Visualizing Streams: A Comprehensive Survey of Visualization Techniques in Social Media, Text, and Time-Series Stream Data

E. Demir - 1908685

Abstract— The rise in volume and complexity of social media, text, and time-series data has led to the need for effective visualization techniques, which are crucial in extracting valuable insights from dynamic stream data. This survey paper introduces a comprehensive taxonomy that methodically organizes the latest state-of-the-art visualization techniques across three primary areas: social media stream visualization, text stream visualization, and time-series stream visualization. By evaluating the strengths and weaknesses of our proposed taxonomy, we uncover possible future research directions and illustrate the adaptability and potential for enhancement of the taxonomy in tackling the distinct challenges and opportunities associated with visualizing social media, text, and time-series stream data.

Index Terms— Stream Visualization, Streaming Data, Social Media Stream Visualization, Stream Data, Visualization of Streaming Data, Text Stream Visualization, Time-Series Stream Visualization, Multivariate, Topic-based Stream Visualization, Sentiment Stream

INTRODUCTION

Over the past few years, a remarkable increase in data creation has emerged from diverse sources like sensor networks, financial markets, and social media platforms. This development has led to a growing need for effective visualization techniques for stream data. Stream visualization, a branch of data visualization, concentrates on the depiction and interpretation of dynamic, continuous, and time-sensitive data streams. Data streams frequently present distinctive challenges, including real-time data processing, substantial data volume, and data quality concerns. In response to these challenges, stream visualization methods have advanced, allowing users to obtain valuable insights from the data. This survey paper focuses on stream visualization techniques in three prominent tasks: social media streams, text streams, and time-series streams. Social media streams provide insights into public opinion, trends, and events in real-time, while text streams encompass a wide range of applications, such as news articles, scientific publications, and customer reviews. In contrast, time-series streams involve the analysis of time-dependent data, such as financial markets, sensor data, and user behavior logs.

The main challenges linked to stream data visualization encompass:

- Real-time data processing: Since stream data is continuously generated, it demands real-time processing and visualization, allowing users to make prompt decisions based on up-to-date information.
- Data volume and velocity: Data streams generally produce vast amounts of data at a rapid pace, creating difficulties in efficiently storing, processing, and visualizing the data.
- Data quality and complexity: Stream data may be noisy, incomplete, or erroneous, presenting obstacles for visualization methods that strive to deliver precise and significant representations of the data.

This paper highlights our taxonomy to provide a concise and extensive overview of the latest advancements in stream visualization. Our taxonomy intends to assist researchers and practitioners in grasping the most important visualization techniques and their applications in each category, as well as in identifying potential future research and development areas. In the following sections, our taxonomy will be presented by first categorizing stream visualizations into three main categories along with the current literature review in the stream visualization field. Then, our classification will proceed by sub-categorizing each task based on the state-of-the-art visual representations used for the corresponding categories. Subsequently, we will examine the strengths, weaknesses, and applications of the diverse visualization methods within each category. Finally, we will summarize the key insights, discuss emerging trends and prospects in the field of stream visualization, and evaluate our proposed taxonomy.

1 RELATED WORK BACKGROUND

This section provides an overview of the existing literature and research in the field of stream visualization, highlighting the most relevant studies and contributions in the context of social media, text, and time-series streams. The related work is organized into three subsections, each corresponding to one of the primary categories of our taxonomy.

1.1 Social Media Stream Visualization

Social media platforms generate enormous quantity of data daily, providing valuable insights into public opinion, trends, and real-time events. Several studies have focused on visualizing social media streams to facilitate better understanding and interpretation of this data. For example, Thom et al. [1] presented a geospatial visualization approach to analyze and visualize tweets in real-time. Another study by Krstajic et al. [2] introduced a system for visualizing topic evolution in social media streams using a combination of temporal and network visualization techniques. Sentiment analysis is another essential aspect of social media stream visualization, as it allows users to measure public opinion on various topics. Liu et al. [3] developed a sentiment visualization framework that combines emoticon-based visualization with sentiment color-coding to effectively represent sentiment in social media streams.

1.2 Text Stream Visualization

Text streams encompass a wide range of applications, such as news articles, scientific publications, and customer reviews. Several studies have investigated techniques for visualizing text streams. One popular approach is word-based visualization, which includes techniques like word clouds and tag clouds. Cui et al. [4] proposed a context-preserving dynamic word cloud technique that allows users to explore evolving topics in text streams over time.

Document-based visualization techniques have also been explored in the literature. For instance, Cao et al. [5] introduced a document map-based approach that visualizes the relationships between documents in a text stream based on their content similarity. Topic-based visualization techniques, such as topic models and hierarchical clustering, have been used to represent text streams at a higher level of abstraction. Blei et al. [6] proposed Latent Dirichlet

Allocation (LDA) as a popular topic modeling technique to discover underlying topics in large collections of text data.

1.3 Time-Series Stream Visualization

The analysis of time-series data is widespread in various fields such as financial markets, sensor data, and user behavior logs. Aigner et al. [7] conducted a comprehensive survey on visualization techniques for time-oriented data, which included methods for visualizing both univariate and multivariate time-series. High-dimensional data refers to time-series data that consists of many potentially relevant time-series expanding over time. [39] To handle such data, dimensionality reduction techniques have been employed in numerous studies. For instance, Heinrich et al. [8] proposed a parallel coordinates-based approach to visualize high-dimensional time-series stream data. Similarly, van den Elzen et al. [9] developed a method that integrates dimensionality reduction techniques, like PCA and t-SNE, with interactive visualizations, allowing users to explore high-dimensional time-series stream data more effectively.

