1 Introduction

In a broad context, *sensemaking* reflects how we make sense of the world so that we can take further actions in it [160]. More specifically, sensemaking is described as a motivated and continuous effort to understand the relationships among people, places and events around us in order to act more effectively [94]. It is the process of collecting, organizing and representing complex information sets with an emphasis on some problem we need to solve [149]. For example, intelligence analysts may need to examine thousands of reports to establish deep understanding of particular persons or organizations before identifying possible threats from them. In an everyday context such as selecting a smartwatch ~~for purchase~~, a person may search for different models on the Internet, learn unknown technical terminologies and consider pros and cons between these identified models. Automated techniques in information retrieval and data mining may speed up the sensemaking process. For instance, named-entity recognition technique [119] can identify and classifies entities in text into predefined categories such as persons, organizations and locations. It saves analysts a considerable amount of time in common tasks such as findings all reports mentioning a particular person. However, human still involves heavily in sensemaking to synthesize knowledge and draw inferences before making informed decisions.

Visualization is a computer-based system designed to help people perform tasks more effectively through visual representations of datasets [118]. It can display a large amount of data in a small space and allow for close examination of details through interaction. During the data exploration, many insights, questions and (conflict) hypotheses may appear. Unfortunately, human with limited capacity of working memory cannot simultaneously hold all of these findings. Visualization can help expand working memory by allowing people to externalize internal cognition and memory usage to the visual displays, and organize them in a meaningful structure. For instance, research has shown that spatially grouping of findings and drawing links between them to indicate their relationships can facilitate reasoning and sensemaking [153].

Not only are the final outcomes of the sensemaking process important, but the process itself is also of great value [141]. *Analytic provenance* captures both the interactive data exploration process and the accompanied human reasoning process during sensemaking [191]. This provenance information allows users to recall and revisit the process, typically combining with undo/redo and bookmarking features [74]. It also allows reproduction of the process, possibly with a different dataset and parameter settings [39]. Provenance can facilitate learning through building tutorials [63], presentation [48] and improving communication between teachers and students [139].

Analytic provenance also has the potential to support the on-going sensemaking process, besides many post hoc applications it provides. When a user employs a system to make sense of a problem, her interaction and discovery (both are referred as provenance data) are captured and visualized for providing support back to the sensemaking process. The visualization of provenance data could provide an overview of the sensemaking process including what has been done, how has it been done, and why has it been done like that. It could show what knowledge the user has learned and offer ways to structure it meaningfully. A ~~through~~ understanding of the past process may guide the user to the next step in making sense of the task. The provenance visualization should be able to communicate with the sensemaking system to facilitate the interplay between the user and these two. Figure 1.1 illustrates this sensemaking-supporting pipeline.

1.1 Research Problem and Approach

1.1.1 Problem

In this thesis, we focus on the **visualization** part of the sensemaking-supporting pipeline as shown in Figure 1.1.

How to design interactive visualizations of provenance data for supporting sensemaking?

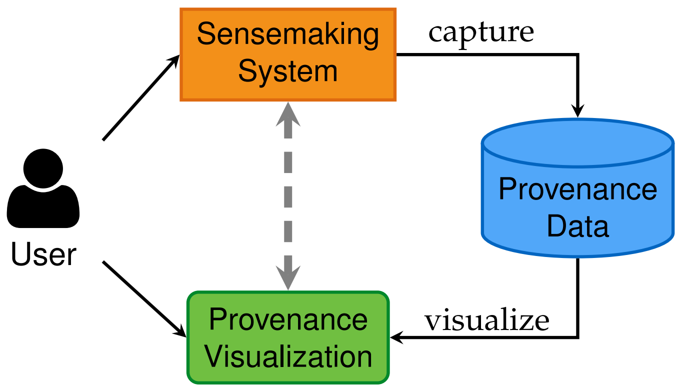

Figure 1.1: A pipeline of supporting sensemaking through analytic provenance. While employing a sensemaking system to solve a task, user interaction and discovery are captured and visualized to provide support back to the sensemaking process.

We elaborate the research question by seeking answers to the three practical questions what – why – how. This elaboration allows a clear understanding of what data will be visualized and what is the purpose of the visualization before diving into the design and implementation of such visualizations. This general approach has been used by Aigner et al. [6] in their analysis of visualization techniques for time-oriented data and by Brehmer and Munzner [17] in their characterization of visualization tasks.

1. **What is presented?**

   Based on the taxonomy of dataset types by Munzner [118], provenance data can be classified as a *table*, where each row represents an item of data and each column describes an attribute of the dataset. One essential attribute of provenance data is *time* providing when a data item is collected. A *categorical* attribute, such as *topics* from a topic modeling technique [14], may be added to help make sense of more complex relationship.

