Spatiotemporal Visualization Literature Review

Claire Pang

September 28, 2018

Abstract — Spatiotemporal data is data involving location, time, and another value. This type of data exists as social media data, cell phone data, traffic data, medical data, and in other domains. To extract insights, visual analytics, topic modeling, anomaly detection, and other techniques are performed on spatiotemporal data. These findings are then visualized using maps, word clouds, heatmaps, and other techniques. This paper discusses the applications of spatiotemporal data in various domains, related visualization techniques, and future areas of research in this field.

1 Introduction

Data can be considered one of the most valuable resources for its ability, when processed, to answer research questions or to be used in predictive capacities. One particular type of data, spatiotemporal data, is essential in providing knowledge through visual analysis. Specifically, spatiotemporal visualization involves visualizing data containing location, time, and another value. The amount of spatiotemporal data has increased with the rise of social media over the years, but is not limited to just social media – spatiotemporal data also exists as cell phone data [1], traffic data [2, 3, 4], medical data [5, 6], and in other domains. As spatiotemporal data has to do with location, most visualizations involve geographic heatmaps, but may also involve other techniques such as word clouds, bar charts, or a combination of the aforementioned.

Often, spatiotemporal visualization is used as a form of visual analysis. Keim et al. describes visual analytics as "an integral approach to decision-making, combining visualization, human factors and data analysis" [7]. By simultaneously using visualizations, human cognition, and perception, pattern analysis can easily be performed. Pattern analysis can then be extended to anomaly detection, which is a technique used to identify abnormal events. These are events that deviate from the expected pattern/behavior. Areas where these abnormal events are located are called "hotspots." The most common use for spatiotemporal data is anomaly detection, but other applications, such as urban and traffic analysis, will be discussed in this paper.

This paper discusses the various applications of spatiotemporal data and related visualization techniques. The type of machine learning models and techniques that have been applied will be discussed, but there will be no evaluation of the best algorithms. This paper will also touch upon existing research in data storage, management, and querying of spatiotemporal datasets specifically, but other topics related to data processing (such as data cleaning) will not be covered.

2 Existing research

2.1 Resources

Much of the research in spatiotemporal visualization can be found in prestigious academic journals or conference papers, such as the Transactions on Visualization and Computer Graphics (TVCG), Visual Analytics Science and Technology (VAST) Conference, and in the Proceedings of the CHI Conference on Human Factors in Computing Systems. Of existing research in spatiotemporal visualization, many papers fall into the following categories: topic modeling, anomaly detection, traffic analysis, urban analysis, storage/data management, and other applications. Overlaps between categories are not uncommon, with research frequently being applied to datasets from multiple domains.

2.2 Topic modeling and anomaly detection

2.2.1 Topic modeling

For many research papers, the dataset used consists of tweets from Twitter. As tweets contain text, text mining is often applied to tweets to categorize content and/or identify frequency of topics. This type of text mining is referred to as "topic modeling," which is a machine learning technique used to perform text classification by deriving phrases/keywords related to a specific topic. There are many algorithms for performing topic modeling. One example is Andrienko et al. [8], who performed term usage clustering on a dataset of Seattle tweets to extract general topics. After that, keyword-based categorization was performed. With the goal of discovering thematic tweeting behavior by spatial region, the keyword-based categorization allows for more extensive analysis of what is tweeted about, when, and at what frequency. The patterns by topic are then visualized, using symbol maps and word clouds.

2.2.2 Anomaly detection with topic modeling

Building upon steps taken by Andrienko et al. [8], most research papers also apply anomaly detection after deriving the topics. Many algorithms for anomaly detection exist, which is extensively discussed by Chandola et al [9]. In terms of tweets, anomaly detection would be able to identify unusual events

