

INVITED ARTICLE

Spatio-temporal visual analytics: a vision for 2020s

Natalia Andrienko and Gennady Andrienko

Fraunhofer Institute IAIS, Sankt Augustin, Germany

City, University of London, UK

Received: February 29, 2020; accepted: April 17, 2020

Abstract: Visual analytics is a research discipline that is based on acknowledging the power and the necessity of the human vision, understanding, and reasoning in data analysis and problem solving. Visual analytics develops methods, analytical workflows, and software tools for analysing data of various types, particularly, spatio-temporal data, which can describe the processes going on in the environment, society, and economy. We briefly overview the achievements of the visual analytics research concerning spatio-temporal data analysis and discuss the major open problems.

Keywords: analytical reasoning, visualisation literacy, visual analytics, big data, knowledge transfer

1 Introduction

Whatever algorithms and technologies for computerized data processing appear, human understanding and reasoning remains the principal and irreplaceable instrument of analysis, modelling, and problem solving. Visual representation of information is acknowledged as the most effective way of supplying information to the human's mind and as a promoter of ideation and analytical reasoning. Visual analytics is a research discipline developing methods, analytical workflows, and software systems that can support unique capabilities of humans by providing appropriate visual displays of relevant information and involving as much as possible the capabilities of computers to store, process, analyse, and visualize data [18].

Understanding of the processes going on in the environment, society, and economy is crucial for the survival of the human civilization. All these are spatio-temporal phenomena;

hence, there is a high demand for approaches to supporting humans in analysis of spatio-temporal data [2]. Given the importance and complexity of such data, they have always been one of the major focuses in the visual analytics research.

2 State of the art

Particularly, much research has been done in the last decade on analysis of *movement data* [1] (i.e., sequences of spatial positions of moving entities), but substantial attention has also been given to *spatial events* (i.e., entities positioned in space and time) and *spatial time series* (i.e., temporal variation of spatially distributed attribute values). Visual analytics researchers proposed generalizable and reproducible analytical workflows involving data selection and filtering, visualization of movement trajectories, aggregate flows, and spatio-temporal distributions of events and attribute values, clustering based on the spatial, temporal, and thematic components of the data, transformations between spatio-temporal data types, and derivation of new data objects representing extracted pieces of task-relevant information. For example, spatial event data can be generated as a result of interactive detection of significant changes, anomalies, or critical circumstances in movement [4] or in time series of spatial situations [5]. Conceptual models have been proposed to address the diversity of spatio-temporal data and systematically consider the possible transformations between data types [1].

Before performing any analysis, it is necessary to ensure that the data are of appropriate *coverage* and *quality*. While there exist quite elaborated tools for inspecting properties of tabular data (e.g., Trifacta data wrangling tools¹), they are not yet addressing the specifics of spatial and temporal data such as the time cycles, spatial and temporal autocorrelation, smoothness of many phenomena in space and time, existence of geographic borders and barriers, just to name a few. Researchers are developing recommendations on how to deal with time series data [14, 15] and movement data [3], as well as other types of spatio-temporal data [22]. These recommendations gradually find their ways to open-source implementations [13].

One of the recent developments in visual analytics that may have potentially wide applicability and appreciable utility for spatio-temporal data analysis is a novel method of *time-based data querying and filtering* called Time Mask [6]. Based on any kind of time-related data, it selects all time intervals in which particular conditions of interest are fulfilled. Next, from all currently available temporal data, it selects the portions where the temporal references fit in the selected intervals. The Time Mask filter can serve as a tool for integrated analysis of several time-related phenomena: analysts can select time intervals based on the state of one of them and investigate how the other phenomena developed during these intervals.

The idea and the work of the Time Mask query method is illustrated in Fig. 1. The example dataset consists of events of taking photos of cherry blossoms that occurred on the territory of the USA during 10 years from 2007 till 2016. The data (metadata of the photos) was obtained from the Flickr photo sharing web service using its API. The whole set of events is shown in a space-time cube in the upper part of Fig. 1. By means of density-based clustering, large spatio-temporal clusters of the events were detected. These clusters signify time periods of mass photo taking in particular areas. The central image in Fig. 1

