EuroVis Workshop on Visual Analytics (2024)
M. El-Assady and H.-J. Schulz (Editors)

Exploring Relationships between Events in Context

N. Andrienko[1,2] and G. Andrienko[1,2]

[1]Fraunhofer Institute IAIS and Lamarr Institute for Machine Learning and Artificial Intelligence, Germany
[2]City University of London, UK

Abstract

We propose an approach to exploring interrelationships between two or more sequences of events when events occurring in one sequence both affect and are affected by events occurring in another sequences. We present the approach by example of exploring the dynamic relationships between COVID pandemic events and changes in population mobility behaviours across various countries. The key idea is to generate data capturing the temporal context of each event, i.e., what types of events occurred in different sequences within a specified time buffer around this event. An application of 2D space embedding to the context data reveals groups of events occurring in similar contexts. We can investigate the types of events each group consists of and see when and where these events and these contexts took place. By interactive or algorithmic clustering of the context data, we categorise event contexts based on their similarities, which allows us to compute, visualise, explore, and compare summary statistics of the context clusters, as well as exploring their distribution over time and other data dimensions.

CCS Concepts

• ***Human-centered computing*** → *Visual Analytics;*

1. Introduction

Understanding the dynamic interplay between events unfolding over time is essential for comprehending behaviours and evolution of complex systems, such as socio-economic trends and epidemiological outbreaks [AMST11, DAA*19]. Traditional approaches either aim to reveal recurrent patterns of event succession or strive to extract and model causal relationships between occurrences of different event types. However, these approaches often make the assumption of stationarity of event sequences, potentially overlooking the context-dependent and time-variant nature of sequential patterns and interrelationships between events. We argue that exploratory analysis must precede any attempts at computational extraction of patterns or causal links [AAF*20].

This paper presents a novel approach for exploring relationships between multiple event sequences. By leveraging event contexts, which encapsulate the temporal and situational surroundings of events, our methodology facilitates the elucidation of dynamic interrelationships and the observation of how they evolve over time and vary across different data dimensions. Through a detailed analysis of concurrent streams of COVID pandemic events and changes in population mobility behaviours, we demonstrate the utility of our approach in uncovering nuanced patterns of event succession and interplay. Furthermore, we outline potential future research directions, particularly in predictive modelling using event contexts to anticipate future events. Overall, our work provides a practical framework for understanding and leveraging event sequences in complex systems.

2. Problem statement

An event can be defined as a discrete entity or happening that exists or takes place within a certain time interval $[t_1, t_2]$ and is associated with a particular type or category. An event sequence can be defined as a series of observed occurrences of various types of events linked to the same dynamic phenomenon unfolding over time. The events occurring in a particular temporal order reflect the evolution or progression of the phenomenon through identifiable stages or manifestations.

The analysis goal we are concerned with is to understand the interplay of two or more phenomena as manifested through respective event sequences, such that each phenomenon may influence and be influenced by others. Our research objective is to develop a methodological framework that facilitates the discovery of interrelationships between events across different sequences and tracks the evolution of these interrelationships over time.

Here are examples of potential applications of the framework we aim to develop:

- Mutual impacts of pandemic events and population mobility behaviours.
- Interplay between raises and drops of stock prices and buying or selling behaviours of stock traders.
- Bidirectional influences of achievements of sports teams and decisions concerning changes in team composition or coaching strategies.

DOI: 10.2312/eurova.20241114

- Interactions between advertising campaigns and changes in consumers' purchasing trends.
- Reciprocal effects of research funding strategies and advances in different research areas.

In developing our framework, we used the example of the changes in population mobility behaviours in different countries during the unfolding of the COVID pandemic. We utilised the publicly available data sets of mobility trends and disease severity indicators available in the Google's COVID-19 Open Data Repository [Goo22]. The original data are multivariate time series, as described in [AAS23]. For each country, we transformed the time series into two event sequences: pandemic development events and occurrences of different mobility behaviour patterns.