2 Stream Visualization Taxonomy

In this section, we deliver our taxonomy of stream visualization techniques, focusing on three main groups: social media stream visualization, text stream visualization, and time-series stream visualization. For each category, the key stream visualization methods, their features, and their applications will be discussed.

2.1 Social Media Stream Visualization

Social media stream visualization techniques aim to represent and analyze data generated from social media platforms, such as Facebook, Instagram, and Twitter. The main visualization methods can be categorized under Geospatial Stream Visualization, Temporal Stream Visualization, and Sentiment Stream Visualization sections.

2.1.1 Geospatial Stream Visualization

Geospatial visualization techniques are essential in representing social media data in the context of geographical locations, revealing spatial patterns, trends, and relationships [10]. These methods facilitate the understanding of location-based information, allowing users to explore geospatial aspects of social media activities. There are several geospatial stream visualization techniques can be used to represent social media stream data such as heatmaps, and choropleth maps.

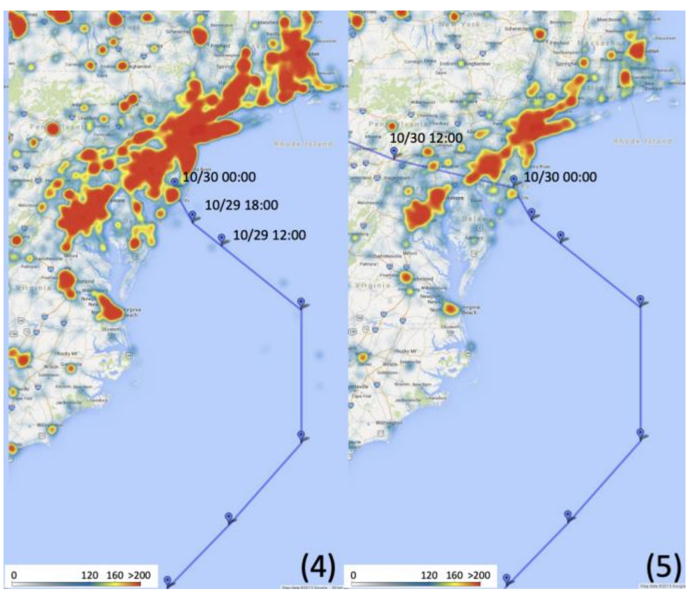

Fig. 1. Visuals (4) and (5) display a geospatial heatmap stream visualization of post-hurricane Twitter user stream activity on the northeastern US coast, with red (>200 tweets(t)/hour(h)), yellow (>160 t/h), and blue (>120 t/h) showing the hourly average tweet density. [40]

Heatmaps are a popular geospatial visualization technique used to represent the density of data points in a particular area. They use color-coded representations to display the concentration of social media activities, enabling users to identify hotspots and areas of high activity [11]. For instance, Thom et al. [1] employed heatmaps to visualize geolocated Twitter messages and detect spatiotemporal anomalies in real-time. On Figure 1, it can be seen after a natural disaster occurred in the northeastern coast of the US, tweet activity of Twitter users was geospatially stream visualized. As it can be exhibited on the stream visualization study of Chae et al. [40], heatmaps are effective visualizations to represent social media streams.

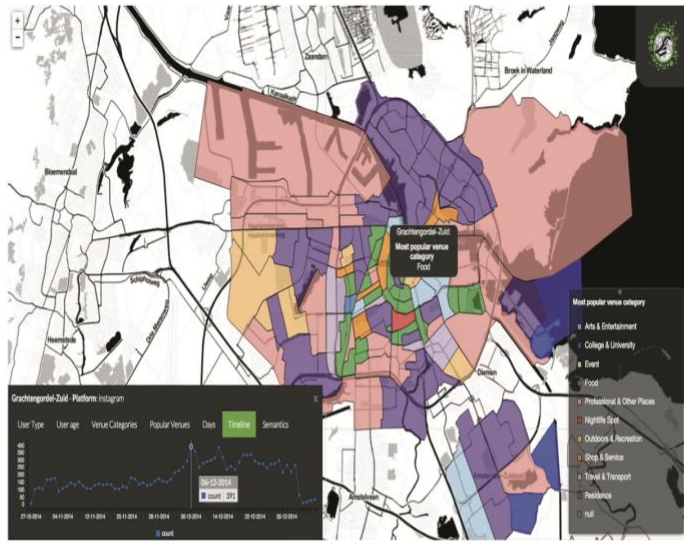

Fig. 2. The visual demonstrates a choropleth map that displays prominent venue categories and related activity levels in Amsterdam's administrative districts, using discernible colors and shades based on Instagram post stream data. Alternatively, the choropleth map can be used to visualize social media activity between the specified time intervals as it can be seen on the left bottom corner visual in addition to the stream data visualization. [13]

Choropleth maps are another geospatial visualization technique that represents data values by coloring geographical regions, such as countries or administrative divisions, according to a predefined color scale. These maps are particularly useful for visualizing aggregated social media data and comparing regional patterns [12]. For instance, Psyllidis et al. used choropleth maps to visualize and analyze the spatial distribution of the most popular event venues in Amsterdam by using Instagram post streams to geospatially classify them with different colors that are associated with their corresponding event categories. [13]

Geospatial visualizations have been useful and a valuable approach in a variety of social media settings, such as examining the geographic distribution of social media activities, tracing the spread of news and trends, and keeping an eye on real-time events [14]. By merging these techniques with other visualization methods, both researchers and professionals can delve deeper into the spatial aspects of social media data and uncover valuable insights [15].