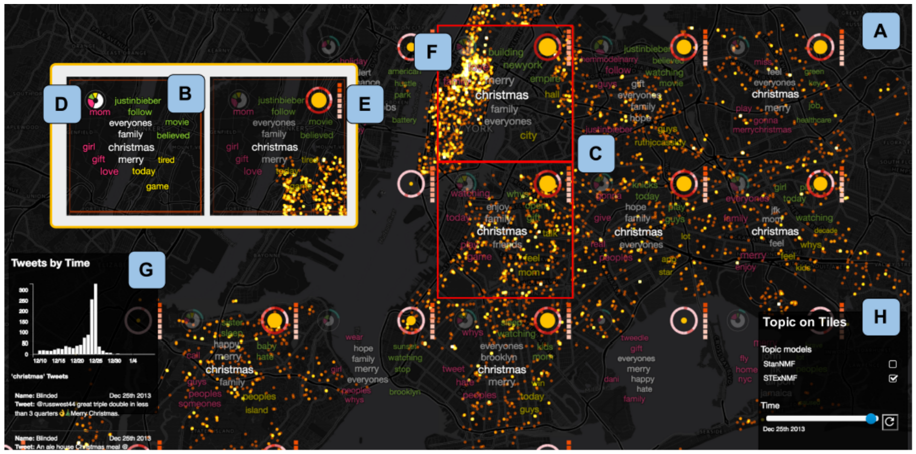

Figure 1: Interface from TopicOnTiles [12]: The topics derived from topic modeling are displayed in a word cloud and a heatmap of the tweets. Red borders indicate that that tile has a high anomaly score.

happening at that time. Another method for topic modeling is LDA (Latent Dirichlet Allocation), as performed by Chae et al. [10]. After deriving the topics using LDA, a seasonal-trend decomposition technique known as STL [11] (seasonal trend decomposition using LOESS, where LOESS is form of regression) is applied to find anomalies. Their visual interface allows for exploring the topics derived from LDA, and visualizing the degree of abnormality.

Choi et al. [12] also used tweets to develop TopicOnTiles, a system that visualizes keywords derived from topic modeling using STExNMF (geospatio-temporally exclusive nonnegative matrix factorization) [13]. Choi et al. divides the spatial region of the tweets into "tiles," which allows STExNMF to find topics that are exclusive to a specific tile, relative to the neighboring tiles. This unique visualization style overlays the topic keywords in tiles, including a heatmap showing the locations of tweets containing those topics. Anomaly scores for each tile are also calculated, with visual cues (red bordered tiles) indicating anomalies (Figure 1).

Another example of social media data usage is with Instagram, Foursquare, and Twitter data in CityBeat [14]. Xia et al. uses data (images and captions from Instagram, topics from Twitter API, and check-ins from Foursquare) from all three applications to build a dashboard identifying hyper-local (restricted to a small region) events, such as concerts or emergencies. This system incorporates 6 algorithms in their event detection system in order to identify the events and generate the relevant data in the visualization.

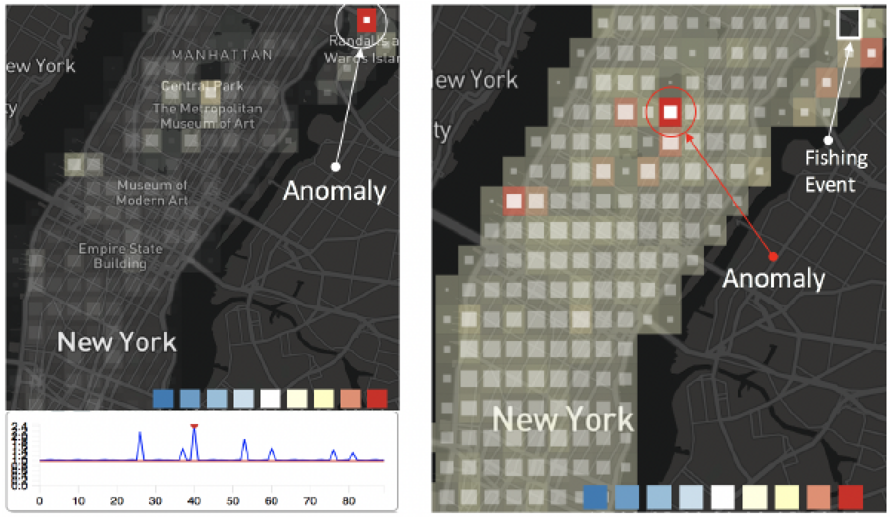

Figure 2: VOILA System [15]. **Left:** Anomaly detection identifying the fishing event. **Right:** Re-updating of anomaly scores and identification of new anomalies after the fishing event is marked as normal.