3. Related work

The exploration and analysis of event sequences have received significant attention of visualisation researchers, particularly spurred by the pioneering research conducted at the B.Shneiderman's Lab [WGGP*11, WG12, MLL*13]. These works introduced approaches aimed at handling vast quantities of sequences and large variety of event categories. A survey on visual analysis of event sequence data [GGJ*21] outlines the key analysis tasks, including summarisation, prediction, recommendation, anomaly detection, comparison, and causality analysis. For comparison tasks, a dedicated survey has been conducted [vdvV23].

The survey [GGJ*21] identifies major analytical techniques, including pattern discovery, sequence inference, and sequence modeling. These techniques aim to describe or predict the progression *within* individual event sequences. Similarly, the existing visualisation and interaction techniques mostly focus on relationships (including causal) and patterns within sequences. The only kind of between-sequence relationship that has been sought is similarity of subsequences, which is crucial for summarisation and pattern discovery. In all works on visually supported analysis of multiple event sequences, there has been a common set of event categories that appear in these sequences. To the best of our knowledge, the task of relating sequences with event categories from distinct "vocabularies" has not been addressed so far in the visual analytics research.

The causality analysis task, which has been relatively underexplored in visual analytics research until recently [GGJ*21], now tends to gain prominence [JGC*20, XHW20, ZSZ*22]. Computational methods are employed to discover causal relationships between events and subsequently visualise the results, which have the form of directed graphs. The researchers are concerned with the problems of incorporating users' feedback [JGC*20], capturing changes of causal relationships over time [XHW20], and detecting combined effects of several events [ZSZ*22].

We contend that computational extraction of causal relationships should be preceded by exploratory analysis of the data to assess the stability of associations between events across different contexts and check whether intuitive expectations regarding event interrelationships hold true.

In our problem setting, we refrain from a making prior assumption about the existence of causal relationships between events of different sequences. We recognise that interplay between sequences may be complex and dynamic, evolving over time and varying across other data dimensions. For instance, the relationships between pandemic events and mobility behaviours may exhibit variability across different time periods and geographical regions. We develop an approach that allows an analyst to explore the variation of the relationships between sequences.

4. Example dataset

To illustrate our approach, we use an example of events derived from a subset of the Google mobility and health data encompassing 60 countries across Europe, Asia, and North America. The subset spans 77 weeks from February 17, 2020 (Monday) to June 27, 2021 (Sunday). The mobility data consist of daily time series representing deviations in visit frequencies to six categories of places, such as groceries, transit stations, work, and residential areas, relative to baseline values established before the onset of the COVID-19 pandemic. The health data comprise daily counts of new COVID-19 cases and COVID-related fatalities. Prior to analysis, we standardised the absolute counts to reflect values per 100,000 individuals for each country.

Given that the process of deriving events from multivariate time series falls outside the primary focus of this paper, we provide a concise overview of what we did. Initially, we applied smoothing techniques to the time series to mitigate the influence of weekly variations, which are irrelevant to our analysis. Subsequently, we partitioned the time series into episodes [AAS23] of 14 days length using a sliding window incrementally shifted by 7 days. This process yielded a total of 4062 episodes, excluding those with few or no data. We then computed descriptive statistics to capture the variation in attributes within each episode.

We employed clustering and dimensionality reduction techniques to group episodes based on similarities of their descriptors, separately for mobility and health data. These groups were subsequently labelled to delineate various types of mobility and COVID-19 events. The mobility event types are 'normal mobility', 'decreasing mobility', 'stay home', and 'increasing mobility', and COVID event types are 'almost no disease', 'low to medium morbidity', 'high morbidity', and 'high morbidity and mortality'. By merging consecutive episodes of the same type for each country, we identified a total of 481 COVID events and 530 mobility events of varying duration. The resulting events are shown in Figs. 1, 2.

