, the more pink and lower the layer is, the closer to "Trump", and the more blue and upper the layer, the closer to Biden. The horizontal placement of the layers indicates the public attention of the central keyword and its selected contextual keywords. If the layer is closer to the central keyword horizontally, it means the larger attention they have. Besides, the width of the links to the layer indicates the total frequency of the patterns on that layer.

Layout contextual keywords on each layer. As Figure 3a shows, we place words appearing sequentially before the central keyword on a left layer and the rest of the words on a right layer in order and arrange the patterns in chronological order from the top to the bottom of the layer. We align the contextual keywords vertically with the central keyword. Keywords in the same horizontal line with the central keyword form a pattern. The time ticks on the side of layers show the corresponding time period of the patterns in the same line. The keywords' size encodes the pattern's frequency after containing this word. The last keyword frequency, therefore, represents the pattern frequency.

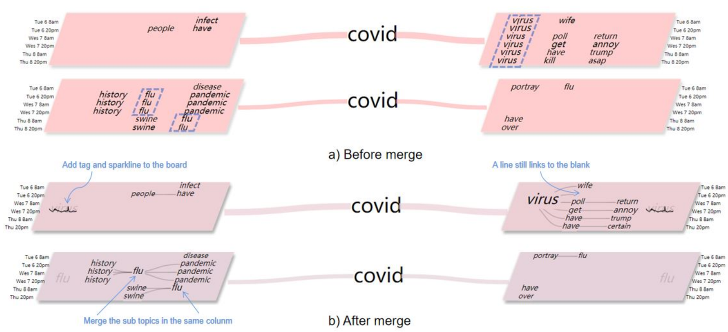

Fig. 3. (a) The visualization before merging the same contextual word. The first contextual word is "virus" while the second is "flu". (b) It's the visualization after merging the contextual words. We can find that there is only one "virus" left on the right layer. As for "flu", the words in the same column are merged.

Merge the selected contextual keywords. In the placement process, we find there are many repeated keywords in the same column, which is not easy to compare the common and distinct semantics. For example, the selected contextual keywords like "flu" are not obviously displayed since "pandemic" also occurs repeatedly and is closer to the central keyword (Figure 3a). Therefore, we need to avoid influence from other contextual keywords and emphasize the selected contextual keyword of the layer. As Figure 3b shows, we merge such keywords in the same column, which maintains the overall structure and avoids misunderstanding. After merging, the information of the frequency evolution of the contextual keyword will lose. So we also add a spark-line to visualize the evolving frequency to make up for the info loss.

Add tree-style links. Employing the idea of the tree structure, we add lines to link the words of the same pattern for better understanding. However, as Figure 3a shows, there exists the situation that the contextual keyword is the final word of a pattern, like "virus" on the second horizontal line. If we merge the word, the place of that word may become blank, and it will look like no words on the right layer at that time, which is a misunderstanding. So we add a line to link the contextual keyword and the next blank position on the horizontal line to express the existence of the contextual keyword as Figure 3b shows.

5.3 Design Alternatives

In the creative process of ContextWing, we also have tried two alternative designs to explore the optimal design that satisfies the analysis goals. (Figure 4)

One alternative is to extract some patterns for each time period and aggregate them into a tree by the central keyword (Figure 4a). Although each tree corresponds to a time, if we want to compare the evolution of a pattern containing a specific keyword at different times, we need to go through each tree to check which pattern is the first. If the time period increases (the number of trees increases), the above process becomes more tedious and the arrangement of multiple trees of the same central word is redundant for the layout. In addition, the design cannot visually show the evolution in two or even more streams of information due to the limitations of encoding. Overall, static sequential tree comparison cannot support our analysis tasks.

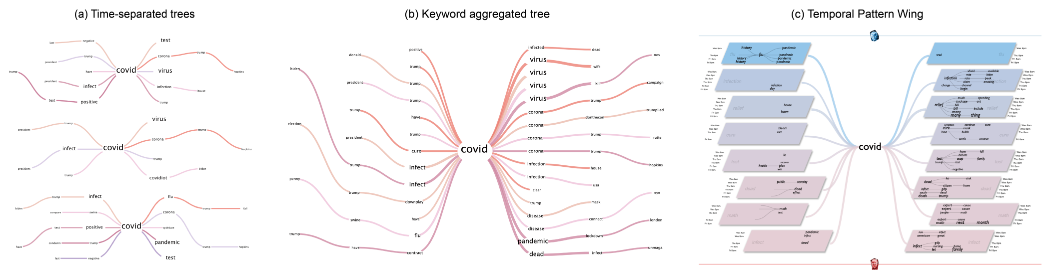

Fig. 4. Two design alternatives and our final design of pattern view: (a) time-separated trees. (b) A keyword aggregated tree (c) A temporal patterns wing.

The second design combines multiple count trees by central terms, saving layout (Figure 4b). It also aggregates by contextual keywords, facilitating the classification of observed patterns, and the vertical position of clusters can encode the relative relevance in two streams. However, it is not easy to make temporal comparisons while keeping keyword aggregation. If each pattern corresponds to a point in time, we can label each pattern with a time tag; but if a pattern appears multiple times, it takes a lot of time to observe the temporal features, which is not an intuitive visual approach.

Therefore, we propose the final design to reserve the most frequent pattern for each time period (Figure 4c). We merge patterns containing a specific keyword if they appear in multiple time periods and on the same layer, so as to compare intuitively according to the aggregation method. Therefore, the design solves both temporal and semantic comparisons of time-varying semantic sequences and supports pair-wise comparisons of two data streams.

6 SYSTEM OVERVIEW

We present a system leveraging the wing design for dynamic semantic information comparison. The system includes the topic view, control view, pattern view, and detail view. Figure 5 shows the system architecture which consists of data processing and analysis workflow.

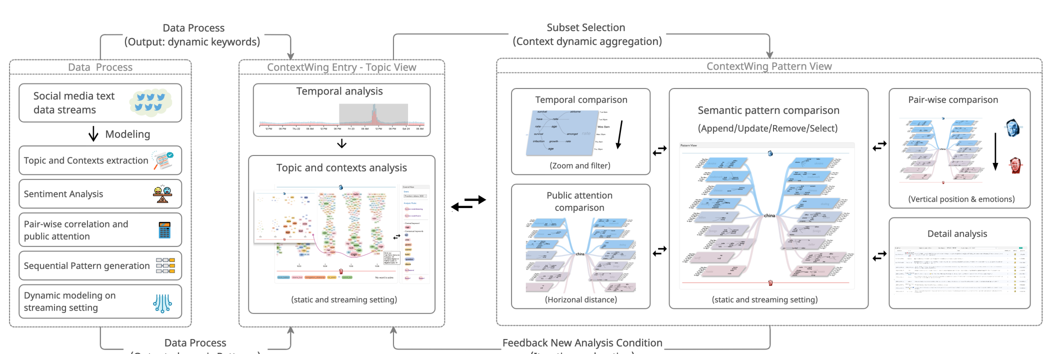

Fig. 5. Overview of the system architecture. The analytics workflow starts with the topic view, supporting iteratively visual analysis on both static and streaming mode.

6.1 Topic View

We provide a topic view to extract semantics and allow users to select keywords as inputs for the pattern view. As Figure 6 shows, the top view is a histogram showing the evolving percentage of tweets of two data streams. Two symbols of streams are placed at the top and bottom of the view respectively. The topics are labeled on buttons at the bottom left with different colors. In the bubble chart, keywords are aggregated and divided into several periods. As each keyword can actually generate a pattern wing structure, we design the keyword bubble as a wing-like glyph. The size and opacity indicate the frequency of the word, and the color indicates the topic it belongs. The vertical position of the bubble represents its correlation with the two key actors. Some important metrics, such as frequency and sentiment distribution, are shown in the tooltip. In order to visually observe the coherence of the topic, we added connecting lines for the keywords frequently appearing in different stages. Users can hover over keywords to observe the frequencies and correlation degrees. The histogram can be brushed to select the time period, and the data in the selected period will be re-aggregated and displayed in several pools.