¹<https://www.trifacta.com/>

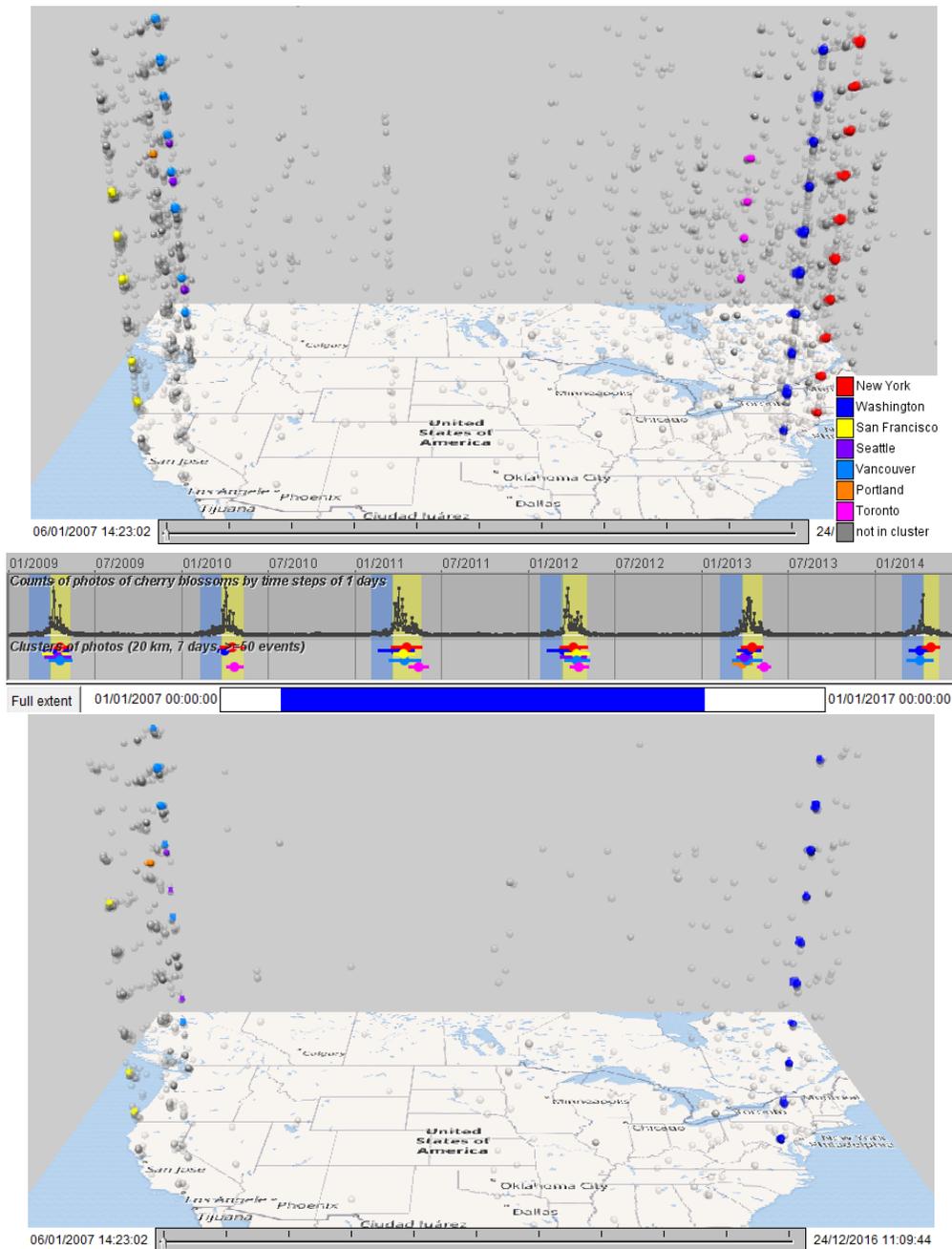


Figure 1: Illustration of the spatio-temporal filtering by Time Mask. Top: All photo taking events are shown in a space-time cube. Middle: A Time Mask query tool selects the 45 days long time intervals before the start times of the large spatio-temporal clusters of the events occurring in New York. Bottom: The space-time cube includes only the events satisfying the query.

demonstrates an interactive visual tool for making Time Mask queries. The user interface (UI) shows the distribution of chosen datasets or subsets over time. In our example, the upper part of the UI shows the time series of the daily counts of the individual events, and the lower part shows the times and durations of the event clusters. In this visual interface, the user first selected the time intervals when mass photo taking was happening in New York. These intervals are highlighted by background painting in yellow. Then, the user selected the 45-days long time intervals before the beginnings of the event clusters in New York. These intervals are highlighted by background painting in blue. In response to this query, the data (individual events and clusters) has been filtered. Only the data items whose time references fit in the time intervals marked in blue have been selected. This data are shown in the space-time cube at the bottom of Fig. 1. This example demonstrates how a Time Mask query can select data from multiple disjoint time intervals, thus enabling flexible data selection and sophisticated analysis workflows.