In Fig. 1, the events are depicted in space-time cubes, where the horizontal plane represents the geographical space, and the vertical dimension reflects the passage of time. The time axis is oriented from bottom to top. The events manifest as vertical sticks painted in colours corresponding to their assigned event types (see the legends on top of Fig. 1). However, due to occlusions inherent in cube representations, we have devised an alternative method to visualise the distribution of events over time and across countries, as illustrated in Fig. 2. The display takes the form of a matrix, with rows corresponding to countries and columns to weekly time steps. The events are depicted as coloured horizontal bars. The absence of occlusions ensures a clearer representation of the event distributions.

The matrix display of COVID-19 events (Fig. 2, left) reveals dis-

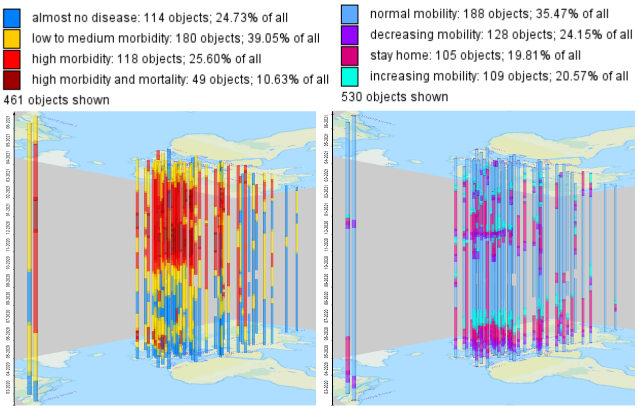

Figure 1: *The events of COVID (left) and mobility (right) are visualised in space-time cube displays. The colours correspond to the event types, as specified in the legends above the cubes.*

Figure 2: *The events of COVID (left) and mobility (right) are represented in an artificial matrix space with rows corresponding to the countries and columns to week-long time steps along the study period. The colours are the same as in Fig. 1.*

tinct patterns corresponding to two notable waves of the pandemic. A relatively brief wave occurred in spring 2020, followed by a considerably longer wave extending from autumn 2020 to early summer 2021. During the initial wave, only a few countries experienced severe COVID situations, as indicated by red and dark red colouring. In contrast, the second wave was not only prolonged but also more intense, with instances of 'high morbidity' (red) and 'high morbidity and mortality' (dark red) events occurring in numerous countries.

Upon comparing the matrix of COVID events with that of mobility events (Fig. 2, right), significant disparities in event distributions during the pandemic's first and second waves become apparent. Despite the milder nature of the initial wave, occurrences of 'decreasing mobility' and 'stay home' events were widespread across nearly all countries and exhibited synchronicity, irrespective of individual countries' COVID situations. Conversely, during the more severe second wave, these mobility event types were less prevalent across countries, and their occurrences were largely asynchronous, with exceptions noted during the Christmas period.

These observations underscore the dynamic and context-dependent nature of the relationships between COVID and mobility events, highlighting the importance of incorporating contextual information in the analysis of event relationships.

5. Approach

We introduce the concept of the local temporal context of an event. Let t_0 be the time of the event appearance and $[t_0 - \Delta_1, t_0 + \Delta_2]$ be a time interval called *temporal neighbourhood* of the event, where Δ_1 and Δ_2 are chosen numbers of time units. The *context* of a given event (henceforth denoted as *target event*) encompasses the time intervals during which all event types included in the analysis are present within its temporal neighbourhood. In our analysis, we represent the individual event context by dividing the temporal neighbourhood into a specified number of equal time steps, or bins. Subsequently, we construct a matrix with columns for the time steps, rows for all event types, and cells containing either 1 or 0 to indicate the presence or absence of events of the respective types in each time step. The target event is also included in the context; the interval of its existence is represented in the matrix row corresponding to the event type.

We construct the local temporal context for each COVID event and each mobility event. Here, we define the temporal neighbourhood of an event by setting $\Delta_1 = 4$ weeks and $\Delta_2 = 6$ weeks. This neighbourhood is divided into 10 time steps, each spanning one week. The context encompasses both COVID event types and mobility event types, 8 types in total. Consequently, an event context is represented by a matrix with 8 rows and 10 columns. Despite the uniform representation, the contexts of COVID and mobility events are analysed separately. The reason is the conceptual differences between the two event families, necessitating distinct approaches to interpreting event contexts and any artefacts obtained during the analysis.