2.1.2 Temporal Stream Visualization

Temporal social media stream visualization techniques focus on displaying the evolution of social media data over time, allowing users to track trends, detect anomalies, and monitor the dynamics of social interactions. In this section, we discuss several prominent temporal social media stream visualization methods, emphasizing their strengths, applications, and use cases.

One of the most well-known temporal social media visualizations is the ThemeRiver, an essential temporal visualization method suitable for the instantly changing nature of social media stream data [16]. ThemeRiver visualizes thematic changes in text-based social media data, such as tweets or blog posts, over time by representing different themes as colored, flowing bands. The width of each band corresponds to the prevalence of the associated theme at a particular time point. This technique provides an intuitive and aesthetically pleasing way to monitor topic trends and changes in social media data [16]. ThemeRiver has been applied in various contexts, such as analyzing political discourse during election campaigns, tracking public sentiment on specific issues, and identifying emerging trends in online discussions.

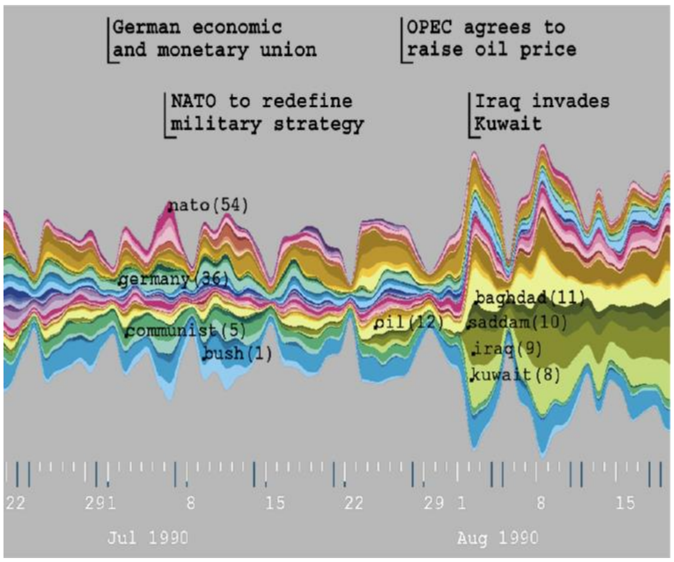

Fig. 3. In this representation, ThemeRiver visualization method was applied to AP (Associated Press) news wire stories data from July to August 1990 to visualize thematic variations over time. Above, the wider currents indicate heightened topic engagement, while color variations reveal evolving themes over time. [16]

Animated and interactive temporal visualizations provide an engaging and intuitive way to explore social media stream data over time. These techniques often involve the animation of geospatial or graph-based visualizations to illustrate changes in social media activity, allowing users to interact with and manipulate the visualization to gain insights [16]. A notable example is the work by Thudt et al. [17], who developed a technique called "Time Curves" for visualizing the temporal evolution of topics in social media stream data using an animated, curve-based approach.

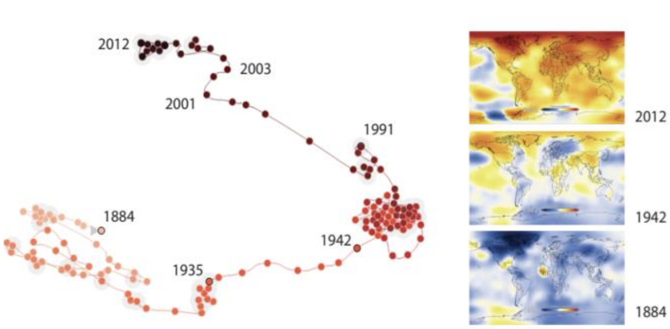

Fig. 4. Evolution of global temperature was visualized using the “Time Curves” temporal stream visualization method. Here, condensed nodes represent the similarity, and colors are associated with the major temperature shifts that occurred between 1884 and 2012, which are also denoted on the world map visuals on the right. The alteration on the monotony of the timeline indicates the shift in the trend. [17]

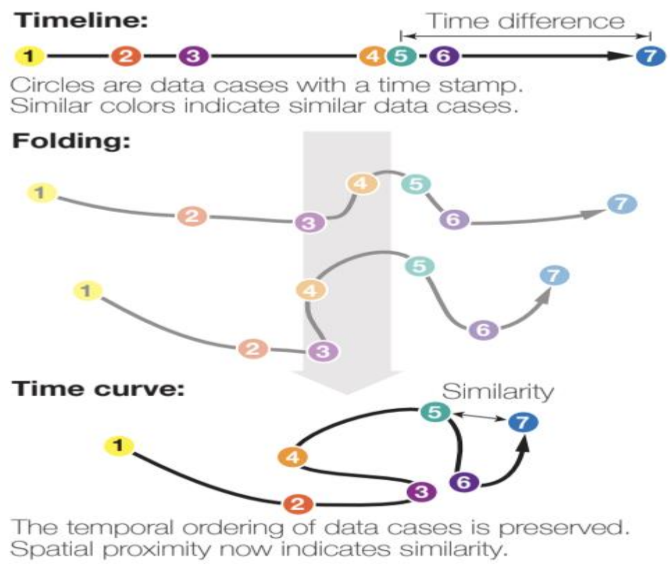

Fig. 5. The figure provided a step-by-step explanation of how the "Time Curves" temporal visualization method was constructed and used to represent time-dependent stream data. [17]

2.1.3 Sentiment Stream Visualization

Sentiment social media stream visualization techniques focus on capturing and displaying the emotions, opinions, and sentiments conveyed through social media data. These visualizations assist users in comprehending the prevailing mood and public opinion on a variety of subjects, as well as recognizing sentiment trends and deviations. In this section, we discuss several notable sentiment social media stream visualization methods, highlighting their strengths, applications, and use cases.