   The user may also provide additional information to the raw collected data to indicate relationships between data items. This transforms the tabular dataset to a *network* dataset.

2. **Why is it presented?**

   The characteristics of provenance data suggests the tasks it can support the users. First, the inherently temporal aspect of provenance data provides an opportunity to allow users to identify *temporal* patterns and relationships of the sensemaking task. More challengingly, can provenance data help user identify *rational* relationships of sensemaking?

3. **How is it presented?**

   The visualization design depends on the data being used and the task it aims to support, and will be discussed in depth in the corresponding chapters.

1.1.2 Approach

The research problem is broken into the following research questions based on the *what*–*why* combination elaborated previously. Note that in our text, we color code the data (what) using orange and task (why) using blue if we want to emphasize them.

1. How to design interactive visualization of time-oriented provenance data enabling users to explore temporal relationship of sensemaking?
2. How to design interactive visualization that can utilize both temporal and thematic provenance data to reveal complex temporal relationship of sensemaking?
3. How to design interactive visualization that can exploit time-oriented provenance data enabling users to explore rational relationship of sensemaking?
4. How to design interactive visualization that can utilize both temporal and relational provenance data enabling users to explore and express complex rational relationship of sensemaking?

Figure 1.2 summarizes these four research questions.

We take a user-centered design approach in seeking solutions to all the research questions. For each question, we elicit the design requirements by conducting a user study and/or drawing from the literature. Visual encoding and interaction are designed to meet those requirements, and the designs are implemented into a working prototype. Finally, an empirical study is conducted to explore how the tool is used by target audience and check whether it provides the intended support.

We aim to design general solutions that can be applied to sensemaking in various domains. However, as the solutions need to be empirically validated using target audience and real-world data, our implementation and evaluation focus on the two following domains.

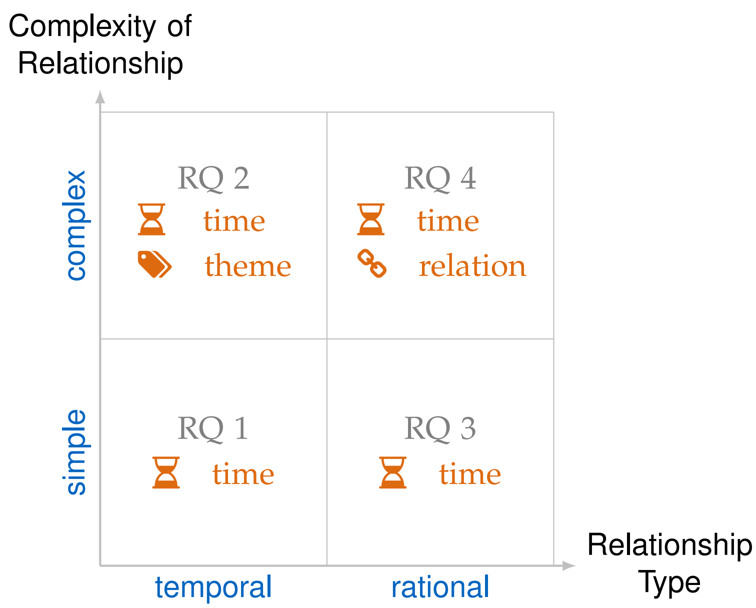

Figure 1.2: Research problem split by data and task. The horizontal axis represents different tasks and the vertical axis represents the complexity involved in the task. The cells show the characteristics of data will be used to support the task.

1. **Intelligence analysis**. This is the domain where sensemaking was introduced into the visualization community with notable sensemaking models [95, 135].
2. **Everyday, online sensemaking**. This hugely popular domain is how people usually make sense of their daily problems.

1.2 Thesis Contributions

Toward the overall goal of supporting users in their sensemaking processes through the visualizations of provenance data, this thesis contributes:

- A timeline visualization technique – *SchemaLine* – that enables users to examine information in chronological order, identify temporal patterns and construct narratives from relevant user annotations. SchemaLine produces a compact but aesthetically pleasing layout and a set of fluid interactions allowing users to perform various sensemaking activities described in the Data–Frame model [95]. This is to address Research Question 1.
- A timeline visualization technique – *TimeSets* – that enables users to explore complex temporal relationship by effectively representing both temporal and

thematic provenance data. TimeSets visually groups data items that share the same theme but still preserves their temporal order. It color codes the backgrounds of the entire themes to distinguish them and uses colored gradient backgrounds for the intersections among those themes. It also adjusts the level of details of each data item dynamically to accommodate more items within a given display estate. This is to address Research Question 2.