A notably large amount of visual analytics research was done on analysis of data from social media [10, 12, 30, 33], in particular, geographically referenced data. Luo and MacEachren [24] developed conceptual and methodological foundations for geo-social visual analytics. To support studies of people's reactions to significant events, visualizations and analytical workflows were proposed for integrated analysis of the *temporal, spatial, social, and thematic facets* of the social media data [20, 27–29]. There was also research on visualization and analysis of people's movements and long-term mobility behaviors derived from geo-tagged social media data [9, 31]. In this respect, attention was given to development of privacy-preserving approaches to visual analysis [7]. The research on analysing social media data contributes to better understanding of the society and people's life.

Visual analytics research topics also include support to development of *predictive models* [23]. In relation to spatio-temporal phenomena, approaches have been proposed for forecasting hotspots of an epidemic outbreak [25], modelling of road traffic flows under usual and extraordinary conditions [8], prediction of crime incident distribution and decision support for resource allocation [26], to name a few. On the one hand, these approaches enable human analysts to incorporate their domain knowledge and insights gained from the analysis in the model development, on the other hand, they help analysts to use the models for improving their understanding of the phenomena at hand, verifying their insights and hypotheses, and developing appropriate decision options.

A strong current trend in visual analytics research is development of machine learning models with understandable and *explainable* behavior (often called as “eXplainable AI”, or xAI) [11]; however, modelling of spatio-temporal phenomena has not been addressed in this research, yet [16].

It has become usual nowadays to study and predict the behavior of complex real-world phenomena by means of computer *simulation models*. Since the latter are affected by configuration parameters, they are usually executed multiple times with different parameter settings to generate ensembles of simulated behaviors. Supporting scientists to explore and comprehend ensemble simulation data is a prominent research topic in visual analytics [32]. It is important not only to portray complex behaviors but also help the analysts to assess the uncertainties arising due to divergences between different predictions and to investigate the variations of the uncertainty over space and time [17, 19].

To summarize the state of the art in space-time visual analytics, we would like to highlight the following trends:

- the advanced state of conceptualization of spatio-temporal data and emerging systematic approaches to data analysis;
- detailed consideration of data properties and semantic interpretation of data;
- active development of methods and tools for task-oriented data transformation and analysis; and
- bridging the gaps between data exploration, analysis, and model building.

3 Open problems and possible approaches

Current research in visual analytics strives to address the challenges of *big data*. Apart from developing technical solutions for enabling rapid visual representation of and fluid interaction with large amounts of dynamically changing complex data, visual analytics must care about the natural limitations in the human capability to comprehend information regarding its amount, complexity, and rate of change. Moreover, since the human is not a passive recipient of information but an important active force in analysis and problem solving, visual analytics must also care about the limited speed of human cognitive processes. The involvement of the “human factor” makes the challenges of big data especially hard for visual analytics. The key idea for addressing these challenges is *effective division of labor between the human and the computer*. Thus, computational techniques can be utilized for *partitioning* very large spatio-temporal datasets into internally homogeneous coherent portions and extracting characteristic *patterns* from these portions, and analysts can employ visual and interactive techniques to understand the patterns and their distribution over space and time [21]. This approach, however, only deals with the problem of data volume but does not address the problem of data dynamism.

In dealing with the latter problem, visual analytics researchers can build on the experiences of successful application of visual analytics to model building. In the case of dynamically changing data, the basic approach is to create an initial model based on currently available data, continuously monitor the model appropriateness to the newly arriving data, and modify the model when it fails to accommodate the new data well enough. It may not be feasible to involve a human expert each time when the model requires adaptation. However, an expert may anticipate how the data may change, based on the previous history of the changes, and orchestrate suitable model adaptation mechanisms that can be activated automatically when the prediction fulfills. This can be done for several possible courses of data evolution. A specific task for visual analytics is to support the process of the expert’s analytical reasoning based on the previous history of data changes and model adaptations. While it may not be possible to develop a general solution, a multitude of application- and task-oriented approaches to handling the problems of big data can be expected.