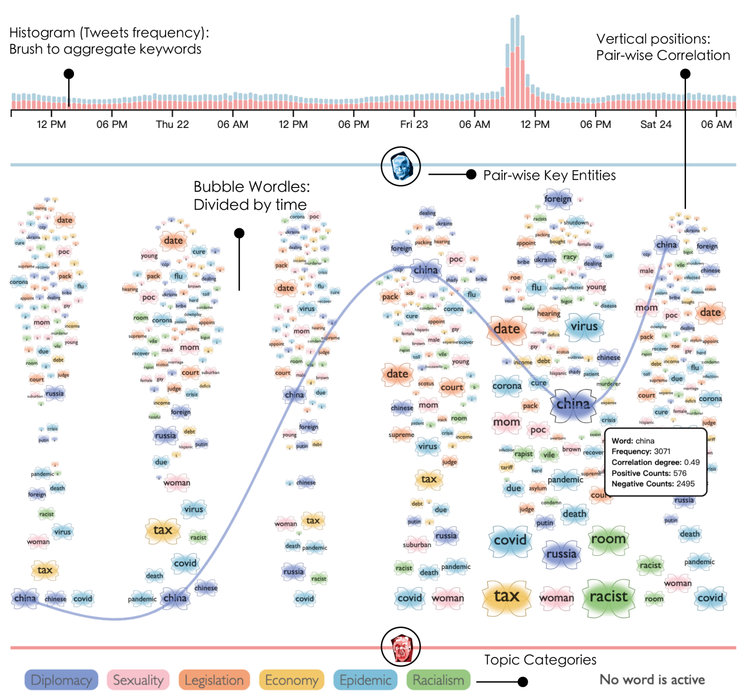

Fig. 6. The overview of the dynamically aggregated keywords for "The US 2020 presidential debate", showing the evolution of classified keywords under different topics.

The design of topic view can also be extended to the streaming setting. The histogram, bubble pools, and topic buttons are updated synchronously at preset intervals (e.g., 1-min). According to the modeling result, the old topic will be replaced and highlighted with new colors if a new topic appears. The color and name of the topic button always correspond to the category of the updated bubble, which can help users perceive the dynamic changes of the topic more intuitively. In dynamic changes, it is difficult for users to keep a mental map of previous information. Therefore, we combine the histogram and bubbles chart, which can help users review real-time historical data. Users can also click "Halt" buttons to pause/continue updates.

6.2 Control View

We set up data set options and analysis modes, and users can choose to switch between static and streaming analysis modes. Moreover, starting from the topic view (Figure 6), and there are two ways to explore a central keyword and its contextual keywords. Users can click "Change Mode" to switch on the iteration control panel (Figure 1B): (1) Explore mode: Users can constantly drill down into the contextual keywords of a keyword by clicking. (2) Pattern mode: Users can click a bubble as the central keyword and then select its contextual keywords. To keep the information coherent, the color of the selected keyword in the control view still indicates its topic. Then, clicking "Go Pattern" can observe the derived patterns in the right pattern view. For example, Figure 1 shows the selection operation under pattern mode, which supports the keywords selection for the central keyword "china". In this process, users can re-click bubbles to update the selection and click "Return" and "Restart" to the previous or initial state to have an iterative exploration.

6.3 Pattern View

Once a wing structure is constructed, users can make comparisons from different aspects. We provide the following interactive ways for detailed comparisons. First, to support the temporal comparison of the patterns, users can hover any keywords and the corresponding patterns will be highlighted while other patterns will be hidden. Thus, users can better observe a single pattern with a time tick on the layer. Second, to fulfill the requirement to compare patterns from the perspective of selected contextual keywords. Users can click any time ticks to highlight patterns of different layers in this period. Besides, a sparkline denoting the evolving frequency of the selected contextual keyword in the whole period will show up upon users' hover on the side of the layer. We also provide a tooltip that show the frequency and emotion distribution of each pattern. Pattern view also supports real-time updates, showing patterns at the same time intervals as in the topic view, arranged vertically to correspond to the few moments from the current moment.

6.4 Detail View

To assist users in understanding patterns, we provide a detail view (Figure 1D), which can show the information of original tweets such as time and sentiment score. In the pattern view, users can select a pattern and the original tweets will be displayed in detail view. Moreover, users can select a time period and type words they are interested in.

7 CASE STUDY

To access the generality of the ContextWing and the approach of comparison techniques, we applied it to two datasets collected from Twitter: "The 2016 Brexit" and "The 2020 US Presidential Debate". Moreover, we verified the effectiveness of the real-time analysis on "The 2020 US Presidential Debate" dataset. We tested whether ContextWing can enable us to obtain interesting findings and accomplish tasks (T1-T5) through visual analytics.

7.1 The 2016 Brexit

In this case, we applied ContextWing to compare the public opinions of "Leave" vs. "Remain" in "The 2016 Brexit" event to verify tasks T1-T4. We collected data related to "Brexit" on June 21, 2016, before the "Brexit" vote on June 23. Our dataset mainly describes the following five topics: *Debate, Economic, Immigrant, Labour* and *Politics*. Text streams, including "Leave" are encoded in blue and "Remain" is in pink. Through the initial exploration of "Immigrant" in the topic view, we found many keywords are related to “Leave”, then we clicked “migrant” and selected several contextual keywords to generate patterns. Figure 7 illustrates the comparison of patterns for “migrant”.

Exploration of different stages. Following the timeline, we found that the patterns of “calais” such as *“migrant-calais-attempt-reach”* arose from 4 AM to 12 AM, while *“hundred-migrant-storm-calais”* appeared at the same period, showing that hundreds of migrants stormed Calais in an attempt to reach the UK (Figure 7-1). Then we found from 12 AM to 8 PM, some people said *“migrant-block-motorway-violent”* and *“hundred-migrant-calais-violent”* indicating that people discussed hundreds of migrants blocked the motorway in a violent (Figure 7-2). The increasing font size also indicated that more people participated in the discussion after 12 AM. Then from 8 PM to 24 PM, many people mentioned *“hundred-migrant-closed-calais”* after the heated discussions of “violent” (Figure 7-3). Therefore, the timeline visualization exemplifies how the discussion on the migrant riots in the French border city of Calais evolved.

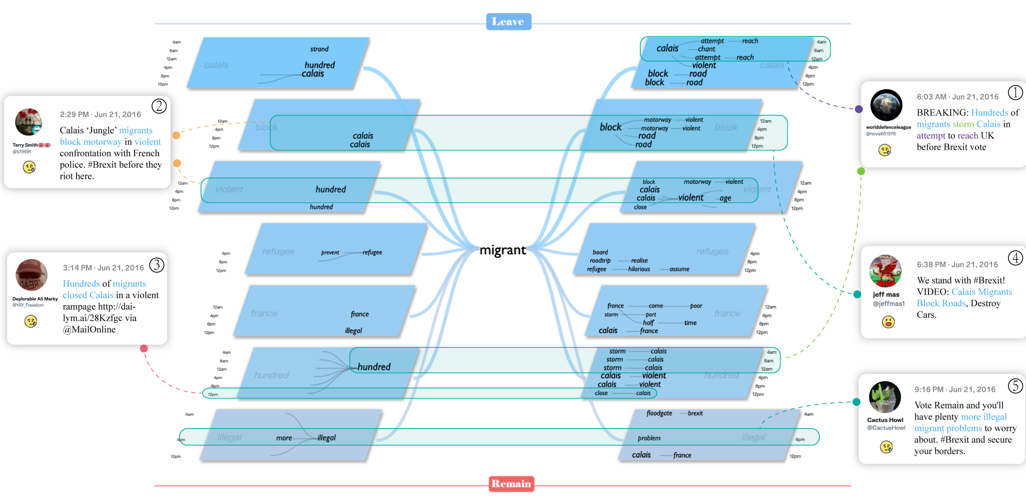

Fig. 7. Exploration of “The 2016 Brexit”, visualizing the patterns of “migrant”. We found the evolution of discussions about Calais migrants’ violence in three stages, which was the reason that people support “Leave”.

Pair-wise comparison of “Leave” vs “Remain”. As for the pair-wise correlations in two streams, we found that patterns of “migrant” were more related to “Leave”, and many people discussed *“more-illegal-migrant-problem”* with negative emotion tendency, which indicated that people were very concerned about border security and tended to vote leave (Figure 7-5). The pair-wise comparison analysis answers our initial question about the propensity of “immigration” and “leave” to be related.

Public attention comparison. When people talked “migrant” with “refugee”, “france” and “hundred”, the tweets were more likely to be retweeted, which shows that people were more concerned about the refugees of migrants, the immigration at the French border and the number of migrants during the whole period.