2.2.3 Anomaly detection through other methods

Anomaly detection is not limited to just social media data, however. In a "smart city" example, Cao et al. [15] developed a system called VOILA (Visual analysis of spatiotemporal data) that combines a tensor-based algorithm and user feedback (using a Bayesian updating rule) for anomaly detection. In addition, this system has two modes: anomaly detection mode and context mode showing expected patterns, with the purpose of using the system as a visual analytics tool. A case study includes using traffic data as a public safety monitoring tool. For the system-identified anomalies, user feedback was incorporated by identifying an anomaly (such as a fishing event) as normal, hence re-updating anomaly scores and then finding other anomalies (Figure 2).

Maciejewski et al. [6] has also developed a system for visual analytics exploration using aggregated heatmaps. An optional contour overlay (Figure 4) also allows for multivariate exploration. The heatmap and contours are aggregated using KDE (kernel density estimation) to ensure privacy of medical data. Using a CUSUM (cumulative summation) algorithm on a time series, alerts are given for potential anomalies (Figure 3). This multivariate view not only allows for searching for hotspots, but also determining correlation among multiple variables. They discuss applications using the multivariate heatmap as a visual analysis tool for examining correlation between medical cases (Figure 4 - left) or analyzing thefts against all crimes (Figure 4 - right).

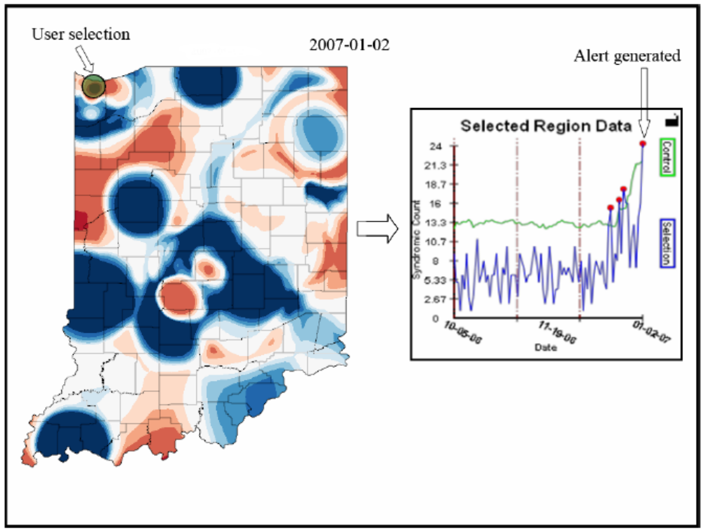

Figure 3: Heatmap of ILI (influenza-like-illness) during flu season. Selection of an area on the heatmap and the related time series graph, with alerts of potential anomalies as red circles [6].

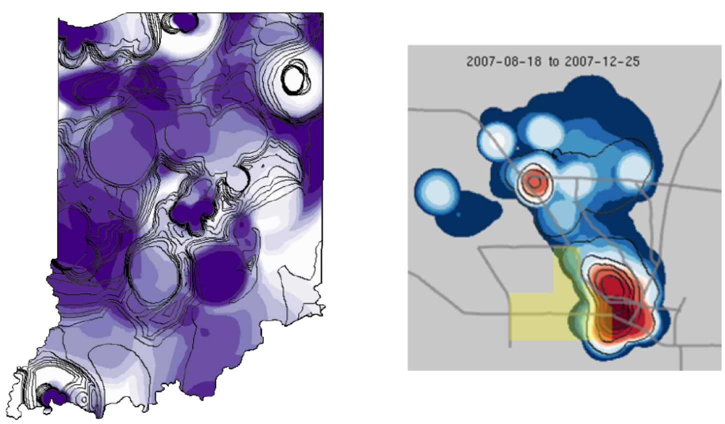

Figure 4: Aggregated heatmaps with contour [6]. This multivariate view allows searching for correlation. **Left**: Shock cases (color) and rash cases (contours) in Indiana. **Right**: All crimes in West Lafayette, Indiana from the 2007-2008 Purdue University school year (color) and 2007 Fall semester thefts (contour).