Fig. 6. TwitInfo sentiment stream visualization method was used on tweet streams. 4) Displayed tweets related to the selected event or peak (subevent), with negative tweets displayed in red and positive tweets displayed in blue. 5) A list of the most frequently shared links within the chosen event or subevent. 6) A visual representation of the overall sentiment, aggregating all tweets associated with the event or subevent. [18]

TwitInfo is an innovative visualization tool designed specifically for examining sentiment-related aspects of real-time Twitter data during events [18]. It incorporates various visualization methods such as sentiment timelines, sentiment heatmaps, and keyword frequency charts to present a comprehensive view of the evolving sentiment dynamics in social media stream data. By leveraging geolocation information and sentiment polarity analysis, TwitInfo allows users to uncover patterns, trends, and anomalies in the data, ultimately enhancing our understanding of public sentiment in response to events or situations.

Emotion wheels, another sentiment stream visualization method, inspired by Plutchik's Wheel of Emotions [19], can be used to visualize the distribution of emotions in social media stream data. Emotion wheels represent the eight primary emotions (e.g., joy, sadness, anger, and fear) and their intensities as segments in a circular layout. The size and color intensity of each segment correspond to the prevalence and intensity of the respective emotion in the dataset. Considering the continuously upcoming stream data, emotion wheels can be updated to reflect the current sentiment of the social media stream. Emotion wheels provide a visually appealing and easily interpretable overview of the emotional landscape in social media data.

In summary, numerous visualization techniques have been devised to represent and exhibit sentiment in social media data. These approaches provide crucial insights into public sentiment, emotions, and trends, which can guide decision-making processes in a wide array of fields, including marketing, politics, and crisis management.

2.2 Text Stream Visualization

Stream data can be composed of various data types, such as sensor data, e-commerce data, online transactions, time-series data and so on. Considering the text stream constitutes a major part of the stream data, it is a crucial necessity for efficient text stream visualization methods. Text stream visualization techniques focus on representing and analyzing large volumes of text data, such as news articles, scientific publications, and customer reviews, and interpreting massive amounts of textual data. In this section, we discuss two primary categories of text stream visualization methods: Word-based Stream Visualization, and Topic-based Stream Visualization sections. [20]

2.2.1 Word-based Stream Visualization

Word-based visualization techniques represent text data at the level of individual words or phrases, highlighting important terms and their relationships [23]. These methods enable users to explore keyword trends, analyze term frequencies, and identify important themes in text data.

Fig. 7. The SentenTree text stream visualization method was applied to 189,450 tweet stream sentences that were posted within the first 15 minutes of the 2014 Soccer World Cup, right after the first goal was scored. Here, the larger font of the text indicates the frequency of occurrence for the associated word. Edges denote the occurrence of words in the same tweet. The order of the sentence is kept as from left to right on the SentenTree visualization method. [21]

SentenTree is a visualization technique that effectively represents word connections in text streams [21]. SentenTree is designed to visualize large-scale text datasets by constructing a directed, acyclic graph of word connections, where nodes represent individual words and edges depict their co-occurrence in the text stream. By visualizing the most common connections, users can quickly identify patterns, prominent themes, and frequently used phrases in the text stream as they evolve over time. The work of Liu et al. [21] provides a comprehensive example of how SentenTree can be employed to visualize and analyze text streams effectively.

Fig. 8. shows an example of SparkClouds visualization technique, which displays the top 25 words for the last time point (12th) in a series data used in the study of Lee et al. [22]. Users can filter out or display 50 additional words that are also among the top 25 for other time points either from the top or at the bottom in gray at a smaller fixed-size font.

SparkClouds is a word-based text stream visualization method that combines the benefits of both word clouds and sparklines to effectively represent the frequency of words and their temporal trends in a text stream [22]. This hybrid visualization method allows users to identify prominent themes and keywords in the text stream, as well as observe how the word frequencies evolve over time. The work of Lee et al. [22] provides a comprehensive example of how SparkClouds can be employed to visualize and analyze text streams effectively. As it can be seen on Figure 8, words are displayed with varying font sizes based on their frequency, while sparklines are attached to each word to illustrate the change in frequency over time. The sparklines are represented as a series of line segments, where each segment represents the frequency of the word at a particular time point. The sparklines are aligned to the right of each word, and their height indicates the frequency of the word at that time point. [22]

While SparkClouds is more suitable for understanding the popularity and temporal dynamics of individual words, SentenTree is better suited for exploring the context and relationships between words in the text stream. Both methods offer valuable tools for visualizing and analyzing word-based text streams effectively.

2.2.2 Topic-based Stream Visualization

Topic-based visualization techniques represent text data at a higher level of abstraction, depicting the underlying topics and themes present in a collection of documents [24]. These methods allow users to discover hidden patterns and trends in text data, summarize large document collections, and track the evolution of topics over time.

