VIPRA – EuroVis Workshop on Visual Process Analytics (2024)
A. Arleo, L. Pufahl, J. Rehse, F. Zerbato, S. van den Elzen, and T. von Landesberger (Organizers)

Interacting with Large Process Data: Challenges in Visual Exploration

G. Shurkhovetskyy[1], P. Schumacher[1] and P. Koytek[1]

[1]Celonis

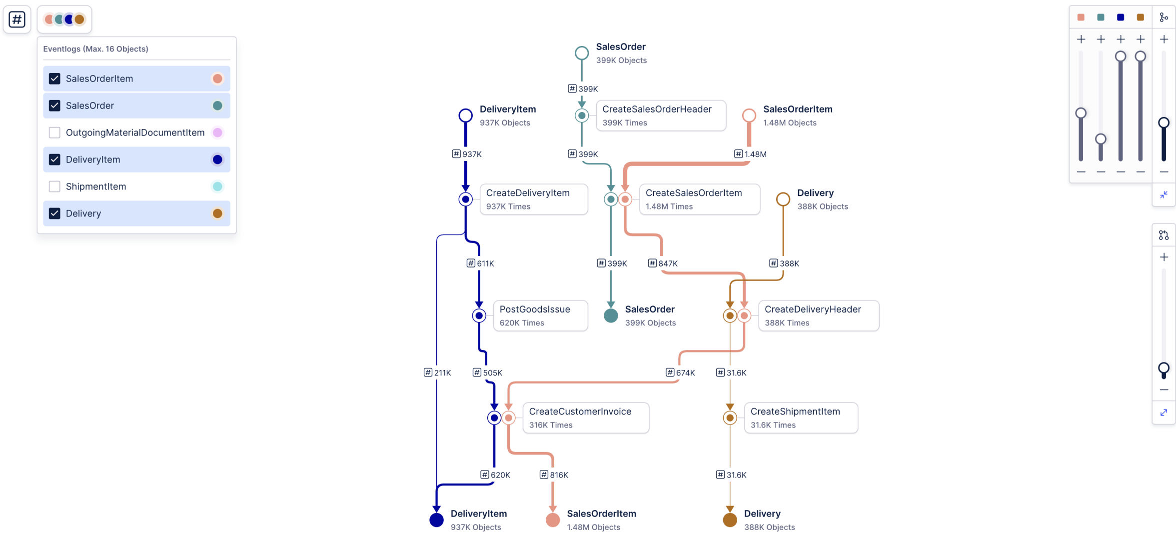


Figure 1: *Graph-based representation of a business process with interactive elements for exploring the process along events and objects.*

Abstract

Information systems across various industries become more common every year. Organizations aim to leverage data collected by such systems to gain useful knowledge and, ultimately, improve their business processes. Process analysis is often facilitated by visual analytics enriched with interactions. However, a systematic evaluation of applicability and implication of use of known interaction techniques to process mining tasks is missing. In this paper, we provide an overview of interaction methods used at Celonis and propose their initial categorization in the context of process mining. We then describe further challenges of interactive visualizations for process analysis from an industry perspective. Finally, we offer directions for future user studies and research to further strengthen the combination of process mining with visual analytics.

CCS Concepts

• ***Human-centered computing*** $\rightarrow$ ***Visual analytics;***

1. Introduction

Data generation across fields such as medicine, security, logistics, manufacturing, and others continues to grow every year. Worldwide digitalization results in adoption of information systems that collect transactional data about everyday business processes.

Process mining is a data-driven methodology that involves extracting knowledge and insights from event logs recorded during the execution of operational processes within an organization. It aims to discover, monitor, and improve real-life processes based on their actual execution data rather than relying on predefined models, assumptions or interviews. It has become a category of business