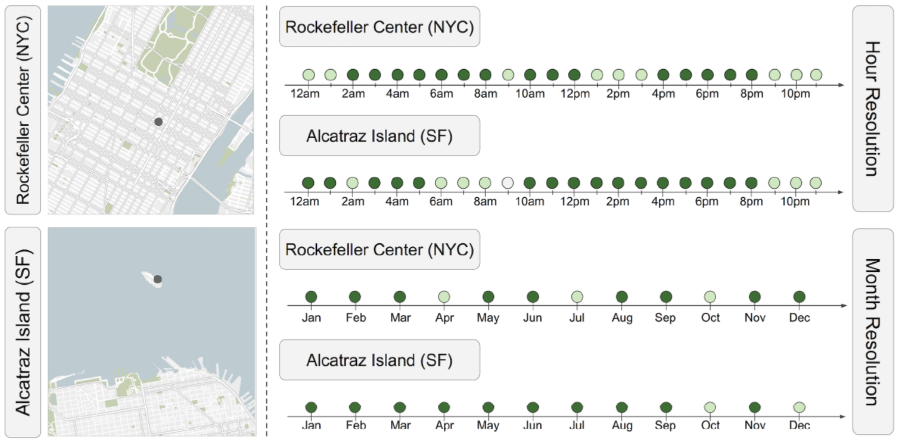

Figure 5: Urban Pulse [17]. Comparing the similar pulses of Rockefeller Center (New York City, New York) and Alcatraz (San Francisco, California). The pulse, denoted through using a gradient of green (high: dark green, low: light green), indicates activity level relative to neighboring locations.

2.3 Urban analysis

Besides anomaly detection, urban analysis is another area that can provide valuable insights. With plentiful amounts of data in this area, urban analysis can tell us about cities. This may include what spots are popular at what times, movement patterns, or general behaviors of the citizens. Understanding those aspects allows cities to be more sustainable and efficient, especially in tasks such as city planning, zoning, improving public transportation, or other services.

Andrienko et al. [16] uses Flickr data and analyzes the seasonal patterns of visits to popular places. This is visualized by using growth rings on a map. Andrienko et al. also examines the flows between visited places by clustering trajectories.

Similar to Andrienko et al. [16], Miranda et al.'s system [17] also uses Flickr data. Miranda et al.'s system allows users to explore the "pulses" of a city, which represent the activity level at a given point in time in relation to the dataset being used and neighboring locations. In doing so, two case studies are presented. The first case study aims to compare the pulses of multiple cities and understand why those locations are popular. The second case study looks at the differing behaviors of the cultural communities in New York based on the locations visited. The visualization of this system includes a map, but also timelines indicating pulses at different times (Figure 5).

Alternatively, another interest in urban analysis may be understanding movement patterns. Chen et al. [18] uses Weibo (a Chinese social media platform) data to analyze movement patterns in Taiwan and mainland China. Because the data used was sparse, the analysis was aided by an uncertainty

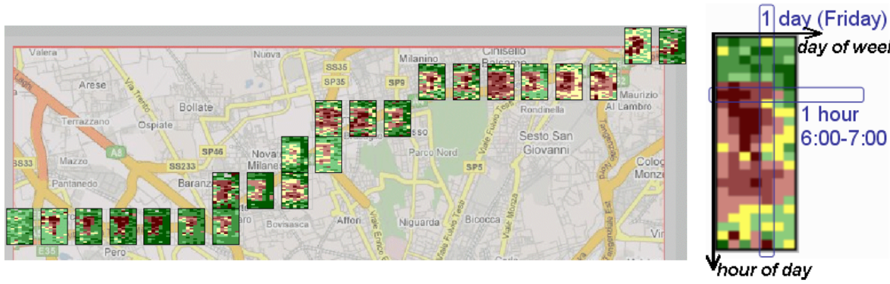

Figure 6: Aggregation from [4] where each heatmap represents median car speed in a spatial area. This particular figure shows aggregation based on a certain road.

model which provided more reliable information. An interface was developed to explore popular areas and movements around Taiwan.

2.3.1 Traffic analysis

With increased urbanization of cities, traffic management is often a major issue cities must deal with. Traffic data such as taxi data, cell phone data, transit data, and road monitoring data can all help cities understand how to better understand the traffic patterns and plan accordingly. Chen et al. [19] covers the many tasks for traffic visualization, including the different visualization techniques currently being employed. There are multiple interfaces developed for analyzing traffic data, such as ones focusing on studying traffic patterns [3] or analyzing traffic jams [2].