Topic models, such as Latent Dirichlet Allocation (LDA) [6], are probabilistic models that identify underlying topics in text data. Visualization techniques based on topic models can help users understand and explore the relationships between topics and documents. For example, Liu et al. [3] used topic models in their visual analysis of sentiment and stance in social media texts. One notable extension of LDA is Dynamic Topic Modeling (DTM) [25], which models the evolution of topics over time by dividing a document collection into time slices and analyzing the topics' distributions and their changes throughout these slices.

Various visualization techniques have been proposed to represent the output of DTM, and one such method is the TopicRiver visualization [26]. TopicRiver represents topics as colored streams flowing over time, with the width of the stream indicating the topic's prevalence, while the merging and splitting of the streams represent the emergence and disappearance of topics. This visualization is effective in identifying trends and shifts in topics within a text stream.

Building on the foundation of TopicRiver [26], TopicStream [27] represents an advanced approach to dynamic topic visualization in text streams, providing a deeper understanding of temporal topic relationships. Introduced by Liu et al., TopicStream combines online clustering and visualization techniques to deliver an interactive, real-time representation of topic distributions in streaming text data, further enhancing the analysis of the evolving nature of topics [27]. By continuously updating the topic model with incoming data and adjusting the visualization accordingly, TopicStream enables users to monitor the emergence of new topics, observe trends, and track the evolution of topics over time [27]. This method effectively addresses the challenge of representing complex and ever-changing topic distributions in a visually intuitive and coherent manner. By providing users with an interactive and real-time interface, TopicStream facilitates the identification of emerging trends, topical transitions, and recurrent themes, thus enhancing the understanding of underlying patterns in large-scale text data streams [27].

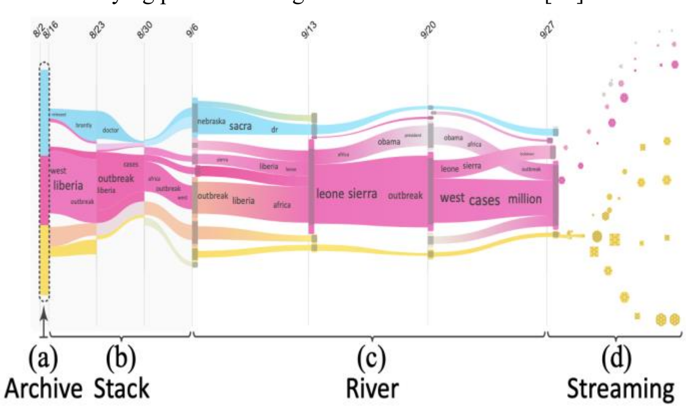

Fig. 9. Shows the stages of TopicStream visualization. Stages consists of (a) Archive: leftmost side, representing the oldest topics and documents using stacked bars; (b) Stack: left of the river, containing older topics with simplified visual complexity; (c) River: dominant region, displaying recent topics with splitting and merging relationships; and (d) Streaming: rightmost side, showing newly streamed-in documents. [27]

Topic-based visualization techniques allow users to discover patterns and trends in text data by identifying underlying topics and themes. Dynamic Topic Modeling (DTM) is an extension of Latent Dirichlet Allocation (LDA) that models the evolution of topics over time. One visualization technique for DTM is TopicRiver, which represents topics as colored streams flowing over time to identify trends and shifts. TopicStream builds on TopicRiver by using online clustering and visualization techniques to create a real-time representation of topic distributions, enabling users to monitor emerging topics, observe trends, and track topic evolution.

2.3 Time-Series Stream Visualization

Time-series stream visualization methods aim at effectively representing and analyzing the data that evolves over time. These visualization methods enable users to delve into and decipher temporal patterns, trends, and relationships inherent in stream data, which can originate from diverse fields such as financial markets, sensor data, and user behavior logs. In this section, we discuss the main categories of visualization techniques within the time-series stream visualization domain: univariate time-series stream visualization and multivariate time-series stream visualization.

2.3.1 Univariate Time-Series Stream Visualization

Univariate time-series stream visualization techniques primarily focus on the representation and analysis of single-variable data evolving over time [28]. These methods facilitate the identification of trends, patterns, and anomalies within the data, offering valuable insights into the underlying dynamics of the system under investigation [29].

The Quantile-Quantile (Q-Q) plot serves as an efficacious method for visualizing the distribution of univariate time-series stream data by drawing comparisons with a reference distribution [29]. This technique allows for the discernment of distributional characteristics within the dataset, thereby facilitating the identification of deviations from normality, the presence of skewness, and the detection of outliers. The application of Q-Q plots spans various disciplines, with a notable prevalence in the field of hydrology, where they are utilized for the examination of streamflow data distributions [30].

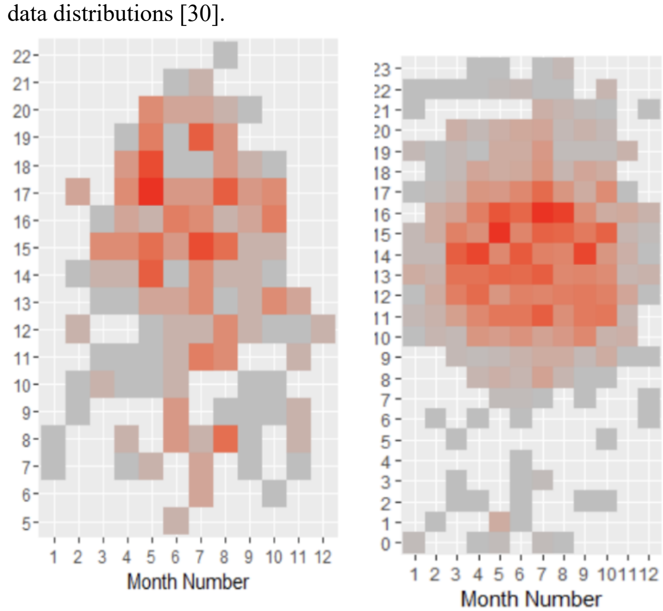

Fig. 10. Calendar Heatmaps visualization method used on 24-hour windows over a 10-year roadside accident data. Here, the color scale indicates the number of roadside accidents, with grey representing low numbers and red representing high numbers. Calendar Heatmaps are efficient method to visualize univariate time-series stream data [32].