In this case, we analyzed the evolution of Calais immigration by comparing the temporal features of patterns. We also analyzed the public’s preference for “Leave” and “Remain” by pair-wise comparison. Moreover, the analysis of public attention shows that people pay significantly different attention to different contexts.

7.2 The 2020 US Presidential Debate

We collected 347,282 tweets from October 21 to October 24 in 2020, which cover the debate period on Thursday, October 22, 9:00 PM–10:30 PM. This case aims to verify the effectiveness of ContextWing in tasks T1-T4 to gain more comprehensive insights with multi-wings.

With ContextWing, we found some interesting stories by comparing the tweets related to two data streams containing "Donald Trump" and "Joe Biden" (Figure 1). From the overview, we found that "China" was mentioned most (3,205) under the discussions of the topic "diplomacy" (Figure 6, so we are curious about what details people usually said when they mentioned "China" in two streams. ContextWing provides an overview of the discussions of Trump and Biden under "China" (Figure 1). In the left topic view, after clicking "China", we found there were several keywords with relative high frequency, such as "ukraine", "foreign", "trademark", "tariff", "dollar", and "russia", etc, were mentioned between Biden and Trump (Figure 1A). Therefore, we selected them to further derive patterns in the right pattern view (Figure 1C).

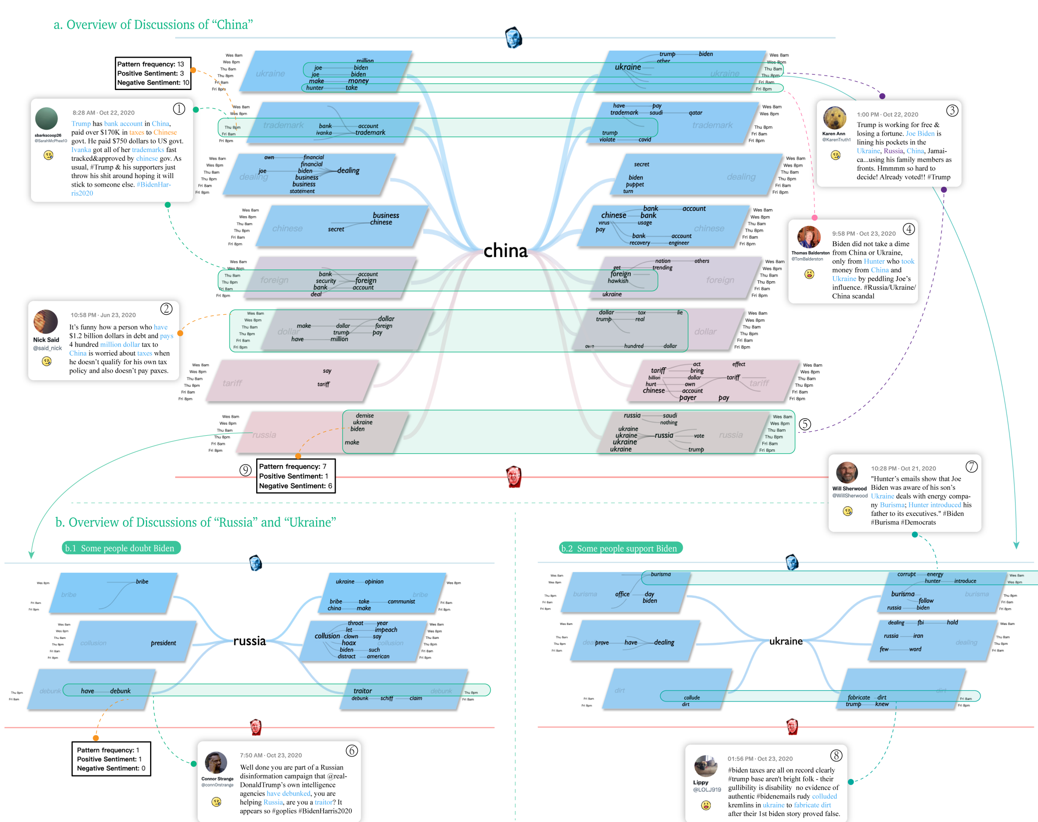

Fig. 8. Comparative analysis of the tweets related to Joe Biden and Donald Trump in the 2020 US Presidential Debate. We analyzed the discussions of "China", indicating that people discussed the presidential candidate's personal finances (a). We then iteratively explored and compared the discussions of "russia" (b.1) and "ukraine" (b.2) and found both negative and positive comments about the candidate's finance issues. Thus, some interesting stories can be detected by comparison of multi-wings.

From the data streams related to Trump (Figure 8a-1), such as "foreign" and "trademark", we found some similar patterns such as *"bank-account-china-foreign"*, *"bank-account-china-trademark"* and *"ivanka-trademark-china-trump"* showed up on Thursday, which indicated that in this period many people discussed a bank account and the trademark in China related to Trump and his daughter Ivanka with negative emotions. Some people did not vote him and were more in favor of Biden ("trademark" was more related to Biden). Besides, by exploring patterns of "dollar" on Wednesday 8 AM and Thursday 8 PM (Figure 8a-2), we found people mentioned *"trump-pay-china-dollar"* and *"china-dollar-tax-lie"*, showing that people discussed Trump paid dollars to china related to taxes. Thus, we drilled down to explore the reason. From Thursday 8 AM to Friday 8 AM, we found people discussed Trump had millions of dollars in a china bank account and paid hundreds of millions of dollars in taxes to China, but he may not qualify for his tax policy to pay taxes. Meanwhile, by comparing the public attention of each contextual keyword, we found "dollar" was the closest keyword at the horizontal level to "china", showing that tweets containing "china" and "dollar" attract more public attention and people concerned the number of dollars Trump had related to China.

In the data stream related to Biden, both patterns of "ukraine" and "russia" showed "biden" in the same period (Figure 8a-3). We first focused on patterns of "ukraine". From Thursday 8 AM to Friday 8 PM, *"joe-biden-china-ukraine"* and *"hunter-take-china-ukraine"* showed up, which suggested that people discussed Biden and his son (Hunter) might make money from China, Ukraine and Russia, and tagged this event as "Russia/Ukraine/China scandal" (Figure 8a-4). Meanwhile, on the other side of the wing, we found that "ukraine" showed up with "russia" when the debate started (Figure 8a-5). Before Wednesday 8 PM, there were frequent references to *"demise-China-Russia-Saudi (Arabia)"* that were not intuitively related to Ukraine. However, after the debate, *"Biden-China-Ukraine-Russia"*, *"China-Ukraine-Russia-vote"* and *"China-Ukraine-Russia-trump"*appeared continuously in the following periods. We also found that ‘russia" was more related to Trump but more people showed negative comments and they were angry with Biden's Ukraine scandal and turned to vote for Trump (Figure 8a-3,9).

To have a deeper understanding of public opinions on Biden's scandal, we further compared the wings of "ukraine" and "russia". In the pattern view of "russia", we saw that people were mainly discussing such scandal issues, and people would like to comment on Biden as a "traitor" rather than standing up for him (Figure 8 b.1-6). However, from patterns of "burisma" on "ukraine", we found that Joe Biden and his son Hunter was described in *"burisma-ukraine-corrupt-energy"* *"ukraine-burisma-hunter-introduce"*, showing that Biden might link with corrupt activities related to energy company Burisma in Ukraine (Figure 8 b.2-7). However, on layer "dirt", some people expressed positive tendencies to stand up for him by illustrating the scandal was dirt fabricated by Trump instead of doubting him (Figure 8 b.2-8).

In this case, we comprehensively analyzed the two data streams containing Biden and Trump regarding semantic features of dynamic patterns. The iterative exploration and comparison of multi-wings enable the formation of a story by understanding multiple public opinions and tendencies for the two data streams.

7.3 Analysis on Streaming Setting

In this section, we applied ContextWing on the streaming setting with "The presidential debate" data set. We focus on the period from 2020-10-23 10:01 AM - 11:00 AM, after the formal debate time (2020-10-22 09:00 PM). We chose the 1-minute interval to update data. Our analysis goal is to summarize the dynamic topic evolution (T1, T2, T3) and compare topics in two data streams (T4) under the streaming setting (T5).