Another major challenge for visual analytics is achieving wide whilst appropriate and effective utilization of visual analytics techniques and workflows by analysis practitioners. A vast amount of research in visualization and visual analytics is concerned with development of user-oriented tools and systems, personalization, automated identification of users’ needs and intentions, as well as user guidance. However, it is not very likely that the research prototypes developed along these directions will soon be converted into widely accessible and highly reliable visual analytics software suitable for a large variety of applications and analysis tasks. Some commercially available systems are quite user-friendly, but their analytical capabilities are very limited, especially with respect to spatio-temporal

data. A practicable solution is proliferation of open-access libraries of combinable software codes implementing various methods for data processing, analysis, modelling, and visualization. This trend is now actively developing, being stimulated by the emergence of the data science languages *R* and *Python* and the interactive web-based computational notebook environments.

A large number and variety of notebooks with examples of application of different data analysis and visualization methods have been published on the web by numerous people. These notebooks can be relatively easily adapted to one's individual needs. As a result, analytical functionality becomes accessible to and usable by nearly everyone. This generally positive trend has its back side, however. The notebooks are often created or adapted by people having quite little idea of how to choose appropriate visualization techniques and design correct and effective visualizations of the data they deal with, and also have no good understanding of why, when, and how visualizations need to be used in analysis and what is their right place in analysis workflows. Some visualizations occurring in the publicly accessible example notebooks may look impressive and convincing to non-specialists, but, in fact, they may communicate spurious patterns in inadequate ways. Those who view these visualizations and think of doing the same for their data and tasks often lack knowledge that would enable critical assessment and understanding of the suitability of the techniques. Other notebooks include only basic graphics having little analytical value, whereas better ways exist for representing the relevant information. Furthermore, often notebooks apply computational methods without checking pre-requisites (e.g., data properties and quality) and investigation if the parameter settings are suitable and results are meaningful.

Besides insufficient visualization and data analysis *literacy*, there is a danger of uncritical trust in what is produced by computers and taking the outcome of a single run of an analysis algorithm with default parameter settings, or with settings previously used by someone else, as the final result. Naive analysts may not realise that a slight change in the data or parameters can sometimes significantly change the result; therefore, they may not bother to examine the reaction of the algorithm to such changes and to check results of several runs for consistency. More experienced and critically-minded analysts, who usually take the trouble to evaluate and compare what they get from computers, may tend to rely solely on statistical measures rather than trying to gain better understanding with the help of visualizations.

Hence, there is a strong need in disseminating the knowledge on how to create meaningful visualizations and how to use them effectively in data analysis together with computer operations. It is also necessary to spread the philosophy of visual analytics, where the main principles are the primacy of human understanding and reasoning and awareness of the weaknesses of computers, which cannot see, understand, and think, and thus need to be led and controlled by humans.

This challenge can be addressed by creation of easily understandable, practitioner-oriented textbooks and open online courses. Visual analytics researchers should not only strive for advancing the research but also take the responsibility for transferring the operational knowledge to practitioners and casual analysts. This is especially important for achieving wider and better understanding of the environmental, economic, and societal processes, their interrelations and effects.

Acknowledgments

This research was supported by the German Priority Research Program SPP 1894 on Volunteered Geographic Information, EU projects SoBigData++ and Track&Know, EU SESAR project TAPAS, and Fraunhofer Cluster of Excellence on “Cognitive Internet Technologies”.

The dataset of photo taking events was acquired and kindly provided to us by Alexander Dunkel (Institute of Cartography, TU Dresden, Germany).