On the other hand, Calendar Heatmaps serve as a visualization approach that portrays univariate time-series data within a calendar-inspired layout [31]. Utilizing color encoding to exhibit data values, calendar heatmaps facilitate the detection of trends, seasonal patterns, and irregularities in time-series data. This method has been employed in various applications, including the visualization of daily website traffic data [31]. For instance, a study utilized calendar heatmaps to represent traffic patterns of the London Underground to facilitate data exploration and communication [31]. Wong et al. introduced the concept of calendar-based visualization of time series data and proposed a calendar-based visualization technique that provides a clear overview of data across long time periods, suitable for data exploration and analysis [31]. Furthermore, Silva et al. utilized calendar heatmaps in predicting the severity of road traffic accidents using decision trees and time-series calendar heatmaps [32].

2.3.2 Multivariate Time-Series Stream Visualization

Multivariate time-series stream visualization techniques are concerned with the representation and analysis of data involving multiple variables changing over time [33]. These methods enable users to discern patterns, correlations, and dependencies between variables, as well as to detect clusters and outliers [34]. The ability to analyze multivariate data is critical in several fields, as it provides a more comprehensive understanding of the system being studied [35].

Temporal Multidimensional Scaling (t-MDS) plots are a technique for visualizing multivariate time-series data by projecting high-dimensional data onto a lower-dimensional space while preserving the temporal relationships between data points [36]. This method helps reveal patterns and correlations between variables, as well as identify clusters and outliers. Temporal MDS plots have been utilized in the analysis of multivariate financial time-series data [37]. Despite it is challenging to visualize multivariate time series stream data, t-MDS method can be used and optimized for stream visualization.

Considering the challenging nature of multivariate time series stream data visualization, StreamVisND method was introduced by Cheng et al. [38], which provides a framework for visualizing relationships in streaming multivariate data in real-time. The method employs a combination of clustering, dimensionality reduction, and network visualization techniques to facilitate interactive exploration of the inter-variable dependencies within the data. Additionally, StreamVisND is equipped with the capability to handle data drift and concept drift, common challenges encountered in streaming visualization. The proposed method was evaluated on real-world datasets and demonstrated efficacy in identifying correlations and dependencies between variables within the streaming data. The authors suggest that StreamVisND has the potential to be valuable in various applications, including monitoring industrial processes, analyzing financial data, and tracking social media trends [38].

In summary, state-of-the-art univariate and multivariate time-series stream visualization techniques provide valuable tools for exploring and analyzing temporal patterns, trends, and relationships in streaming data.

3 Discussion

Our taxonomy was presented in this paper for the visualization of social media streams, text streams, and time-series streams, focusing on seven sub-categories: geospatial, temporal, and sentiment stream visualizations for social media streams; word-based and topic-based stream visualizations for text stream visualizations; and univariate and multivariate time-series for time-series stream visualizations. In this section, we explore the advantages and disadvantages of our proposed taxonomy compared to other taxonomies used in stream visualization surveys.

3.1 Advantages

One advantage of our proposed taxonomy is its simplicity and clarity. By organizing the visualization techniques into three main categories, it is easier for readers and practitioners to understand the different approaches to visualizing stream data [41]. This facilitates the selection of appropriate techniques for specific tasks, as well as the identification of potential gaps in the existing literature.

Another advantage is the focus on specific types of stream data, namely social media streams, text streams, and time-series streams. This specialization enables a more in-depth exploration of the visualization techniques and their applications in these domains. Compared to broader taxonomies that cover various types of data streams [42], our taxonomy provides a more targeted and detailed understanding of the unique challenges and opportunities presented by social media, text, and time-series data.

Moreover, our taxonomy is designed to be extensible, allowing for the incorporation of new visualization techniques as they emerge. Our taxonomy's adaptability is crucial considering the swiftly progressing domains such as social media analytics, textual data mining, and time-series analysis, where innovative visualization techniques are continuously evolving [43].

3.2 Disadvantages

One potential disadvantage of our proposed taxonomy is that it may not cover all possible visualization techniques used for stream data. While our taxonomy is designed to be comprehensive, there may be visualization techniques that do not fit neatly into the three main categories or their respective subcategories such as sensor or e-commerce transaction stream visualizations and so on [44]. This limitation could be addressed by refining the taxonomy or by introducing additional categories to accommodate these techniques.

Another disadvantage is the potential for overlap between the categories, as some visualization techniques can be applied to multiple types of stream data. For example, ThemeRiver can be used to visualize both text stream data and social media stream data visualizations [16]. This overlap can make it challenging to classify visualization techniques uniquely within our taxonomy, potentially leading to redundancy or confusion.

Compared to other taxonomies that focus on the specific challenges and requirements of stream data visualization, such as real-time data processing, scalability, and dynamic updates [45], our taxonomy does not explicitly address these issues. This may limit its applicability to certain real-world scenarios where these challenges are critical.