Other research in traffic includes AllAboard, created by Lorenzo et al [1]. AllAboard uses cell phone data from Abidjan, Ivory Coast, to understand demand for public transportation. Using a dataset containing users' SMS and call data with antenna location over 5 months, the system performs data mining to generate transit route recommendations. This is available in a user interface which displays travel flows and patterns, and allows analysts to tests optimization by altering the transit network.

One of the problems with large datasets is how to visualize the data without information overload, or needing to load the entire dataset. Techniques for visualizing big movement datasets has been studied by Andrienko and Andrienko [4], who recommend aggregation. By using a dataset of one week's worth of car movement data in Milan, Italy, aggregation is done by placing multiple heatmaps representing a certain spatial area over a map. The data in the heatmap is aggregated by displaying average car speed (Figure 6). Other aggregation techniques are also demonstrated in this paper, such as aggregation by compass direction and clustering by trajectory.

2.4 Data management and storage strategies

Given that spatiotemporal datasets are often massive, research has also been done in the most efficient way to store and query for the purposes of visualization. Additionally, spatiotemporal data is often unstructured – it typically involves trajectories or location data mapped to both value and time, where values may be multidimensional.

Chen et al. [20] developed VAUD (Visual Analyzer for Urban Data), a system focused on efficient querying of spatiotemporal data from different sources. VAUD utilizes space-time-cubes in order to store the urban dataset. Ferreira et al. [21] also focuses their efforts on spatiotemporal querying through their New York taxi dataset of over 520 million taxi trips. The large dataset is managed by using a k-d tree.

Another approach by researchers on data management is with nanocubes. Lins et al. [22] developed a nanocube, which is a data cube designed to be stored on a laptop. It allowed for more efficient querying, but still required more memory than desired.

Another focus of these papers is visual querying [20, 21], or allowing users to view information specific to their interests/desired area of exploration. These systems often incorporate a graphical interface that does not require knowledge of any database languages. This helps prevent information overload but also is more efficient in the visual analytics process by being able to view and compare select information, or view the results of 2 queries side by side.

2.5 Other applications

There are many other applications for spatiotemporal datasets which do not fit in the above categories. Bryan et al. [5] describes a system that builds emulators/predictive models for simulating the spread of disease. The focus of this research is to test the simultaneous use of visual analytics with predictive models. Maciejewski et al. [6] also presents a case study that deals with the spread of disease, or "syndromic surveillance." Through their heatmap visualizations (Figure 4), users can search by syndrome for any hotspots and determine if an outbreak is occurring.

Another example of what can be done with Twitter data is Whisper, created by Cao et al. [23]. Whisper traces the spread of information over Twitter of major events. Using a sunflower metaphor (Figure 7) as the design of their visualization, Whisper is able to track the spread by time, location, media outlet, and sentiment.

Lastly, another application of spatiotemporal visualization is for learning purposes. Sifakis et al. [24] present ViSTPro, which combines Google Earth and other animations to describe the flow of an event visually, and with annotations on the side. One of the projects implemented using ViSTPro is to demonstrate the Battle of Crete by overlaying symbols at their respective locations on the map (Figure 8). By visually presenting the information with related locations, it helps students to better understand historical events.

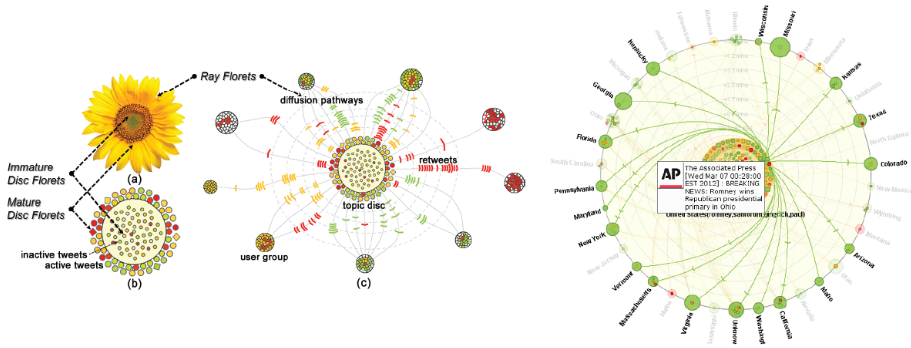

Figure 7: **Left**: the structure of Whisper [23], a visualization modeled after sunflowers to trace the dispersion of information on Twitter. **Right**: Example visualization showing information dispersion by the Associated Press of the Republican presidential primaries.