For the application of computational analysis methods, we transform the matrices into vectors by concatenating the rows; hence, the dimensionality of the vectors equals the product of the number of event types and the number of the time steps. To analyse the diversity of the local contexts and identify groups of similar contexts, we employ dimensionality reduction on the vectors. In the presented example, we use the UMAP method [MHSG18, MHM20], which prioritises placing close neighbours in proximity in the embedding space at the cost of potentially distorting distances between non-neighbouring objects. Consequently, groups of similar data items appear as compact clusters of points in the projection plot, facilitating visual detection and interactive selection through brushing.

UMAP has two main parameters, number of neighbours to consider *n_neighbors* and the minimal distance between points in the embedding *min_dist*, which controls how tightly UMAP may pack points together. To explore the variability of the projections, we iteratively run UMAP with different parameter combinations, resulting in multiple projection variants. Subsequently, we compare the resulting projection plots to assess their consistency and select the one best suited for further analysis. In our example, UMAP consistently produced robust results across various parameter combinations. We chose projections with well-separated point clusters for the convenience of subsequent analysis, allowing for easy selection and interpretation.

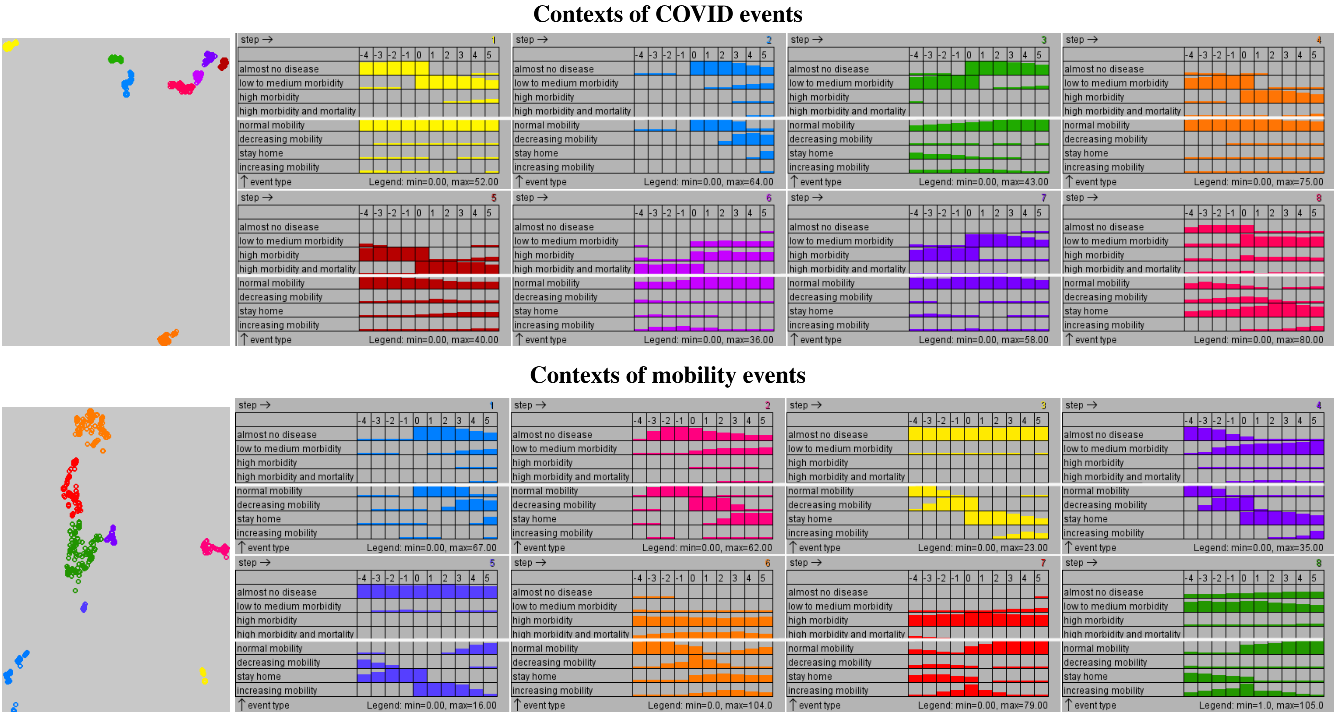

Figure 3: *Exploration of the temporal contexts of the COVID events (upper panel) and mobility events (lower panel). On the left are UMAP projections with interactively defined classes of contexts. On their right, profiles of the context classes are represented by matrices with rows corresponding to the types of COVID and mobility events and columns to the relative times within the context time windows.*

The UMAP projections of the contexts of the COVID events and mobility events are displayed on the left of Fig. 3. The shown projection variant for the COVID events was obtained with $n\_neighbors = 25$ and $min\_dist = 0.3$; for the mobility events, $n\_neighbors = 25$ and $min\_dist = 0.05$. In the projection plots, we interactively selected clusters of closely positioned points and assigned distinct labels and colours to the corresponding groups of local contexts, thus defining *classes of contexts*. Please note that one context class may include local contexts of target events of different types.