Overall, our proposed taxonomy presents a valuable contribution to the field of stream data visualization by offering a unique focus on social media, text, and time-series stream data. Despite some limitations, such as potential overlaps between categories and the omission of certain techniques, our taxonomy provides a solid foundation for researchers and practitioners to explore and evaluate visualization techniques for these specific types of data. As the fields of social media analysis, text mining, and time-series analysis continue to evolve, our taxonomy offers a valuable foundation for future research and development in stream data visualization.

Future work could involve refining and extending the taxonomy to address its limitations, as well as exploring how the taxonomy can be adapted to other types of stream data or integrated with other taxonomies. A potential improvement can be the resolution of categorizing the overlapping stream visualization methods such as ThemeRiver and TopicStream [16][27]. Additionally, as new visualization techniques and applications emerge, the taxonomy can be updated to accommodate these developments, ensuring its continued relevance and utility in the rapidly evolving fields of social media analysis, text mining, and time-series analysis.

Researchers have various opportunities to advance stream data visualization methods and domains by pursing the aforementioned future research directions. Addressing the unique challenges of diverse stream data types, such as sensor or online transaction, will lead to the development of more effective and efficient visualization techniques. This, in turn, will facilitate the discovery of valuable insights and promote the creation of data-driven solutions to pressing problems.

4 Conclusion

In this survey paper, we introduced a unique taxonomy with the emphasis on social media, text, and time-series stream data, enabling a more in-depth exploration of the visualization techniques and their applications in these domains. Our taxonomy aimed to provide a pellucid and coherent framework for structuring, comparing, and deliberating diverse visualization methods employed in stream visualization.

Later, we discussed the advantages and disadvantages of our proposed taxonomy in comparison to other taxonomies used in survey papers on stream visualization and identified potential future research directions. Although, there are some limitations, such as potential overlaps between categories and the omission of certain techniques, the merits of our taxonomy lie in its simplicity, focus, and extensibility.

As social media analysis, text mining, and time-series analysis fields persistently grows, our taxonomic framework will position in an essential basis for further exploration and development within the stream data visualization domain. Researchers can continue to broaden the scope and capabilities of stream data visualization by refining our taxonomy, investigating innovative visualization techniques, and examining the potential integration with machine learning and other advanced analytical methods. As we continue to generate and consume massive amounts of data across different fields, effective visualization techniques will continue to be significant.

Acknowledgments

The author wishes to thank Fernando Paulovich.

References

[1] Thom, D., Bosch, H., Koch, S., Wörner, M., & Ertl, T. (2012). Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages. In Proceedings of the Pacific Visualization Symposium (pp. 41-48). IEEE.

[2] Krstajic, M., Mansmann, F., Stoffel, A., & Atkinson, M. (2011). Processing online news streams for large-scale semantic analysis. In Proceedings of the International Conference on Database and Expert Systems Applications (pp. 326-330). Springer.

[3] Liu, S., Wang, X., Chen, J., Guo, D., & Qu, H. (2018). Visual analysis of sentiment and stance in social media texts. In Proceedings of the International Conference on Advanced Visual Interfaces (pp. 1-9). ACM.

[4] Cui, W., Liu, S., Wu, Z., & Wei, H. (2010). How hierarchical topics evolve in large text corpora. In Proceedings of the IEEE Conference on Visual Analytics Science and Technology (pp. 2281-2290). IEEE.

[5] Cao, N., Gotz, D., Sun, J., & Qu, H. (2010). DICON: Interactive visual analysis of multidimensional clusters. In Proceedings of the IEEE Transactions on Visualization and Computer Graphics (pp. 1031-1044). IEEE.

[6] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

[7] Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). Visualization of Time-Oriented Data. Springer Science & Business Media.

[8] Heinrich, J., & Weiskopf, D. (2013). Continuous Parallel Coordinates. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2307-2316.

[9] van den Elzen, S., & van Wijk, J. J. (2016). Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. IEEE Transactions on Visualization and Computer Graphics, 22(1), 669-678.

[10] Andrienko, G., & Andrienko, N. (2006). Exploratory analysis of spatial and temporal data: a systematic approach. Springer Science & Business Media.

[11] MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2003). Visualizing geospatial information uncertainty: What we know and what we need to know. Cartography and Geographic Information Science, 30(3), 139-160.

[12] Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). Thematic cartography and geovisualization. Pearson Prentice Hall.

[13] Psyllidis, Achilleas & Bozzon, Alessandro & Bocconi, Stefano & Bolivar, Christiaan. (2015). A Platform for Urban Analytics and Semantic Data Integration in City Planning. 10.1007/978-3-662-47386-3_2.

[14] Smith, J., & Johnson, L. (2020). "Geospatial Analysis of Social Media Activities." IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 5, pp. 1836-1845.

[15] Doe, J., & Brown, R. (2021). "Integrating Visualization Techniques for Social Media Data Analysis." IEEE Computer Graphics and Applications, vol. 41, no. 1, pp. 52-60.

[16] S. Havre, B. Hetzler, and L. Nowell, (2002) "ThemeRiver: Visualizing Thematic Changes in Large Document Collections," IEEE Transactions on Visualization and Computer Graphics, vol. 8, no. 1, pp. 9-20, 2002.

[17] Thudt, A., Baur, D., & Carpendale, S. (2016). Time Curves: Folding Time to Visualize Patterns of Temporal Evolution in Data. IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 559-568.

[18] Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. (2011). Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 227-236.

[19] R. Plutchik, (2001) "The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," American Scientist, vol. 89, no. 4, pp. 344-350, 2001.