Streaming visual analytics can provide us high flexibility to explore tiny but important features in a short time, which are easily lost in long-term analysis. For example, we found that before 10:03 AM, the topic "wage-minimum" was discussed most frequently, with a frequency of 135 at 10:02 AM and the percentage of negative sentiment was 74% (100/135) at 10:02 AM (Figure 9a.1). However, the keyword "cage" (belonging to "Obama-cage") was heating up from 10:04 AM, and the sentiment analysis showed that 87.6% tweets were negative (197/225) (Figure 9a.2), which was more negative than "wage-minimum". Therefore, we clicked "cage" and further observed its contextual keywords (Figure 9b). We found that "cage" frequently appears with words like "build", "child" and was very close to Trump, while "Obama", "answer" were closer to Biden. These findings gave us an initial understanding that this was a negative topic related to both Trump and Biden.

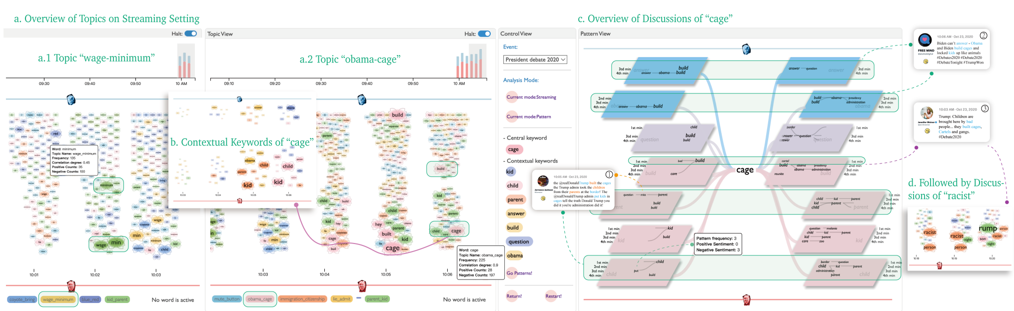

Fig. 9. An example that investigated the real-time changes in the public discussions on "cage": (a) The analysis of topic view indicated that "cage" started to appear after topic "wage-minimum" (a1), and gain significant attention from 10:03 AM (a2); (b-c) The analysis of contextual keywords and semantic patterns showed that the highlight of discussions was "Who build the cage for the child? Trump or Biden-Obama's administration?" Many people condemned Trump, while others believed that Biden-Obama built it (c-1, c-2); (d) Since Trump's description of this topic involved racism, which caused the public to criticize it (c-3) and the topic of racism heated up later.

To obtain more in-depth understanding, we chose some relatively high-frequency contextual keywords to generate patterns. From 10:03AM to 10:07AM, people usually discussed "Trump" with "child" such as *"put-child-cage-border"*, *"build-cage-child-parent"*, from which we can roughly understand that trump was related to building cages for children at the border leading to their separation from their parents (Figure 9c-1). The specific tweets in the detail view confirm this. On the other hand, we also found that there were references to patterns such as *"answer-obama-build-cage, cage-build-obama-administration"*, which may indicate that people were discussing whether Obama administration built it, and we found it close to Biden and with a strong negative sentiment (Figure 9c-2). By comparing the discussions of Trump and Biden, we knew that people were discussing "Who built the cages for child?".

The benefits of real-time analytics are that we can keep track of new topics as they arise and the connections between topics. By reading the pattern (*"bad-people-build-cages-cartel"*) and related tweets, we found that, in previous day's debate, Trump's explanation of building cages triggered a racist discussion and taken a lot of criticism. From the topic view, we observed "racist" became an explosive controversial topic from 10:15 AM.

Through this case, we can not only understand the updates of the topic, but also trace the relationship between the topics and key entities to understand the evolution of an event.

8 EVALUATION

This section introduces a user study and an expert study to evaluate ContextWing. To the best of our knowledge, there are no similar systems for comparison. Some visualization designs for text such as ThemeRiver [32] based on a river metaphor, enables users to observe the changes of topic categories, but it does not support the user to see the contexts to the topic and analyze the pair-wise relationship. Therefore, it cannot achieve the analysis purpose of an in-depth understanding of the evolution of topics in pair-wise social media data streams. At the same time, we found that compared with our baseline designs (Figure 4) was unfair as well. Since the baseline designs failed to clearly display time, topic and pair-wise relationship at the same time, users are unable to make more informative summaries quickly. Therefore, instead of carrying out an unfair comparison, we evaluate ContextWing on its own.

8.1 User Study Set-up

The introduction of user study including participants and apparatus, questionnaire, procedure and results analysis.

Participants and apparatus. We recruited 18 students (13 male and 5 female, age 19−27, $\mu = 21.88$, $\sigma = 1.94$), denoted as P1-P18. Each experiment was conducted on their 13-inch monitor, with a keyboard and a mouse.

Questionnaire. Each participant was asked to answer 15 objective questions in Table 1 and 14 subjective questions shown in Figure 11.

- Objective questions (O1-O15) investigate users' understanding of the system functions and the analysis results, which are derived from five analytical tasks (T1-T5) described in Section 3. The questions are proposed based on the presidential debate data set. Each objective question is accompanied by a screenshot of the system, and participants need to answer a single- or multiple-choice question.
- Subjective questions (S1-S14) investigate users' comprehensive evaluation of the system, which cover four aspects including intelligibility, analysis functionality, design effectiveness and usability. These aspects are chosen based on suggestions of Rossi et al. [57]. The questions are 5-point Likert scale (1 for strongly disagree and 5 for strongly agree).

Procedure. The study was composed of three sessions, beginning with a 15 minutes tutorial session, in which the analytical tasks and the usage of ContextWing were introduced to the participants. Participants can follow the experimenter to use the system, familiarize themselves with system functions and freely explore the system. When performing the tasks, participants were free to ask questions and were encouraged to think aloud. The formal assessment was conducted using a questionnaire. Each participant was first asked to answer objective questions, followed by subjective questions. A post-study interview was conducted to collect more detailed feedback from the participants. The whole session lasted about 40 minutes for each user.

8.2 Results and Analysis

The average time to complete the questionnaire was around 24 minutes ($\mu = 15.27$, $\sigma = 5.59$). Overall, the average accuracy for objective evaluation reached 89.56% ($\sigma = 0.09$) and the average score of subjective evaluation is around 4.55 ($\sigma = 0.34$). The detailed results of the questionnaire are summarized in Figure 10 and Figure 11.

8.2.1 Objective Evaluation. The analysis of user objective answers helps us to reflect on the analytical tasks in Section 3. Figure 10 shows that all users achieved relatively high accuracies

Table 1. Objective questions correspond to five analytical tasks in Section 3, denoted as O1 to O15.

	Semantics extraction	**Temporal analysis of semantics**	**Comparative analysis of semantics**	**Pair-wise comparison**	**Analysis on streaming setting**
Objective questions	O1: Which keyword is discussed most under the theme of "Epidemic" in the topic view?	O4: In which time period does this pattern start to appear?	O7: Which layer has patterns other layers don't in this pattern view?	O10: Which keyword is most associated with Trump in this topic view?	O13: What is the sentiment tendency of keywords related to Biden at 8:19 AM in this real-time topic view?
	O2: Which pattern is discussed most in this pattern view?	O5: Which keyword shows a decreasing trend in the selected time period in the topic view?	O8: Which layers have the same time distribution?	O11: Which layer is most associated with Biden in this pattern view?	O14: Which topic is newly added at 8:08 AM in this real-time topic view?
	O3: Do you agree with understanding the patterns in this pattern view?	O6: Which pattern is most discussed in this period?	O9: Which layer has the most public attention?	O12: Are the patterns related to Trump more positive or negative in this pattern view?	O15: Do you agree with the below conclusions about "laptop" in real-time pattern view?

(>0.7). Among the five analytical tasks, T1 (Semantics extraction) had the highest accuracy, 96.08%, followed by T4 (Pair-wise comparison, 92.37%).

Semantics extraction (T1). Among the three questions, the first two questions (O1, O2) achieved 100% accuracy, and only the third question (O3) obtained a lower accuracy (μ = 88.24%, σ = 0.32, with two users (P8 and P14) answering incorrectly. In the feedback of P8 and P14, they suggested that they did not remember the meaning of some functions, such as how to generate patterns by selecting contextual keywords. After we re-explained the functions to them, they gave a correct answer to O3. Overall, the results indicate that the participants were generally able to understand the design of patterns connection.