To enable viewing, interpretation, and comparison of the context classes, we devised a visual display comprising multiple matrices, each representing the contexts of one class in an aggregated form. These multi-matrix displays are depicted on the right of the projection plots in Fig.3. In the matrices, the rows correspond to the event types, and the columns represent the time steps within the temporal neighbourhoods of the target events. The top four rows of the matrices correspond to the types of COVID events, while the bottom four rows represent mobility event types. The first four columns represent the four-week period preceding the onset of the target events, with the fifth column depicting the week when the target events commenced. The remaining five columns correspond to the subsequent weeks. The contents of the cells are determined by summing the individual context matrices included in the respective groups. The resulting numbers are depicted by the proportional heights of vertical bars, which are painted in the colours assigned to the context classes.

Each matrix reveals the prevailing trends in a group of 10-week intervals. For instance, let us take class 1 of the COVID event contexts depicted in the upper panel of Fig.3, indicated by the yellow colour. The top two rows of the matrix show a shift from the event 'almost no disease' to 'low to medium morbidity'. In the fifth row, 'normal mobility' remains the prevailing mobility event throughout the 10-week period. Hence, this context class signifies periods where worsening of the pandemic situation does not significantly impact the mobility behaviour. Conversely, in class 5 (dark red), we observe a decrease in the frequency of 'normal mobility' as the COVID event transitions from 'high morbidity' to 'high morbidity and mortality'. Among the context classes of the mobility events in the lower panel of Fig.3, we encounter an intriguing pattern in class 3 (yellow). Here, we observe consecutive transitions 'normal mobility' – 'decreasing mobility' – 'stay home' – 'increasing mobility', despite the predominant COVID event type being 'almost no disease' throughout the entire period. The matrix of class 4 (purple) exhibits a similar sequence of transitions between the mobility events as the pandemic situation changed from 'almost no disease' to 'low to medium morbidity'.

To gain a clearer understanding of the contexts in which each event type occurred, we devised co-occurrence plots shown in Fig. 4. Here, the rows starting from the second one correspond to the types of the target events, namely COVID events in the upper panel and mobility events in the lower panel. The top rows represent the complete sets of the target events. The frequencies of event occurrence within the defined context classes are depicted by the

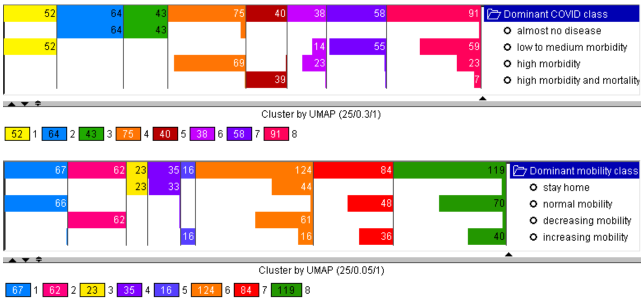

Figure 4: *Occurrence frequencies of the types of target events in the context classes. Top: COVID events; bottom: mobility events.*