References

- [1] ANDRIENKO, G., ANDRIENKO, N., BAK, P., KEIM, D., AND WROBEL, S. *Visual Analytics of Movement*. Springer Publishing Company, Incorporated, 2013.
- [2] ANDRIENKO, G., ANDRIENKO, N., DEMSAR, U., DRANSCH, D., DYKES, J., FABRIKANT, S. I., JERN, M., KRAAK, M.-J., SCHUMANN, H., AND TOMINSKI, C. Space, time and visual analytics. *International Journal of Geographical Information Science* 24, 10 (2010), 1577–1600. doi:10.1080/13658816.2010.508043.
- [3] ANDRIENKO, G., ANDRIENKO, N., AND FUCHS, G. Understanding movement data quality. *Journal of Location Based Services* 10, 1 (2016), 31–46. doi:10.1080/17489725.2016.1169322.
- [4] ANDRIENKO, G., ANDRIENKO, N., HURTER, C., RINZIVILLO, S., AND WROBEL, S. Scalable analysis of movement data for extracting and exploring significant places. *IEEE Transactions on Visualization and Computer Graphics* 19, 7 (2013), 1078–1094. doi:10.1109/TVCG.2012.311.
- [5] ANDRIENKO, G., ANDRIENKO, N., MLADENOV, M., MOCK, M., AND POLITZ, C. Identifying place histories from activity traces with an eye to parameter impact. *IEEE Transactions on Visualization and Computer Graphics* 18, 5 (2012), 675–688. doi:10.1109/TVCG.2011.153.
- [6] ANDRIENKO, N., ANDRIENKO, G., CAMOSSO, E., CLARAMUNT, C., GARCIA, J. M. C., FUCHS, G., HADZAGIC, M., JOUSSELME, A.-L., RAY, C., SCARLATTI, D., AND VOUIROS, G. Visual exploration of movement and event data with interactive time masks. *Visual Informatics* 1, 1 (2017), 25–39. doi:10.1016/j.visinf.2017.01.004.
- [7] ANDRIENKO, N., ANDRIENKO, G., FUCHS, G., AND JANKOWSKI, P. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Information Visualization* 15, 2 (2016), 117–153. doi:10.1177/1473871615581216.
- [8] ANDRIENKO, N., ANDRIENKO, G., AND RINZIVILLO, S. Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics. *Information Systems* 57 (2016), 172–194. doi:10.1016/j.is.2015.08.007.
- [9] BLANFORD, J. I., HUANG, Z., SAVELYEV, A., AND MACEACHREN, A. M. Geo-located tweets. Enhancing mobility maps and capturing cross-border movement. *PLOS ONE* 10, 6 (2015), 1–16. doi:10.1371/journal.pone.0129202.

- [10] CHEN, S., LIN, L., AND YUAN, X. Social media visual analytics. *Computer Graphics Forum* 36 (2017), 563–587. doi:10.1111/cgf.13211.
- [11] CHOO, J., AND LIU, S. Visual analytics for explainable deep learning. *IEEE Computer Graphics and Applications* 38, 4 (2018), 84–92. doi:10.1109/MCG.2018.042731661.
- [12] DUNKEL, A., ANDRIENKO, G., ANDRIENKO, N., BURGHARDT, D., HAUTHAL, E., AND PURVES, R. A conceptual framework for studying collective reactions to events in location-based social media. *International Journal of Geographical Information Science* 33, 4 (2019), 780–804. doi:10.1080/13658816.2018.1546390.
- [13] GRASER, A. Movement data in GIS #26: towards a template for exploring movement data. <https://anitagraser.com/2020/01/03/movement-data-in-gis-26-towards-a-template-for-exploring-movement-data/>, 2020.
- [14] GSCHWANDTNER, T., AIGNER, W., MIKSCH, S., GÄRTNER, J., KRIGLSTEIN, S., POHL, M., AND SUCHY, N. TimeCleanser: A visual analytics approach for data cleansing of time-oriented data. In *Proceedings of the 14th International Conference on Knowledge Technologies and Data-Driven Business* (New York, NY, USA, 2014), Association for Computing Machinery. doi:10.1145/2637748.2638423.
- [15] GSCHWANDTNER, T., GÄRTNER, J., AIGNER, W., AND MIKSCH, S. A taxonomy of dirty time-oriented data. In *Multidisciplinary Research and Practice for Information Systems* (Berlin, Heidelberg, 2012), G. Quirchmayr, J. Basl, I. You, L. Xu, and E. Weippl, Eds., Springer Berlin Heidelberg, pp. 58–72.
- [16] GUIDOTTI, R., MONREALE, A., RUGGIERI, S., TURINI, F., GIANNOTTI, F., AND PEDRESCHI, D. A survey of methods for explaining black box models. *ACM Comput. Surv.* 51, 5 (2018). doi:10.1145/3236009.
- [17] KAPPE, C. P., BÖTTINGER, M., AND LEITTE, H. Exploring variability within ensembles of decadal climate predictions. *IEEE Transactions on Visualization and Computer Graphics* 25, 3 (2019), 1499–1512. doi:10.1109/TVCG.2018.2810919.
- [18] KEIM, D. A., KOHLHAMMER, J., ELLIS, G. P., AND MANSMANN, F. *Mastering the Information Age - Solving Problems with Visual Analytics*. Eurographics Association, 2010.
- [19] KUMPF, A., RAUTENHAUS, M., RIEMER, M., AND WESTERMANN, R. Visual analysis of the temporal evolution of ensemble forecast sensitivities. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 98–108. doi:10.1109/TVCG.2018.2864901.
- [20] LI, J., CHEN, S., CHEN, W., ANDRIENKO, G., AND ANDRIENKO, N. Semantics-space-time cube: a conceptual framework for systematic analysis of texts in space and time. *IEEE Transactions on Visualization and Computer Graphics PP* (2018), 1. doi:10.1109/TVCG.2018.2882449.
- [21] LIU, D., XU, P., AND REN, L. TPFlow: Progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 1–11. doi:10.1109/TVCG.2018.2865018.