Temporal analysis of semantics (T2). O5 shows that the topic view is very effective for temporal comparisons at the keyword level (100%). O4 and O6 show the accuracy of temporal comparison at the pattern level. The accuracy of O4 is μ = 88.24%, σ = 0.32 (wrong answers from P8 and P14) and that of O6 is μ = 82.35%, σ = 0.38 (wrong answers from P5, P8 and P14). Apart from P8 and P14, one wrong answer is from P5, who misunderstood the order of time. Overall, the result shows that the majority of people were able to understand the interaction design of temporal comparison and answer correctly.

Comparative analysis of semantics (T3). We set questions to cover three aspects of the proposed task T3: semantic similarity of patterns (O7), semantic and temporal relevance (O8) and public attention (O9). The accuracy of the above questions both are larger than 72%. O8 and O9 show higher performance (82.35%). Among the five errors in O7, two were careless mistakes (according to interviews of P4 and P9), three were because the aggregation principle of the pattern is not fully understood (P10, P14 and P17). P14 disagreed that ContextWing was easy to use, commenting that *"I cannot fully remember the visual encoding in pattern view"*. Overall, the result confirmed that aggregating patterns by selected contextual keywords help to compare semantics.

Pair-wise comparison (T4). There are three questions related to this task (O10-O12) and they both have an accuracy above 88%. The high accuracies of O10 and O11 show that ContextWing enables users to compare relationships of semantics in two data streams. Meanwhile, Participants' responses to O12 can prove that sentiment analysis can help to understand public attitudes towards two key entities in the evolution of a whole event.

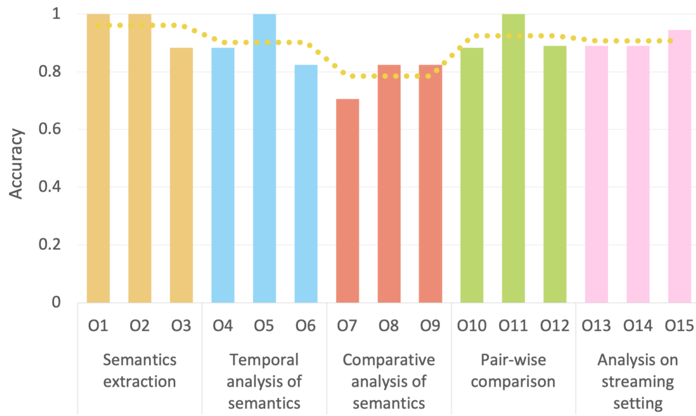

Fig. 10. The results of objective evaluation. The histogram shows accuracy for each question and the line chart shows the averaged accuracy for five analysis tasks.

Analysis on streaming setting (T5). To further verify the effectiveness of dynamic topic analysis, contextual patterns and pair-wise analysis in the streaming setting, we set three questions (O13-O15). The accuracies of the three questions are around 90%, which shows that most users can find the newly coming topics by our real-time topic (O14) view and they can understand the public tendency by understanding and comparing patterns in pair-wise data streams with sentiment analysis result (O13, O15).

8.2.2 Subjective Evaluation. Overall, ContextWing receives high scores in four assessment metrics, in which the analysis functionality is most outstanding and design effectiveness has also been widely recognized. Details are featured in Figure 11.

Intelligibility. To verify the effectiveness of different levels of semantics comparison, four questions were designed. S1 is to evaluate the comprehensibility of the pattern design, and the results show that for S1 ($\mu = 4.44$, $\sigma = 0.62$), 17 people suggested that the definition of pattern and its visualization design was easy to understand. S2 investigated the system for multiple sub-event summarizing of pattern, which was unanimously agreed ($\mu = 4.16$, $\sigma = 0.92$). We also set up S3 to verify if the system supports iterative exploration, which received a higher score than above two ($\mu = 4.5$, $\sigma = 0.78$). S4 shows the highest rating among four questions ($\mu = 4.94$, $\sigma = 0.24$). P2 commented that sentiment analysis was beneficial for analyzing pair-wise opinions and the tendencies of opinions.

Analysis functionality. In general, all participants are neutral, agree or strongly agree that the system can support the analytical tasks proposed in Section 3. Analysis functionality gets the highest score among four aspects. S6 (semantic comparison) has the highest approval rating ($\mu = 4.83$, $\sigma = 0.38$) , which shows that the design of patterns and group patterns by contextual keywords are intuitive and helpful for semantic analysis. Meanwhile, S5 (temporal comparison, $\mu = 4.83$, $\sigma = 0.38$) and S7 ($\mu = 4.67$, $\sigma = 0.59$) shows relatively lower average scores, for which users suggested adding some tips to prevent forgetting function. S8's scores are all above 4, which indicates that users generally recognize the usefulness of the system in real-time analysis.

Visual design effectiveness. The system design is considered effective on all four dimensions, especially in terms of interactivity (S11). Multi-view analysis (S9) and aesthetics (S10) obtained relatively higher scores (S7: $\mu = 4.56$, $\sigma = 0.61$, S8: $\mu = 4.67$, $\sigma = 0.59$). Moreover, for the design of

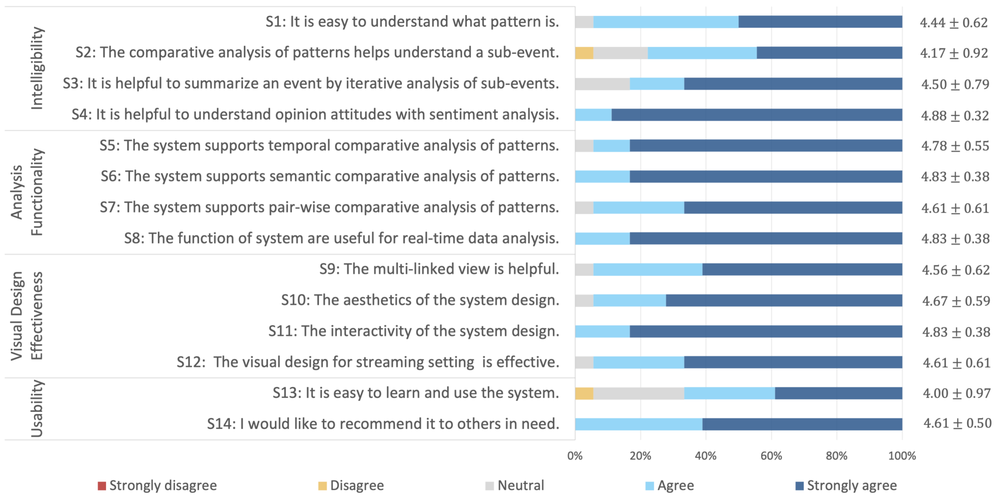

Fig. 11. Description and results of subjective questionnaires. The leftmost column indicates four assessment aspects. The rightmost column denotes Mean ± SD.

streaming analysis (S12), P8 commented that the dynamic animation and contrasting colors are very intuitive to distinguish the updating incremental topics from clustering results.

Usability. Most participants were confident that ContextWing was easy to learn and easy to use (S13) and willing to recommend it to others in need (S14). Meanwhile, some users indicated that it is very useful that the system can support both static and streaming exploration, which is why the system's functions require more exploration time to learn, but they are willing to use it in the future. As an overall comment, P1 said *"This system surprised me because it gave me a rapid and clear understanding of the presidential debates and can find some sub-events such as the scandal of Biden that I did not know before"*.

9 EXPERT STUDY

We conducted an expert study with three experts whom we introduced in Section 3.1 to evaluate the usability and effectiveness of our system.

9.1 Study Process

The interview consists of three sessions. We first introduced our work within 15 minutes and demonstrated the interaction methods and the analysis process within 10 minutes, similar to Case 2 (7.2). Then, in the exploration phase, we invited them to explore the system around 30 minutes, using the think-aloud method by Van et al. [60]. In the exploration phase, EA was interested in "The Brexit" data set (Case 1), while EB and EC were interested in "The 2020 US Presidential Debate" (Case 2 and 3) and were willing to try to use our system under our guidance. We observed how they interacted with ContextWing and collected their feedback with a semi-structured interview lasting around 20-30 minutes.

9.2 Exploration Phase

In the exploration phase, we expected that experts could understand the functions of the system through full exploration, and verify whether the summarized challenges are addressed by our system and help them gain some effective findings, and further discuss how to combine them with their research.