DOI: 10.2312/vipra.20241102

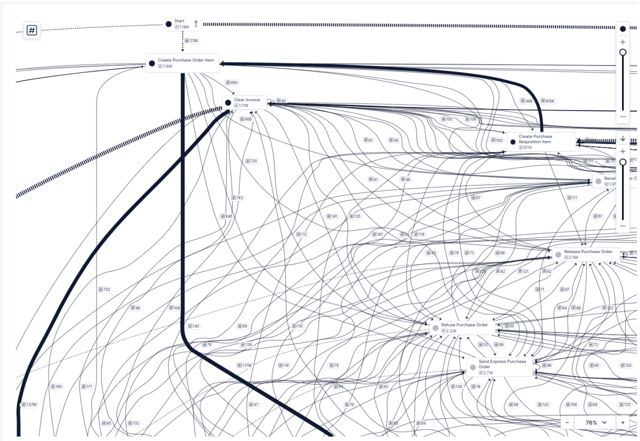


Figure 2: *Partial view of the process model data in Celonis. Large volume of events and connections clutters the display and renders it unusable for generating actionable insights unless modifications are applied. Extraction and iteration are some of the strategies that help alleviate the visual overload.*

technology that has already been successfully applied by a large number of organizations around the world [vdA22].

In its most basic form, process data consists of an event log with three pieces of information for every event: a name of the performed activity, a unique reference to the manipulated business object, and a timestamp. Now end-to-end visualizations of the process can be created by Process Mining tools superimposing every step that every case took as it moved through the cycle. Each unique path is called a variant. In large real-world data sets (i.e. several billion events) variants often number in the thousands or tens of thousands, showing all possible variants at once creates a "spaghetti" diagram, see Figure 2. This provocative visualization is often used at the beginning of a Process Mining project to convincingly demonstrate the delta between expected happy path. The happy path is an informally agreed-upon concept within the process mining community, which describes the most desirable variant, often being the most frequent one.

Due to the spaghetti diagram's complexity, it is very difficult to get actionable insights from it. Appropriate storage, abstraction, modeling and visual representation of event data can enable operators and analysts to extract actionable insights. Therefore different ways of interacting, slicing, aggregating and visualizing the process data are required to drive action and improvements in real-world business processes [DLRMR13].

Visual analytics is defined as "the science of analytical reasoning facilitated by visual interactive interfaces" [TC05] and combines the computation methods of knowledge discovery with the abilities of human perception facilitated by visual representation of complex data. This makes it a useful tool for process analysis and exploration as it enables quicker overview and identification of patterns and outliers.

In this paper we discuss challenges of designing interactions for exploratory process visualizations we faced to make complex business processes approachable to a large and diverse user base with the goal of reaching insights easier. For this, we report on two exploratory interaction mechanics designed to gradually unfold process graphs and share our reflections on combining different techniques. The learnings we share are informed by our experience designing interactions for process exploration and are augmented by user feedback. They can be used as foundation for further research and can help to discuss and evaluate more interaction techniques for the exploration of process graphs.

This paper is organized as follows. In Section 2, we provide overview of the related work. Section 3 provides an overview of the two interaction techniques and the criteria according to which their benefits and drawbacks are drawn. Section 4 extends the discussion of interactions with challenges and prospects of interaction design and visual analytics more broadly in the context of process mining. It is followed by a conclusion and suggestions for future research in Section 5.

2. Background and Related Work

Interaction is an important tool for understanding large process data. Du et al. [DSP*17] describe and group 15 strategies for addressing large data volume. One of them is *iterative strategies* that resemble exploratory analysis often performed on process data. Process mining research further confirms that analysts often resort to interactive methods within the software they use to explain their findings to different audiences [KMN22]. Yi et al. [YKSJ07] demonstrate the importance of interaction in visualization and provide a categorization based on the user intent.

Several works cover the intersection of visual analytics and process mining [KPRMS16, Gsc17, ZYG*23]. Notably, Gschwandtner [Gsc17] provides an overview of challenges found across both fields. One of them is "Interaction to Support Process Discovery and Enhancement".

Kriglstein et al. [KPRMS16] made important steps towards classification of process mining techniques against visual analytics aspects. They evaluate process mining views according to criteria such as interaction and visualization data type.

Sirgmets et al. [SMP18] provide a methodological framework for developing visual representations of process diagrams. Their approach of breaking down stages of the data visualization pipeline [Chi00] into three levels of granularity allow for more systematic tackling of process mining problems and designing interactions around them.

Yeschenko et al. [YM24] provide useful overview of visualization techniques applied to event sequence data. Authors iterate through numerous papers that describe methods to visualize process data. They group reviewed visualizations by representation type and discuss their applicability to different variations of event sequences.