- Qualitative research methods are often used in understanding rational relationship of sensemaking. This is a manual and time-consuming process: researchers collect observation data, transcribe screen capture videos and think-aloud recordings, identify recurring patterns, and eventually abstract the sensemaking process into a general model. We contribute a visual sensemaking tool – *SensePath* – that offers an alternative and possibly faster approach in performing transcription and coding. In stead of having to transcribe the video, SensePath automatically captures and detects participant's sensemaking actions, and provides multi-linked visualizations to support further analysis. It visualizes provenance data in a timeline that enables researchers to quickly gain an overview of the sensemaking process and identify recurring sensemaking patterns. It also links with a screen capture video to allow researchers to examine additional context when necessary. Finally, to enable researchers to continue working on later stages of analysis using their normal workflow, SensePath exports its coded transcript in a common format that can be used by other popular qualitative data analysis software packages. This is to address Research Question 3.

- A visual sensemaking tool – *SenseMap* – that enables users to explore and express complex rational relationship of sensemaking. It automatically captures and detects sensemaking actions and relationships between these actions before visualizing both of them in a branching history tree. This allows users to examine the rational relationship between the actions they performed and potentially helps them remind of what have been done earlier. SenseMap offers users to assign additional meaning to the automatically collected data by spatially grouping actions or adding rational links between them, in order to help explain complex relationship. Finally, SenseMap allow users to communicate their analysis results at different levels of granularity including a big picture of user-organized findings, a more detailed analysis process and raw provenance data captured. This is to address Research Question 4.

1.3 Related Publications

1.3.1 Primary Publications

The following publications contain the core content presented in this thesis, from Chapter 3 to Chapter 6.

- **P. H. Nguyen**, K. Xu, R.Walker, and B. L.W.Wong. SchemaLine: Timeline Visualization for Sensemaking. In *International Conference on Information Visualization*, pages 225–233. IEEE, jul 2014.

  This is the main content of Chapter 3.

- **P. H. Nguyen**, K. Xu, R. Walker, and B. L. W. Wong. TimeSets: Timeline visualization with set relations. *Information Visualization*, 15(3):253–269, jul 2016.

  This is the main content of Chapter 4.

- **P. H. Nguyen**, K. Xu, A. Wheat, B. L. W. Wong, S. Attfield, and B. Fields. SensePath: Understanding the Sensemaking Process through Analytic Provenance. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):41–50, jan 2016.

  This is the main content of Chapter 5.

- **P. H. Nguyen**, K. Xu, A. Bardill, S. Betul, K. Herd, and B. L.W.Wong. SenseMap: Supporting Browser-based Online Sensemaking through Analytic Provenance. In *IEEE Conference on Visual Analytics Science and Technology*, 2016.

  This is the main content of Chapter 6.

1.3.2 Secondary Publications

The following publications also contribute to the ideas about provenance presented in this thesis.

- R. Walker, A. Slingsby, J. Dykes, K. Xu, J. Wood, **P. H. Nguyen**, D. Stephens, B. L. W. Wong, and Y. Zheng. An extensible framework for provenance in human terrain visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2139–2148, dec 2013.

- K. Xu, S. Attfield, T. J. Jankun-Kelly, A. Wheat, **P. H. Nguyen**, and N. Selvaraj. Analytic provenance for sensemaking: a research agenda. IEEE Computer Graphics and Applications, 35(3):56–64, jan 2015.

- K. Xu, **P. H. Nguyen**, and B. Fields. Visual analysis of streaming data with SAVI and SenseMAP. In *IEEE Conference on Visual Analytics Science and Technology*, pages 389–390. IEEE, oct 2014.

1.4 Thesis Outline

The remainder of this thesis is organized as follows.

First, Chapter 2 reviews the core work related to sensemaking, analytic provenance and visualization. Then, it emphasizes on the visualization of provenance data for supporting sensemaking. At the end, this chapter presents visualization techniques of general time-oriented and network data because these types of data have similar characteristics as provenance data.

Chapter 3 discusses the SchemaLine timeline visualization of user annotations enabling the users to explore temporal relationship of sensemaking – addressing Research Question 1.

Chapter 4 extends Chapter 3 to present the TimeSets visualization technique that can effectively show both temporal and thematic provenance data in order to reveal complex temporal relationship of sensemaking – addressing Research Question 2.

Chapter 5 discusses the SensePath visualization tool that can exploit time-oriented provenance data, enabling users to explore rational relationship of sensemaking – addressing Research Question 3.

Chapter 6 describes the SenseMap visualization tool that can utilize both temporal and relational provenance data enabling users to explore and express complex rational relationship of sensemaking – addressing Research Question 4.

Finally, **??** concludes the thesis with a discussion on its contributions and future research directions triggered from this work.