Public policy perspective. We first tested the application of ContextWing on public policy research. For this field, EA was more concerned with the rationality of the policy and the public comments of the policy. EA noticed the immigration problem brought by "Brexit" and understood that people wanted to support "Leave" due to riots and conflicts (C1), which fits the findings in Case 1. EA believed that from this example we can quickly know *"the public attitudes on "Brexit"* (C3) *and the specific and continuous impact* (C2) *on people's lives, which can help the assessing the feasibility of alternative policies"*. For the streaming analysis, EA commented that real-time tracking of topics could reveal *"the situation of the public and the causal mechanism of topics"* (C4), which provides the reference for governments to *"adjust policies in time, and actually is an effective risk communication strategy"*. Traditional questionnaires capture the attitudes at the time and cannot effectively reflect changes in public opinions in real-time.

Political science perspective. For the study of political science, EB was interested in researching how Trump and Biden competed for support in the social media battlefield in previous debate. Through comparing patterns in Trump and Biden under "china", he found how the public criticized Biden and Trump's (e.g., the trademarks in foreign countries with corruption) (C1, C3), and with different attention to them before and after the debate (C2), and these personal business issue influenced people's voting intentions (Case 1). EB concluded that these findings suggest that *"social media messages (even if unverified) are just as likely to stir up opposition*, and that *"parties with more online buzz have more opportunities for direct dialogue with the public* and are therefore more likely to *"stir up mass emotions"*. Therefore, these findings by pair-wise comparison analysis are helpful in studying *"how to use the media wisely to win online battles"*.

Media and communication perspective. As a journalism scholar and journalist, EC has analytical needs for static historical data analysis and is more interested in real-time analysis. After successfully conducting Case 2, EC turned to real-time analysis mode and found that in the initial stage, the public was talking about the topic "wage-minimum", then the topic "cage-kid' 'was coming up with higher frequency (C2). For the patterns of "cage", EC also found people discussed who builds cages for the children (C1, C3) and EC also found. However, "black" "racist" emerged later, and the topic of "cage" still exists, which indicates that "cage" was a focus of this period. By fully exploring the system, EC was interested in using this real-time function to *"select focus topics and survey public attitudes for preparing writing news reports"*(C4), especially for *"the timely objective reporting of controversial topics"*, to *"avoid the partial intensification of public opinion"*.

9.3 Semi-structured Interviews

The results from interviews showed that all experts confirmed that they understood the visual design and analysis methodology of our system.

Visual design. We asked if our visual design and interaction were easy to use and understand, and they were all positive about our design. They acknowledged that the colors were clearly distinguished, and pair-wise entities could be easily associated with topics and patterns. EA suggested that by using color and connecting lines, the topics are coherent over time so that we can fully understand the semantic evolution of the topic. Besides, EB commented that on top of topic visualization, the positions and colors show the topic-entity relevance that fits our analysis goals in many pair-wise political events.

Analysis workflow. We interviewed experts to see if they understood our analysis workflow and the interaction methods. They all confirmed that they understood the workflow and learned to use the system in both static and real-time analysis. EC encouraged that analysis from topics, contexts and patterns can obtain an overview and details of an event and be able to go back to the original text, which prevented spending a lot of time on social media software. They also commented that the system supported multiple interactions, such as the update and removed function, and the return and restart button was very convenient for exploration in a loop.

Insight and inspiration. Encouragingly, all experts agreed that ContextWing deepened their understanding of a social media event and helped to improve their research. For the opinions understanding and evolution analysis, EA commented that *"Previously, we had only to use questionnaire methods to collect public opinions about a policy, which is more targeted but less timely. ContextWing has greatly helped me to gain a broad understanding of the comments, which has inspired me to combine them in future research"*. EB also mentioned that the patterns of different times might be replicated. If the time granularity was proper, comparing patterns could be very efficient. As for pair-wise comparison analysis, EB said that *"The system was very helpful in exploring contrasting political events. By analyzing the topics' evolution and opinions' distribution, we can clearly know how key entities influenced the public and how social media determines the political election"*. For analysis on both static and streaming mode, EC emphasized that *"The system inspired me in two aspects. On the one hand, it can help me select focused or controversial topics and learn about public opinions' evolution to write in-depth reports or lead the right views promptly. On the other hand, since ContextWing concludes quantitative and qualitative results with real-time visual analysis, which can be effectively used to write data news [34]"*. We also asked experts about what other information they would like to know. They all mentioned that if supporting the comparison of more key entities, the system could be applied to a wider range of cases.

10 DISCUSSION AND CONCLUSION

This section introduces our work from multi facets, including implications for design, comparison and generalization, limitations, future work, and conclusion.

10.1 Implications for design

While the evaluations have reflected that our system can address the initial research questions, we also want to highlight what can be learned from this study from a broader perspective.

Pair-wise comparison. In most events, there will be more than one entity or more than one attitude toward an entity. Two entities (attitudes) are common among these events, which is a fundamental research task. Therefore, we provide pair-wise comparison by setting the top and bottom as two different entities and using vertical positions to show semantics-entities relationships, which can help people get an intuitive understanding of which entity a topic is more related to.

Sentiment adds context. People's exact attitudes are worth reporting. Instead of "relating to ", people are more interested in word sentiment, which reveals their positive or negative views of the topic or entity being discussed. So we combine sentiment information in pair-wise comparison and visualize them by adding word sentiment frequency to the tooltips. It shows whether a word or a pattern relates to an entity positively or negatively.

Get the difference in one visualization. Click a button once and get what you want if everyone's willing. So we combine semantic comparison, temporal comparison and pair wise comparison together in the pattern view so that people are able to compare textual differences in three aspects in one visualization.

Summarize content topics. We provide topic classification in the topic view, which goes beyond just providing keywords to be selected as input for the pattern view. But we find those clear topics can help people better understand the events and fast select interesting words. So in the topic view, we visualize topics by giving colors to the wing-like glyph and adding topic selection.

Combine static and real-time analysis. At first, our system only supports static analysis, which mainly provides a complete event visualization. But social media data is real-time and people's attitudes and focus toward an event change over time. These changes are valuable to catch once they appear. But a static system lacks immediacy. Therefore, we create a streaming mode to visualize streaming data and realize real-time analysis.

Original text is necessary. We want to understand a sentence with only four words. The idea is cool that it can save people a lot of time reading long texts. But sometimes, four words are not enough to restore the meaning of the original text. We have to provide a way for people to observe the texts when they need. Therefore, in text analysis, the detail view is necessary. This is also the benefit of visual analysis, where we can directly observe the analysis results from the system and trace them back to the original text.

10.2 Comparison and Generalization

Comparison. Previous visual analytics tools primarily combined rivers to visualize the overall change of evolving topics, and it would reduce much contextual information. Word cloud can be used to present more contexts but without connection in semantics. Our design supports topics analysis, contextual pattern generation and pair-wise comparison. Considering the online time-varying text data visualization, compared with the existing systems like TwitterScope [26], we not only give a topic-level or word-level visualization but also extract patterns with the selected keywords from the original text, which improves people's understanding. Besides, we use Dynamic-Bertopic-Kmeans++ as our background algorithm to visualize real-time evolution, which is fast and effective. The support of static and streaming analysis enables ContextWing to be applied to more generalized scenarios such as research of public participation in politics, news reporting, public policy evaluation, etc.

Generalization. The system may have the possible applications on the below points.

First, social media data can be used in more scenarios. In the study process, we found that professional data is challenging to obtain from experts due to privacy issues, especially in politics. But they confirmed that the public social media data we analyzed could also be helpful. For media and communication experts, Twitter is the universal data source that they research. Therefore, we provide our Twitter data to demonstrate our proposed method, and these data can be applied in more social science analysis tasks.

Second, the event contents to be visualized are not limited to social media data. It can be any time-varying textual data, such as news reports or interviews. The modeling methods can also be expanded to them. The topic view can capture the topics included in the events, and the pattern view can help people quickly understand the event. For these longer data, pattern generation can be built on the extraction of arguments based on natural language models [28], making it better to summarize the text in shorter patterns.

Moreover, in the expert evaluation, they mentioned that the quantitative and qualitative analysis present in the visual analytics system could be well used as the materials for data news generation [44]. There have been studies on report generation [56, 58], but they are limited to combining social media visual analytics to generate data news. The difference is that data journalism needs to support systems oriented to mining selected topics, polling public opinion, and supporting real-time analysis. Our system can provide a possible application for visual analytics in this direction.

10.3 Limitations and Future Work

The evaluation results confirm that our work has accomplished analytical goals. Domain experts appreciate our system and see the opportunity to utilize it in their research and teaching practices. However, there are still several limitations.