Guo et al. [GGJ*22] take an opposite approach and instead tackle the lack of mapping between data visualization and process mining from the perspective of the latter. They propose categorization of visualizations of event sequence data based on high-level analytical tasks in the context of process mining (summarization,

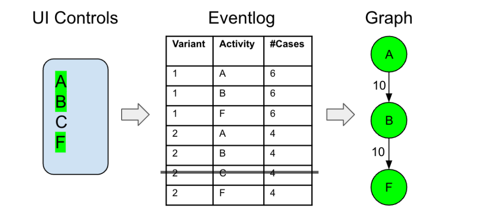
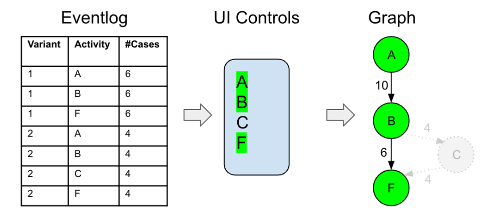
INTERACTION TECHNIQUES		EFFECTS			
Interaction Behaviour	Interaction Interface	Predictability	Freedom of Exploration	Granularity	Ease of Use
 UI Controls manipulating the eventlog	Slider	Low: not clear what appears next	Sequential exploration	Fine or coarse (move slider quickly)	High: switching focus to and from main view is not necessary
	List View	High: explicitly choosing what appears next (but not clear where)	Free exploration	Fine granularity (choosing one by one)	Medium: intuitive to use but requires switching from the main view for every action
 UI Controls manipulating the graph	Slider	Low: not clear what appears next	Sequential exploration	Fine or coarse (move slider quickly)	High: switching focus to and from main view is not necessary
	List View	High: explicitly choosing what appears next (but not clear where)	Semi-free exploration: some restrictions in place to ensure the graph is connected	Fine granularity (choosing one by one)	Medium: intuitive to use but requires switching from the main view for every action

Figure 3: *Interaction techniques represented by two types of supported behaviour (eventlog manipulation, graph manipulation) and two types of visual interface (sliders, list) and assessed according to the effects each of the combinations of behaviour and interface entails.*

comparison, anomaly detection, etc.). They also provide a summary table with an evaluation of each reviewed approach against four dimensions:

- Data scales.
- Automated Sequence Analysis.
- Visual representations.
- Interactions.

We believe that combining the two approaches described in [YM24] and [GGJ*22] with further specialization would improve synergy between visualization and process mining. For example, analysts might be constrained to use graph-based visualization but could have the freedom to choose among several interaction techniques for tackling their process mining task. Mapping interaction types to process mining tasks would provide a structured framework of operation for analysts and a useful roadmap for researchers.

3. Evaluation of Exploratory Interactions

Process data often includes thousands of events. This means that with sufficient number of observations and certain degree of deviation from the standard process, thousands of process variants could be present. Displaying all variations can quickly become non-feasible both computationally and in terms of analyst's ability to visually perceive all information.

To illustrate the challenge of selecting or designing an interaction for process data, we first need to describe the aspects to characterize the options available to us. We classify exploratory interaction methods according to the following two facets:

- **Interaction interface.** The user interface or input elements used for the exploratory interaction of the process.
- **Interaction behaviour.** The behaviour of the interaction, ie. what does the user interface element control (the eventlog or simply the visibility of events).

We then describe the consequences, or **effects** that characterize how the interaction is used, perceived, and how it can be combined with other interaction methods.

In this section, we introduce each of these facets, or dimensions, and evaluate the two interactions methods used in Celonis Process Explorer. We summarize our findings in Figure 3.

3.1. Interaction interface

We distinguish between two different interaction interfaces to control the process exploration which are implemented in the Celonis Process Explorer. Furthermore we are focusing on the interactions which allow to gradually explore the graph. Interactions such as filtering out individual cases are out of the scope of this paper.

Slider. This interaction method allows users to increase or decrease coverage of the events and connections by dragging the thumb of the slider element (right of Figure 1). Buttons (+/-) allow incremental change of the number of events shown. Currently, frequency count is used as the key indicator for what is shown next. Initially, only the most frequent events are displayed. Dragging the slider thumb up, an analyst can increase the number of shown events. As the frequency threshold lowers, more events appear in the graph.

List of events. In order to provide users with more control of what they want to be displayed, we offer the control of visibility of spe-

Figure 4: *Multiple coordinated views in Celonis. On the left, Process Explorer graph shows process data with connection thickness encoding the number of cases. Top right view shows the Bar Chart depicting frequencies of activities categorized by event type. Table on the bottom right shows a list of values sorted by KPI.*

cific events. For that, a textual list of events is provided which allows for granular control of the visualization by simply checking the events explicitly that should be shown in the graph. The list is sorted according to the frequency of the events.