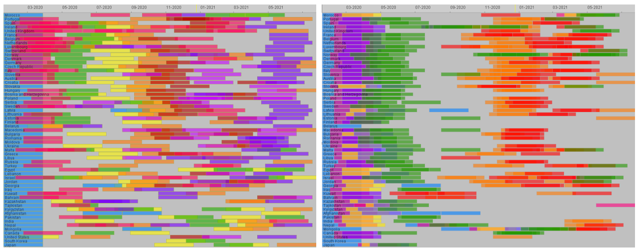

Figure 5: *The contexts of the COVID events (left) and of the mobility events (right) are represented in an artificial matrix space by bars coloured according to the context classes.*

proportional lengths of horizontal bars coloured according to the respective classes.

To examine the distributions of the context classes over time and across countries, we put the contexts in the artificial matrix space analogously to representing the events in Fig. 2. In Fig.5, the coloured bars depict the contexts rather than the events, with horizontal positions corresponding to the times when the contexts occurred, and vertical positions indicating the countries where they took place. As previously, the colours represent the context classes. The grey spaces between the bars indicate instances where the time intervals between the beginnings of consecutive target events exceeded 10 weeks. Consequently, there is considerable grey space in the display of the contexts of the mobility events, with 20% of them (108 out of 530) lasting for more than 10 weeks.

As there is no space for a detailed discussion of the information conveyed by the visualisations, we give only a brief summary.

General observations. The co-occurrence plots in Fig. 4 reveal that the same types of events may occur in at least two distinct classes of contexts. Each context class captures a specific combination of relationships between events, as illustrated by the multi-matrix displays in Fig. 3. Consequently, the relationships of each event type with other event types are context-dependent. The temporal trends observed in the succession of the context classes in Fig. 5 indicate that event relationships evolve over time.

More specific observations. The imposition of stringent mobility restrictions leading to the 'stay home' event often did not solely stem from the severity of COVID situations within the countries where they were implemented, particularly during the initial wave of the pandemic. However, these measures may have helped prevent the emergence of severe COVID conditions. Conversely, maintaining mobility at normal levels amidst 'high morbidity' contexts tended to exacerbate COVID situations further. On the other hand, transitioning to the 'stay home' regime did not immediately ameliorate the COVID situation within the subsequent 6-week period. To observe delayed effects, extending the temporal spans of the event contexts is necessary. It is also reasonable to extend the contexts thematically by including additional information, in particular, reflecting the overall pandemic situation in the world and/or in the neighbouring countries.

6. Discussion and conclusions

The example considered in this paper underscores a critical lesson: intuitive assumptions regarding potential causal relationships between events can often be oversimplified or incorrect. Similarly, assuming that relationships remain static across all contexts and time frames may lead to flawed analyses. Therefore, before applying approaches that rely on computational discovery of causal relations or quantification of predefined causal links, it is imperative to examine whether interrelationships between event types remain consistent across various dimensions. This exploration may unveil disparities between different periods or subsets of data, suggesting the need for separate consideration of these data portions in subsequent analysis and modelling efforts [AA23].

We have proposed an approach to enable such an exploration. The key idea is to represent local temporal contexts of events by a data structure that allows numeric estimation of context similarity and application of computational techniques, such as projection and clustering. This representation, namely, a binary matrix, is also convenient for aggregation of multiple contexts and for visualisation of resulting aggregates. The approach involves delineating distinct classes of event contexts through interactive grouping or algorithmic clustering based on similarity, followed by an examination of the distribution of these context classes across various data dimensions, including time and geographical space.

Our approach poses no principal limitation on the number of event sequences. Moreover, contexts may include not only information about event occurrences but also any information that can be represented by categories or discrete attribute values.