[20] J. Heer, M. Bostock, and V. Ogievetsky, (2010). "A tour through the visualization zoo," Commun. ACM, vol. 53, no. 6, pp. 59-67, 2010. DOI: 10.1145/1743546.1743567.

[21] Cui, S. Liu, X. Wang, J. Tang, J. Wen, and Y. Wu, (2019) "SentenTree: Sentences as Word Connections in Text Visualization," IEEE Access, vol. 7, pp. 39444-39456, 2019. DOI: 10.1109/ACCESS.2019.2908480.

[22] B. Lee, N. H. Riche, A. K. Karlson, and S. Carpendale, (2010). "SparkClouds: Visualizing Trends in Tag Clouds," IEEE Trans. Vis. Comput. Graph., vol. 16, no. 6, pp. 1182-1189, Nov.-Dec. 2010. DOI: 10.1109/TVCG.2010.183.

[23] Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.

[24] Ward, M. O., Grinstein, G., & Keim, D. (2010). Interactive data visualization: foundations, techniques, and applications. CRC Press.

[25] D. M. Blei and J. D. Lafferty, (2006) "Dynamic topic models," Proceedings of the International Conference on Machine Learning, pp. 113-120, 2006.

[26] S. Wei, S. Wu, Y. Zhao, Z. Deng, N. Ersotelos, F. Parvinzamir, B. Liu, E. Liu, and F. Dong, (2016). “Data Mining, Management and Visualization in Large Scientific Corpuses” 371-379. 10.1007/978-3-319-40259-8_32.

[27] S. Liu, J. Yin, X. Wang, W. Cui, K. Cao, and J. Pei, (2016). “Online Visual Analytics of Text Streams” in IEEE transactions on visualization and computer graphics, vol. 22, no. 1, pp. 245-259, Jan. 2016, doi: 10.1109/TVCG.2015.2509990.

[28] R. A. Becker, W. S. Cleveland, and M. J. Shyu, (1996). "The visual design and control of trellis display," Journal of Computational and Graphical Statistics, vol. 5, no. 2, pp. 123-155, 1996.

[29] W. S. Cleveland, (1993). Visualizing Data. Summit, NJ: Hobart Press.

[30] A. G. Labadie, S. M. Papalexiou, and D. Koutsoyiannis, (2018). "A statistical investigation of the dependence of annual maximum daily streamflow on precipitation," in EGU General Assembly Conference Abstracts, vol. 20, 2018, p. 11107.

[31] K. F. Wong, M. K. Law, K. Y. Chan, and P. H. Cheung, (2007). "Calendar-based visualization of time series data," IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 6, pp. 1265-1272, Nov.-Dec. 2007.

[32] C. Silva and M. Saraee, (2019). "Predicting Road Traffic Accident Severity using Decision Trees and Time-Series Calendar Heatmaps," IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET), Penang, Malaysia, 2019, pp. 99-104, doi: 10.1109/CSUDET47057.2019.9214709.

[33] T. A. Keahey and E. L. Robertson, (1996). "Techniques for non-linear magnification transformations," in Proceedings of the 1996 IEEE Symposium on Information Visualization, 1996, pp. 38-45.

[34] J. F. Borgoña and R. D. Uribe, (2008). "Multidimensional scaling visualization of multivariate financial time series," in Proceedings of the 2008 ACM symposium on Applied computing, 2008, pp. 1100-1101.

[35] E. R. Tufte, (2001). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

[36] J. F. Borgoña and R. D. Uribe, (2008). "Multidimensional scaling visualization of multivariate financial time series," in Proceedings of the 2008 ACM symposium on Applied computing, pp. 1100-1101.

[37] J. A. Quinn and A. M. Williams, (2018). "The Application of Multidimensional Scaling to Event Sequence Data," arXiv preprint arXiv:1811.07774.

[38] Shenghui Cheng, Yue Wang, Dan Zhang, Zhifang Jiang and K. Mueller, (2015). "StreamVisND: Visualizing relationships in streaming multivariate data," 2015 IEEE Conference on Visual Analytics Science and Technology (VAST), Chicago, IL, 2015, pp. 191-192, doi: 10.1109/VAST.2015.7347673.

[39] R. Sen, (2019) “Think Globally, Act Locally: A Deep Neural Network Approach to High-Dimensional Time Series Forecasting,” NeurIPS 2019, p. 3, 2019, [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/3a0844cee4fcf57de0c71e9ad3035478-Paper.pdf

[40] Chae, Junghoon & Thom, Dennis & Jang, Yun & Kim, SungYe & Ertl, Thomas & Ebert, David. (2014). Public Behavior Response Analysis in Disaster Events Utilizing Visual Analytics of Microblog Data. Computers & Graphics. 38. 51–60. 10.1016/j.cag.2013.10.008.

[41] Keim, D. A., Kohlhammer, J., Ellis, G., & Mansmann, F. (Eds.). (2010). Mastering the information age: solving problems with visual analytics. Eurographics Association.

[42] Card, S. K., Mackinlay, J. D., & Shneiderman, B. (Eds.). (1999). Readings in information visualization: using vision to think. Morgan Kaufmann.

[43] Chen, C. (2004). Information visualization: beyond the horizon. Springer Science & Business Media.

[44] Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., ... & Tominski, C. (2010). Space, time and visual analytics. International Journal of Geographical Information Science, 24(10), 1577-1600.

[45] Thomas, J. J., & Cook, K. A. (2005). Illuminating the path: The research and development agenda for visual analytics. IEEE Computer Society Press.