3.2. Interaction behaviour

The ways to produce the results of an interaction differ. We describe two possible methods of view manipulation that enable visual abstraction used for generation of both the initial view, and the subsequent views during the iterative exploration.

Eventlog controls. This method essentially manipulates the eventlog behind the visualization with every interaction to only include selected events in the eventlog and exclude the rest from the calculations. Top left of Figure 3 shows how selecting events A, B and F with such eventlog controls results in removal of the event C from the event log entirely. The graph is then drawn based on the modified eventlog. Number of cases from both variants (4 + 6) that end with event F are summed up and yield number 10 (label on the connection between B and F).

Graph controls. Another approach is to modify the visibility of events in the graph, without changing the underlying eventlog. Bottom left of Figure 3 shows how the eventlog remains unaffected by the UI controls. Instead, on selection of events A, B and F, only their visibility in the graph changes. This results in hiding both the event C, and the connections it has to all other events. With this method, the four cases that lead to event F via C become invisible and are no longer reflected in the connection between B and F (6).

3.3. Effects

To compare the impact of choosing one combination of interaction interface and behaviour over another, we identified four dimensions of effects which we assessed to be the most relevant for our use case.

Predictability describes to what degree the user can know before an interaction what changes will be made to the graph. For example, which event will appear next and where? As such, sliders do not reveal what events will appear next when dragging the thumb up. In contrast, with a list view, the analyst can explicitly choose the events they want to appear next.

Freedom of exploration describes how freely an analyst can choose their path of exploration, eg. in which order they want to reveal events. The slider interface only allows a strictly sequential discovery as it reveals events based on their frequency (potentially, another metric can be used instead). The list view, on the other hand, allows for a free selection of events in virtually any order. However, certain restrictions might still apply depending on the interaction behaviour. For example, when behaving as a graph control (ie. manipulating the visibility of events in the graph), even a list view might not allow deselecting of certain events if such action would result in a disconnected graph.

Granularity. Slider allows for many changes to quickly appear on screen if the user drags it fast enough. This speeds up the exploration if the user does not want to spend time at every step. Finer interaction granularity is also available with the slider via the + or - buttons that reveal/hide only one element at a time. List of events control only supports choosing elements one by one.

Ease of use. As both interaction interfaces we use are *instrumental interactions* [BL00] (as opposed to *direct manipulation* [Shn83] where users act on visual objects of interest), it is important to evaluate how cognitively demanding their use is. With the list view, user focus will travel back and forth between the list interface and the visualization to repeatedly add events and see where in the graph they appeared. This is in contrast to the slider which allows the user to fully focus on the visualization while dragging the thumb freely up and down (or, alternatively, repetitively press the +/- button) to modify the view. We measure ease of use by how often users have to switch focus during repeated interactions.

4. Challenges and Prospects

4.1. Filtering Behaviour

Interaction behaviour also has important implications on filtering, which is another interaction technique that is instrumental for data exploration [YKSJ07]. For instance, a user might want to only see cases that end at a specific event. However, the log might contain events that follow the event upon which the user wants to filter the view. In case the user previously applied graph controls to hide the subsequent events, the filter will not work properly because no case in the eventlog actually ends at the event the user selected. Further confusion will arise from discrepancy between the view, and the underlying eventlog. On the other hand, eventlog controls will support such filtering operation because removal of subsequent events at the previous step would have resulted in modification of the underlying eventlog.

Challenge: conveying implications of interaction types to the user.

4.2. Multiple Coordinated Views

Many analysts use a combination of several tools to find, select and communicate insights [KMN22]. At Celonis, we have been receiving consistently positive feedback from customers about the utility of multiple coordinated views. Figure 4 shows how they are used in conjunction with the Process Explorer.

Our observations echo existing research [Rob07] and confirm that ability to explore data from different perspectives enables users to find insightful relationships. To further that, Xu et al. [XMRC17] propose a visualization for a manufacturing process that aims at enhancing readability and rate of successful failure prediction. For this, the authors combine an extended Marey's graph enriched with interactive coordinated views that allow for efficient exploration and understanding of causes for failures or deviations in repetitive processes. Their approach further demonstrates that given well-defined set of data attributes and analysis requirements, complex processes can be effectively visualized.