- [22] LIU, S., ANDRIENKO, G., WU, Y., CAO, N., JIANG, L., SHI, C., WANG, Y.-S., AND HONG, S. Steering data quality with visual analytics: The complexity challenge. *Visual Informatics* 2, 4 (2018), 191–197. doi:10.1016/j.visinf.2018.12.001.
- [23] LU, Y., GARCIA, R., HANSEN, B., GLEICHER, M., AND MACIEJEWSKI, R. The state-of-the-art in predictive visual analytics. *Computer Graphics Forum* 36, 3 (2017), 539–562. doi:10.1111/cgf.13210.
- [24] LUO, W., AND MACEACHREN, A. Geo-social visual analytics. *Journal of Spatial Information Science* 8 (2013), 27–66. doi:10.5311/JOSIS.2014.8.139.
- [25] MACIEJEWSKI, R., HAFEN, R., RUDOLPH, S., LAREW, S. G., MITCHELL, M. A., CLEVELAND, W. S., AND EBERT, D. S. Forecasting hotspots—a predictive analytics approach. *IEEE Transactions on Visualization and Computer Graphics* 17, 4 (2011), 440–453. doi:10.1109/TVCG.2010.82.
- [26] MALIK, A., MACIEJEWSKI, R., TOWERS, S., MCCULLOUGH, S., AND EBERT, D. S. Proactive spatiotemporal resource allocation and predictive visual analytics for community policing and law enforcement. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1863–1872. doi:10.1109/TVCG.2014.2346926.
- [27] NELSON, J., QUINN, S., SWEDBERG, B., CHU, W., AND MACEACHREN, A. Geovisual analytics approach to exploring public political discourse on twitter. *ISPRS International Journal of Geo-Information* 4, 1 (2015), 337–366. doi:10.3390/ijgi4010337.
- [28] PEZANOWSKI, S., MACEACHREN, A. M., SAVELYEV, A., AND ROBINSON, A. C. Senseplace3: a geovisual framework to analyze place-time-attribute information in social media. *Cartography and Geographic Information Science* 45, 5 (2018), 420–437. doi:10.1080/15230406.2017.1370391.
- [29] SAVELYEV, A., AND MACEACHREN, A. M. Augmenting geovisual analytics of social media data with heterogeneous information network mining—cognitive plausibility assessment. *PLOS ONE* 13, 12 (2018), 1–27. doi:10.1371/journal.pone.0206906.
- [30] SCHRECK, T., AND KEIM, D. Visual analysis of social media data. *Computer* 46, 5 (2013), 68–75. doi:10.1109/MC.2012.430.
- [31] VON LANDESBERGER, T., BRODKORB, F., ROSKOSCH, P., ANDRIENKO, N., ANDRIENKO, G., AND KERREN, A. MobilityGraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 11–20. doi:10.1109/TVCG.2015.2468111.
- [32] WANG, J., HAZARIKA, S., LI, C., AND SHEN, H. Visualization and visual analysis of ensemble data: A survey. *IEEE Transactions on Visualization and Computer Graphics* 25, 9 (2019), 2853–2872. doi:10.1109/TVCG.2018.2853721.
- [33] WU, Y., CAO, N., GOTZ, D., TAN, Y.-P., AND KEIM, D. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia* 18 (2016), 2135–2148. doi:10.1109/TMM.2016.2614220.