The setting of parameters. To further enhance the system's ability to explore and analyze freely, users can set thresholds, such as the number of keywords in the topic view, the division of topics into categories, etc. Moreover, according to our expert study, the pair-wise comparison quality depends on the heterogeneity of the two key labels. If the two key actors' attitudes towards a certain topic are significantly different, the result may be better and more prominent. Therefore, the labels of two data streams can be multiple for selection.

The number of data streams. Our visual analysis tasks are currently focused on the pair-wise comparison. The multiple comparisons are a more challenging research question that involves adjusting the computational model and rearranging the design of semantic sequence patterns intuitively, which is worth further study.

Layout comparison of multi-wings. Our system is capable of comparing multiple wings by iterative exploration, but we can further implement such multi-comparisons in a more unified layout. For example, users can choose various wings to observe and compare in one view. Moreover, some recording functions can also be provided in the iteration process to allow users to mark useful information for a straightforward summary.

Time granularity. The temporal comparison of the sequential patterns may not be very meaningful in some situations. Our expert study shows that if the topic's time granularity is improper, the information extracted from different periods may turn out to be homogeneous. The time comparison is useless since the topics of the dataset do not change much. When we extend to streaming analysis and set the short time interval, we see a lot of small topics that start to emerge so that it was not easy to observe in a large temporal granularity. So for those topics that don't change much from a macro perspective, we can solve them in a smaller temporal granularity. A better solution is allowing users to interactively adjust the time granularity depending on their knowledge, which requires online re-aggregation and computing and is worth studying in the future.

10.4 Conclusion

This work presents a novel visualization technique for pair-wise visual comparison called ContextWing. It can help users get an overview of what people are discussing on social media by in-depth pair-wise comparison of evolving sequential patterns of contexts. The key contributions of our system are a novel wing metaphor for pair-wise comparison of contextual patterns, the visual analytics system supporting both static and streaming modes. Three case studies are conducted to evaluate our system. An in-depth user study and expert study demonstrate the effectiveness of our approach in achieving the proposed analytical tasks. The findings of case studies accord with and further supplement the existing theories. Through the evaluation, we demonstrate the system's ability in to tackle the visualization tasks for visual analysis of evolving sequential patterns.

ACKNOWLEDGEMENT

This work is supported by Natural Science Foundation of China (NSFC No.62202105) and Shanghai Municipal Science and Technology Major Project (2021SHZDZX0103), General Program (No. 21ZR1403300), Sailing Program (No. 21YF1402900) and ZJLab.

REFERENCES

[1] Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, and Ke Tao. 2012. Twitcident: fighting fire with information from social web streams. In *Proceedings of the 21st international conference on world wide web*. 305–308.

[2] Danielle Albers, Colin Dewey, and Michael Gleicher. 2011. Sequence surveyor: Leveraging overview for scalable genomic alignment visualization. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2392–2401.

[3] E. Alexander and M. Gleicher. 2016. Task-Driven Comparison of Topic Models. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 320–329. https://doi.org/10.1109/TVCG.2015.2467618

[4] Jamal Alsakran, Yang Chen, Ye Zhao, Jing Yang, and Dongning Luo. 2011. Streamit: Dynamic visualization and interactive exploration of text streams. In *2011 IEEE Pacific Visualization Symposium*. IEEE, 131–138.

[5] David Arthur and Sergei Vassilvitskii. 2006. *k-means++: The advantages of careful seeding*. Technical Report. Stanford.

[6] Angie Boggust, Brandon Carter, and Arvind Satyanarayan. 2019. Embedding Comparator: Visualizing Differences in Global Structure and Local Neighborhoods via Small Multiples. *CoRR* abs/1912.04853 (2019). arXiv:1912.04853 http://arxiv.org/abs/1912.04853

[7] S. Bremm, T. von Landesberger, M. Heß, T. Schreck, P. Weil, and K. Hamacherk. 2011. Interactive visual comparison of multiple trees. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 31–40. https://doi.org/10.1109/VAST.2011.6102439

[8] M. Burch, S. Lohmann, D. Pompe, and D. Weiskopf. 2013. Prefix Tag Clouds. In *2013 17th International Conference on Information Visualisation*. 45–50. https://doi.org/10.1109/IV.2013.5

[9] Nan Cao, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu. 2012. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE transactions on visualization and computer graphics* 18, 12 (2012), 2649–2658.

[10] Quim Castella and Charles Sutton. 2013. Word Storms: Multiples of Word Clouds for Visual Comparison of Documents. *WWW 2014 - Proceedings of the 23rd International Conference on World Wide Web* (01 2013). https://doi.org/10.1145/2566486.2567977

[11] Rohitash Chandra and Ritij Saini. 2021. Biden vs trump: Modeling US general elections using BERT language model. *IEEE Access* 9 (2021), 128494–128505.

[12] S. Chen, N. Andrienko, G. Andrienko, J. Li, and X. Yuan. 2021. Co-Bridges: Pair-wise Visual Connection and Comparison for Multi-item Data Streams. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1612–1622. https://doi.org/10.1109/TVCG.2020.3030411

[13] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, and X. Zhang. 2017. E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 36–47. https://doi.org/10.1109/VAST.2017.8585638

[14] Siming Chen, Lijing Lin, and Xiaoru Yuan. 2017. Social Media Visual Analytics. *Comput. Graph. Forum* 36, 3 (June 2017), 563–587. https://doi.org/10.1111/cgf.13211

[15] C. Collins, F. B. Viegas, and M. Wattenberg. 2009. Parallel Tag Clouds to explore and analyze faceted text corpora. In *2009 IEEE Symposium on Visual Analytics Science and Technology*. 91–98. https://doi.org/10.1109/VAST.2009.5333443

[16] Pier Cortese, Giuseppe Battista, Antonello Moneta, Maurizio Patrignani, and Maurizio Pizzonia. 2006. Topographic Visualization of Prefix Propagation in the Internet. *IEEE transactions on visualization and computer graphics* 12 (10 2006), 725–32. https://doi.org/10.1109/TVCG.2006.185

[17] E. Cuenca, A. Sallaberry, F. Y. Wang, and P. Poncelet. 2018. MultiStream: A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series. *IEEE Transactions on Visualization and Computer Graphics* 24, 12 (2018), 3160–3173. https://doi.org/10.1109/TVCG.2018.2796591

[18] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. 2011. TextFlow: Towards Better Understanding of Evolving Topics in Text. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2412–2421. https://doi.org/10.1109/TVCG.2011.239

[19] Weiwei Cui, Shixia Liu, Zhuofeng Wu, and Hao Wei. 2014. How Hierarchical Topics Evolve in Large Text Corpora. *Visualization and Computer Graphics, IEEE Transactions on* 20 (12 2014), 2281–2290. https://doi.org/10.1109/TVCG.2014.2346433

[20] W. Cui, S. Liu, Z. Wu, and H. Wei. 2014. How Hierarchical Topics Evolve in Large Text Corpora. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2281–2290. https://doi.org/10.1109/TVCG.2014.2346433

[21] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. 2010. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology*. 115–122. https://doi.org/10.1109/VAST.2010.5652922

[22] Marian Dörk, Daniel Gruen, Carey Williamson, and Sheelagh Carpendale. 2010. A visual backchannel for large-scale events. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1129–1138.

[23] Wenwen Dou, Xiaoyu Wang, Drew Skau, William Ribarsky, and Michelle X Zhou. 2012. Leadline: Interactive visual analysis of text data through event identification and exploration. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 93–102.

[24] Marian Dörk, Daniel Gruen, Carey Williamson, and Sheelagh Carpendale. 2011. A Visual Backchannel for Large-Scale Events. *IEEE transactions on visualization and computer graphics* 16 (01 2011), 1129–38. https://doi.org/10.1109/TVCG.2010.129

[25] Mylynn Felt. 2016. Social media and the social sciences: How researchers employ Big Data analytics. *Big data & society* 3, 1 (2016), 2053951716645828.

[26] Emden R Gansner, Yifan Hu, and Stephen North. 2012. Visualizing streaming text data with dynamic graphs and maps. In *International Symposium on Graph Drawing*. Springer, 439–450.

[27] A. Ge, H. Jang, G. Carenini, K. Ho, and Y. J. Lee. 2019. OCTVis: Ontology-Based Comparison of Topic Models. In *2019 IEEE Visualization Conference (VIS)*. 66–70. https://doi.org/10.1109/VISUAL.2019.8933646

[28] Pierre-Etienne Genest and Guy Lapalme. 2011. Framework for abstractive summarization using text-to-text generation. In *Proceedings of the workshop on monolingual text-to-text generation*. 64–73.