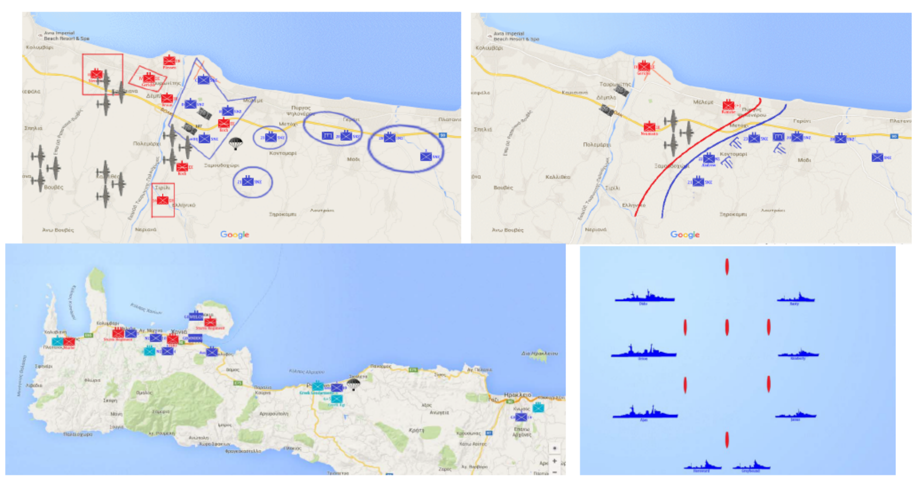

Figure 8: ViSTPro [24], a platform utilizing Google Earth to visualize an event spatiotemporally for learning purposes. This demonstration is of the Battle of Crete.

3 Future Research Topics

Some of the current challenges in spatiotemporal visualization are systems that support big data, real-time data, or data from multiple sources. Additionally, as seen in previous work [20, 21, 22] there is the need for more efficiently storing data so that queries can be even faster. Much of the research discussed in this paper uses datasets that are pre-loaded and cleaned, but do not necessarily support real-time data. There is tremendous value if systems can be created to be used in real-time, especially for anomaly detection. This would allow these systems to be used for real-time surveillance, in cases such as security monitoring or traffic monitoring. It would also be useful for systems to support data from multiple sources for algorithm improvement. Such systems might incorporate image/video processing as well.

As with all algorithms, research should also be done in improving the algorithms used for anomaly detection or predictive modeling. Some research incorporated user feedback into their anomaly detection systems [15], but more should rely upon human visual analytics and allow the user's input in calculating anomaly scores. As a way to improve visual analytics, more guidance should be given to the user. If in real-time monitoring, alerts should be automatically created instead of having the user search for anomalies. To those untrained in statistics, explaining related metrics would also be helpful.

The methods for displaying data can always be improved, as aggregation [4, 6] may be sometimes necessary to visualize large datasets or for privacy reasons. However, different methods for aggregation can be developed. In addition, how to handle multivariate data on heatmaps can be researched, as we have seen only contouring overlay [6] or tiled word clouds [12] so far.

Spatiotemporal visualization should also be applied to other domains, as most of the focus has been on social media, traffic, and urban analysis. It would be beneficial to see the same analysis applied to data from other domains as the value spatiotemporal analysis can provide is immense. Such domains might include animal migration, environmental science, neuroscience, or agriculture.

4 Conclusion

This paper has discussed spatiotemporal visualization by describing sources of spatiotemporal data, where research in spatiotemporal visualization is mostly found, and the currently existing research. Most research can be categorized into topic modeling, anomaly detection, urban analysis, traffic analysis, or data management/querying, but can also belong to another domain entirely. As a next step, more research should be done in different domains due to the valuable knowledge that can be extracted, as seen from the previous research. Scalability, algorithms for prediction or anomaly detection, and new visualization techniques are additional areas in which further research should be pursued.