Our future research direction will focus on leveraging event contexts for predictive modelling. The overarching objective is to develop models capable of forecasting subsequent events or estimating the probabilities of occurrence for different event types given a partial context. By incorporating contextual dynamics and dependencies in the process of model building, we hope to represent in the models the evolution of complex dynamic phenomena and interrelationships between them.

Acknowledgements

This work was supported by Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia as part of the *Lamarr Institute for Machine Learning and Artificial Intelligence* (Lamarr22B), and by EU in projects *SoBigData++* and *CrexData* (grant agreement no. 101092749).

References

[AA23] ANDRIENKO N., ANDRIENKO G.: It's about time: Analytical time periodization. *Computer Graphics Forum 42*, 6 (2023), e14845. doi:10.1111/cgf.14845. 5

[AAF*20] ANDRIENKO N., ANDRIENKO G., FUCHS G., SLINGSBY A., TURKAY C., WROBEL S.: *Visual analytics for data scientists*. Springer, 2020. doi:10.1007/978-3-030-56146-8. 1

[AAS23] ANDRIENKO N., ANDRIENKO G., SHIRATO G.: Episodes and topics in multivariate temporal data. *Computer Graphics Forum 42*, 6 (2023), e14926. doi:10.1111/cgf.14926. 2

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of time-oriented data*, vol. 4. Springer, 2011. 1

[DAA*19] DUNKEL A., ANDRIENKO G., ANDRIENKO N., BURGHARDT D., HAUTHAL E., 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. 1

[GGJ*21] GUO Y., GUO S., JIN Z., KAUL S., GOTZ D., CAO N.: Survey on visual analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics 28*, 12 (2021), 5091–5112. doi:10.1109/TVCG.2021.3100413. 2

[Goo22] GOOGLE: Google COVID-19 open data repository, 2022. [accessed 7-September-2022]. URL: https://health.google.com/covid-19/open-data/. 2

[JGC*20] JIN Z., GUO S., CHEN N., WEISKOPF D., GOTZ D., CAO N.: Visual causality analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2020), 1343–1352. doi:10.1109/TVCG.2020.3030465. 2

[MHM20] MCINNES L., HEALY J., MELVILLE J.: Umap: Uniform manifold approximation and projection for dimension reduction, 2020. arXiv:1802.03426. 3

[MHSG18] MCINNES L., HEALY J., SAUL N., GROSSBERGER L.: Umap: Uniform manifold approximation and projection. *Journal of Open Source Software 3*, 29 (2018), 861. URL: https://doi.org/10.21105/joss.00861, doi:10.21105/joss.00861. 3

[MLL*13] MONROE M., LAN R., LEE H., PLAISANT C., SHNEIDERMAN B.: Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2227–2236. doi:10.1109/TVCG.2013.200. 2

[vdvV23] VAN DER LINDEN S., DE FOUW E., VAN DEN ELZEN S., VILANOVA A.: A survey of visualization techniques for comparing event sequences. *Computers & Graphics 115* (2023), 522–542. doi:10.1016/j.cag.2023.05.016. 2

[WG12] WONGSUPHASAWAT K., GOTZ D.: Exploring flow, factors, and outcomes of temporal event sequences with the Outflow visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2659–2668. doi:10.1109/TVCG.2012.225. 2

[WGGP*11] WONGSUPHASAWAT K., GUERRA GÓMEZ J. A., PLAISANT C., WANG T. D., TAIEB-MAIMON M., SHNEIDERMAN B.: Lifeflow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), pp. 1747–1756. doi:10.1145/1978942.1979196. 2

[XHW20] XIE X., HE M., WU Y.: Causalflow: Visual analytics of causality in event sequences. *arXiv preprint arXiv:2008.11899* (2020). 2

[ZSZ*22] ZHU S., SHEN Y., ZHU Z., XIA W., CHANG B., LIANG R., SUN G.: Vac2: Visual analysis of combined causality in event sequences. *arXiv preprint arXiv:2206.05420* (2022). 2