[29] Michael Gleicher. 2018. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 413–423. https://doi.org/10.1109/TVCG.2017.2744199

[30] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

[31] Frank Ham, Martin Wattenberg, and Fernanda Viégas. 2009. Mapping Text with Phrase Nets. *IEEE transactions on visualization and computer graphics* 15 (11 2009), 1169–76. https://doi.org/10.1109/TVCG.2009.165

[32] S. Havre, B. Hetzler, and L. Nowell. 2000. ThemeRiver: visualizing theme changes over time. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*. 115–123. https://doi.org/10.1109/INFVIS.2000.885098

[33] Marti Hearst, Emily Pedersen, Lekha Patil, Elsie Lee, Paul Laskowski, and Steven Franconeri. 2019. An Evaluation of Semantically Grouped Word Cloud Designs. *IEEE Transactions on Visualization and Computer Graphics* PP (03 2019), 1–1. https://doi.org/10.1109/TVCG.2019.2904683

[34] Alfred Hermida and Mary Lynn Young. 2019. *Data journalism and the regeneration of news*. Routledge.

[35] M. Hu, K. Wongsuphasawat, and J. Stasko. 2017. Visualizing Social Media Content with SentenTree. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 621–630. https://doi.org/10.1109/TVCG.2016.2598590

[36] Markus John, Eduard Marbach, Steffen Lohmann, Florian Heimerl, and Thomas Ertl. 2018. MultiCloud: Interactive Word Cloud Visualization for the Analysis of Multiple Texts. In *Proceedings of the 44th Graphics Interface Conference* (Toronto, Canada) *(GI '18)*. Canadian Human-Computer Communications Society, Waterloo, CAN, 34–41. https://doi.org/10.20380/GI2018.06

[37] J. Knittel, S. Koch, and T. Ertl. 2020. PyramidTags: Context-, Time- and Word Order-Aware Tag Maps to Explore Large Document Collections. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. https://doi.org/10.1109/TVCG.2020.3010095

[38] Johannes Knittel, Steffen Koch, Tan Tang, Wei Chen, Yingcai Wu, Shixia Liu, and Thomas Ertl. 2021. Real-time visual analysis of high-volume social media posts. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 879–889.

[39] Thomas Kraft, Derek X Wang, Jeffrey Delawder, Wenwen Dou, Li Yu, and William Ribarsky. 2013. Less after-the-fact: Investigative visual analysis of events from streaming twitter. In *2013 IEEE Symposium on Large-Scale Data Analysis and Visualization (LDAV)*. IEEE, 95–103.

[40] K Krishna and M Narasimha Murty. 1999. Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 29, 3 (1999), 433–439.

[41] R. Krüger, Guodao Sun, Fabian Beck, Ronghua Liang, and T. Ertl. 2016. TravelDiff: Visual comparison analytics for massive movement patterns derived from Twitter. *2016 IEEE Pacific Visualization Symposium (PacificVis)* (2016), 176–183.

[42] P. M. Law, R. C. Basole, and Y. Wu. 2019. Duet: Helping Data Analysis Novices Conduct Pairwise Comparisons by Minimal Specification. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 427–437. https://doi.org/10.1109/TVCG.2018.2864526

[43] Bongshin Lee, Nathalie Henry Riche, Amy Karlson, and Sheelagh Carpendale. 2010. SparkClouds: Visualizing Trends in Tag Clouds. *IEEE Trans. Vis. Comput. Graph.* 16 (11 2010), 1182–1189. https://doi.org/10.1109/TVCG.2010.194

[44] Leo Leppänen, Myriam Munezero, Mark Granroth-Wilding, and Hannu Toivonen. 2017. Data-driven news generation for automated journalism. In *Proceedings of the 10th international conference on natural language generation*. 188–197.

[45] G. Li, Y. Zhang, Y. Dong, J. Liang, J. Zhang, J. Wang, M. J. Mcguffin, and X. Yuan. 2020. BarcodeTree: Scalable Comparison of Multiple Hierarchies. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 1022–1032. https://doi.org/10.1109/TVCG.2019.2934535

[46] J. Li, S. Chen, K. Zhang, G. Andrienko, and N. Andrienko. 2019. COPE: Interactive Exploration of Co-Occurrence Patterns in Spatial Time Series. *IEEE Transactions on Visualization and Computer Graphics* 25, 8 (2019), 2554–2567. https://doi.org/10.1109/TVCG.2018.2851227

[47] Shixia Liu, Yingcai Wu, Enxun Wei, and Mengchen Liu. 2013. StoryFlow: Tracking the Evolution of Stories. *IEEE Transactions on Visualization & Computer Graphics* 19, 12 (2013), 2436–2445.

[48] S. Liu, J. Yin, X. Wang, W. Cui, K. Cao, and J. Pei. 2016. Online Visual Analytics of Text Streams. *IEEE Transactions on Visualization and Computer Graphics* 22, 11 (2016), 2451–2466. https://doi.org/10.1109/TVCG.2015.2509990

[49] Shixia Liu, Michelle X. Zhou, Shimei Pan, Weihong Qian, Weijia Cai, and Xiaoxiao Lian. 2009. Interactive, Topic-Based Visual Text Summarization and Analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (Hong Kong, China) *(CIKM '09)*. Association for Computing Machinery, New York, NY, USA, 543–552. https://doi.org/10.1145/1645953.1646023

[50] Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* 2, 11 (2017), 205.

[51] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[52] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[53] Tamara Munzner. 2009. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics* 15, 6 (2009), 921–928.

[54] Jennifer Pan, Zijie Shao, and Yiqing Xu. 2022. How government-controlled media shifts policy attitudes through framing. *Political Science Research and Methods* 10, 2 (2022), 317–332.

[55] Celeste Paul, Jessica Chang, Alex Endert, Nick Cramer, David Gillen, Shawn Hampton, Russ Burtner, Ralph Perko, and Kristin Cook. 2018. TexTonic: Interactive visualization for exploration and discovery of very large text collections. *Information Visualization* 18 (07 2018), 147387161878539. https://doi.org/10.1177/1473871618785390

[56] Alejandro Ramos-Soto, Alberto Jose Bugarin, Senén Barro, and Juan Taboada. 2014. Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data. *IEEE Transactions on Fuzzy Systems* 23, 1 (2014), 44–57.

[57] Peter H Rossi, Mark W Lipsey, and Gary T Henry. 2018. *Evaluation: A systematic approach.* Sage publications.

[58] Rita Sevastjanova, Mennatallah El-Assady, Adam James Bradley, Christopher Collins, Miriam Butt, and Daniel Keim. 2021. VisInReport: Complementing Visual Discourse Analytics through Personalized Insight Reports. *IEEE Transactions on Visualization and Computer Graphics* (2021).

[59] Kate Starbird, Dharma Dailey, Owla Mohamed, Gina Lee, and Emma S Spiro. 2018. Engage early, correct more: How journalists participate in false rumors online during crisis events. In *Proceedings of the 2018 CHI conference on human factors in computing systems.* 1–12.

[60] Maarten Van Someren, Yvonne F Barnard, and J Sandberg. 1994. The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress* 11 (1994).

[61] Ji Wang, Jian Zhao, Sheng Guo, Chris North, and Naren Ramakrishnan. 2014. ReCloud: semantics-based word cloud visualization of user reviews.. In *Graphics Interface.* 151–158.

[62] M. Wattenberg and F. B. Viégas. 2008. The Word Tree, an Interactive Visual Concordance. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1221–1228. https://doi.org/10.1109/TVCG.2008.172

[63] Y. Wu, Z. Chen, G. Sun, X. Xie, N. Cao, S. Liu, and W. Cui. 2018. StreamExplorer: A Multi-Stage System for Visually Exploring Events in Social Streams. *IEEE Transactions on Visualization and Computer Graphics* 24, 10 (2018), 2758–2772. https://doi.org/10.1109/TVCG.2017.2764459

[64] Yingcai Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. 2014. Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 1763–1772.

[65] Panpan Xu, Yingcai Wu, Enxun Wei, Tai-Quan Peng, Shixia Liu, Jonathan JH Zhu, and Huamin Qu. 2013. Visual analysis of topic competition on social media. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2012–2021.

Received January 2022; revised July 2022; accepted November 2022