References

[1] G. Di Lorenzo, M. Sbodio, F. Calabrese, M. Berlingerio, F. Pinelli, and R. Nair, "Allaboard: Visual exploration of cellphone mobility data to optimise public transport," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 2, pp. 1036–1050, 2016.

[2] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2159–2168, 2013.

[3] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan, "Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection," in *2011 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 163–170, IEEE, 2011.

[4] G. Andrienko and N. Andrienko, "Spatio-temporal aggregation for visual analysis of movements," in *2008 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 51–58, IEEE, 2008.

[5] C. Bryan, X. Wu, S. Mniszewski, and K.-L. Ma, "Integrating predictive analytics into a spatiotemporal epidemic simulation," in *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 17–24, IEEE, 2015.

[6] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert, "A visual analytics approach to understanding spatiotemporal hotspots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 205–220, 2010.

[7] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," in *Information Visualization*, pp. 154–175, Springer, 2008.

[8] G. Andrienko, N. Andrienko, H. Bosch, T. Ertl, G. Fuchs, P. Jankowski, and D. Thom, "Thematic patterns in georeferenced tweets through space-time visual analytics," *Computing in Science & Engineering*, vol. 15, no. 3, pp. 72–82, 2013.

[9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[10] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl, "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 143–152, IEEE, 2012.

[11] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “Stl: A seasonal-trend decomposition procedure based on loess,” *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[12] M. Choi, S. Shin, J. Choi, S. Langevin, C. Bethune, P. Horne, N. Kronenfeld, R. Kannan, B. Drake, H. Park, and J. Choo, “Topicontiles: Tile-based spatio-temporal event analytics via exclusive topic modeling on social media,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 583, ACM, 2018.

[13] D. S. Shin, M. Choi, J. Choi, S. Langevin, C. Bethune, P. Horne, N. Kronenfeld, R. Kannan, B. Drake, H. Park, and J. Choo, “STExNMF: Spatio-temporally exclusive topic discovery for anomalous event detection,” in *2017 IEEE Conference on Data Mining (ICDM)*, pp. 435–444, IEEE, 2017.

[14] C. Xia, R. Schwartz, K. Xie, A. Krebs, A. Langdon, J. Ting, and M. Naaman, “Citybeat: real-time social media visualization of hyper-local city data,” in *Proceedings of the 23rd International Conference on World Wide Web*, pp. 167–170, ACM, 2014.

[15] N. Cao, C. Lin, Q. Zhu, Y.-R. Lin, X. Teng, and X. Wen, “Voila: Visual anomaly detection and monitoring with streaming spatiotemporal data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 23–33, 2018.

[16] G. Andrienko, N. Andrienko, P. Bak, S. Kisilevich, and D. Keim, “Analysis of community-contributed space-and time-referenced data (example of panoramio photos),” in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 540–541, ACM, 2009.

[17] F. Miranda, H. Doraiswamy, M. Lage, K. Zhao, B. Gonçalves, L. Wilson, M. Hsieh, and C. T. Silva, “Urban pulse: Capturing the rhythm of cities,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 791–800, 2017.

[18] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. L. Zhang, and J. Zhang, “Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 270–279, 2016.

[19] W. Chen, F. Guo, and F.-Y. Wang, “A survey of traffic data visualization,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 2970–2984, 2015.

[20] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan, and R. Maciejewski, “Vaud: A visual analysis approach for exploring spatio-temporal urban data,” *IEEE Transactions on Visualization and Computer Graphics*, no. 9, pp. 2636–2648, 2018.

[21] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2149–2158, 2013.

[22] L. Lins, J. T. Klosowski, and C. Scheidegger, "Nanocubes for real-time exploration of spatiotemporal datasets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2456–2465, 2013.

[23] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu, "Whisper: Tracing the spatiotemporal process of information diffusion in real time," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2649–2658, 2012.

[24] Y. Sifakis, S. Christodoulakis, and N. Moumoutzis, "Vistpro: A platform for visualization of spatiotemporal processes on google earth," in *2016 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, pp. 117–122, IEEE, 2016.