However, proper utilization of multiple views is only possible with their consistent coordination. Depending on the interaction behavior used, different scenarios for view updates must be implemented.

Challenge: view coordination across views that might or might not employ different interaction behaviour.
Prospect: more compact visualizations of complex processes with rich exploration support.

4.3. Conveying the Impact of Interaction

Interacting with the visualization produces new or updates the existing view of the data. Introduction of more visual elements to an already busy view might further complicate data understanding. In addition, it might not be intuitive to predict how instrumental interactions will affect the view. In order to make interactions congruent with the intention of the user, we identify the following relevant questions:

- How to maintain user's perception of the data and focus on parts of the view that they wanted to explore?
- How to convey *what* will appear on screen?
- How to convey *where* will new elements appear?
- How best to introduce new visual elements and highlight changes between the two stages of the exploration?

Willett et al. [WHA07] suggest *social navigation* within visualization to ease navigation of the new data for the user. They embed visual cues based on interactions and findings of other users into the interface controls. Such approach however requires collection and access to activity traces of other exploration sessions.

Challenge: maintain focus and user perception with instrumental-only interaction.
Prospect: support of direct manipulation can make exploration of complex data more intuitive.

4.4. User Adaptation

Introducing new interaction methods poses a separate set of challenges. Our experience shows that users often exhibit familiarity bias when it comes to user interfaces. This is further echoed by research [MFK*21]. Such bias applies across many user groups but especially noticeable among large traditional companies established decades ago. With that, the following questions have to be answered:

- How to measure and quantify if a change is acceptable for users?
- How to best transition customers between interaction patterns?
- How to familiarize customers from traditional companies with the new methods in an evolving field?

Challenge: slow user adaptation of novel interaction methods.
Prospect: faster track to gaining insights using the new techniques.

4.5. Object-centric Event Data

These challenges get even harder with **object-centric event data**. Traditional process mining, although very powerful, is built around the notion of processes running in isolation from each other. Traditionally, events in log data refer to one case. With object-centric event data, events may relate to any number of objects. Such an approach improves scalability due to more efficient data storage, transfer, and aggregation [MFK*21, vdA23]. It also opens up opportunities in visual representation of the data. However, due to different data representation, existing techniques for even basic operations such as filtering, clustering and prediction, have to be reinvented.

This means that approaches for visual representation of object-centric event data have to consider the nuance of underlying data model which makes the use of established interactions methods non-trivial if at all relevant.

Challenge: need to reinvent even basic representation and manipulation methods.
Prospect: ability to leverage advantages of object-centric event data.

5. Conclusion

Process mining is drawing more interest from industry and academia every year. Combining it with visual analytics enables efficient exploration of large volume of process data. Although a growing number of works research how the two fields can be effectively brought together, to the best of our knowledge, there does not exist a framework that maps visual interaction techniques to specific tasks of process understanding. Our paper aims at making the steps towards filling this gap from an industry perspective. For that, we provide an overview of interaction techniques with focus on their behaviour, interface, and effects. We then augment it with discussion of the challenges and prospects of visual analytics related to views' coordination, use and adoption of novel interaction methods by a large and diverse user base. Our overview of the interactions could serve as a foundation for future user studies and research involving more types of interactions and evaluating their utility for specific process mining tasks.

References

[BL00] Beaudouin-Lafon M.: Instrumental interaction: an interaction model for designing post-wimp user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2000), CHI '00, Association for Computing Machinery, p. 446–453. URL: https://doi.org/10.1145/332040.332473, doi:10.1145/332040.332473. 4

[Chi00] CHI E. H.: A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization 2000 (INFOVIS'00), Salt Lake City, Utah, USA, October 9-10, 2000* (2000), Mackinlay J. D., Roth S. F., Keim D. A., (Eds.), IEEE Computer Society, pp. 69–75. URL: https://doi.org/10.1109/INFVIS.2000.885092, doi:10.1109/INFVIS.2000.885092. 2

[DLRMR13] DUMAS M., LA ROSA M., MENDLING J., REIJERS H. A.: *Fundamentals of Business Process Management*. Springer, Berlin, 2013. doi:10.1007/978-3-642-33143-5. 2

[DSP*17] DU F., SHNEIDERMAN B., PLAISANT C., MALIK S., PERER A.: Coping with volume and variety in temporal event sequences: Strategies for sharpening analytic focus. *IEEE Transactions on Visualization and Computer Graphics 23*, 6 (2017), 1636–1649. doi:10.1109/TVCG.2016.2539960. 2

[GGJ*22] 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 (dec 2022), 5091–5112. URL: https://doi.org/10.1109/TVCG.2021.3100413, doi:10.1109/TVCG.2021.3100413. 2, 3

[Gsc17] GSCHWANDTNER T.: Visual analytics meets process mining: Challenges and opportunities. In *Data-Driven Process Discovery and Analysis* (Cham, 2017), Ceravolo P., Rinderle-Ma S., (Eds.), Springer International Publishing, pp. 142–154. 2

[KMN22] KUBRAK K., MILANI F., NOLTE A.: Process mining for process improvement - an evaluation of analysis practices. In *Research Challenges in Information Science* (Cham, 2022), Guizzardi R., Ralyté J., Franch X., (Eds.), Springer International Publishing, pp. 214–230. 2, 4

[KPRMS16] KRIGLSTEIN S., POHL M., RINDERLE-MA S., STALLINGER M.: Visual Analytics in Process Mining: Classification of Process Mining Techniques. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2016), Andrienko N., Sedlmair M., (Eds.), The Eurographics Association. doi:10.2312/eurova.20161123. 2

[MFK*21] MARTIN N., FISCHER D. A., KERPEDZHIEV G. D., GOEL K., LEEMANS S. J. J., RÖGLINGER M., VAN DER AALST W. M. P., DUMAS M., LA ROSA M., WYNN M. T.: Opportunities and challenges for process mining in organizations: Results of a delphi study, 2021. URL: http://dx.doi.org/10.1007/s12599-021-00720-0;https://dl.gi.de/handle/20.500.12116/37550, doi:http://dx.doi.org/10.1007/s12599-021-00720-0. 5

[Rob07] ROBERTS J. C.: State of the art: Coordinated multiple views in exploratory visualization. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)* (2007), pp. 61–71. doi:10.1109/CMV.2007.20. 5

[Shn83] SHNEIDERMAN: Direct manipulation: A step beyond programming languages. *Computer 16*, 8 (1983), 57–69. doi:10.1109/MC.1983.1654471. 4

[SMP18] SIRGMETS M., MILANI F. P., PUNGAS T.: A visualization framework for designing process mining diagrams. URL: https://api.semanticscholar.org/CorpusID:21675213. 2

[TC05] THOMAS J. J., COOK K. A.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005. URL: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0769523234. 2

[vdA22] VAN DER AALST W. M. P.: *Foundations of Process Discovery*. Springer International Publishing, Cham, 2022, pp. 37–75. URL: https://doi.org/10.1007/978-3-031-08848-3_2, doi:10.1007/978-3-031-08848-3_2. 2

[vdA23] VAN DER AALST W. M. P.: Object-centric process mining: Unraveling the fabric of real processes. *Mathematics 11*, 12 (2023). URL: https://www.mdpi.com/2227-7390/11/12/2691, doi:10.3390/math11122691. 5

[WHA07] WILLETT W., HEER J., AGRAWALA M.: Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1129–1136. doi:10.1109/TVCG.2007.70589. 5

[XMRC17] XU P., MEI H., REN L., CHEN W.: Vidx: Visual diagnostics of assembly line performance in smart factories. *IEEE Transactions on Visualization and Computer Graphics 23* (2017), 291–300. URL: https://api.semanticscholar.org/CorpusID:5602193. 5

[YKSJ07] YI J. S., KANG Y. A., STASKO J., JACKO J.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1224–1231. doi:10.1109/TVCG.2007.70515. 2, 4

[YM24] YESHCHENKO A., MENDLING J.: A survey of approaches for event sequence analysis and visualization. *Information Systems 120* (2024), 102283. URL: https://www.sciencedirect.com/science/article/pii/S0306437923001199, doi:https://doi.org/10.1016/j.is.2023.102283. 2, 3

[ZYG*23] ZINAT K. T., YANG J., GANDHI A., MITRA N., LIU Z.: A comparative evaluation of visual summarization techniques for event sequences, 2023. arXiv:2306.02489. 2