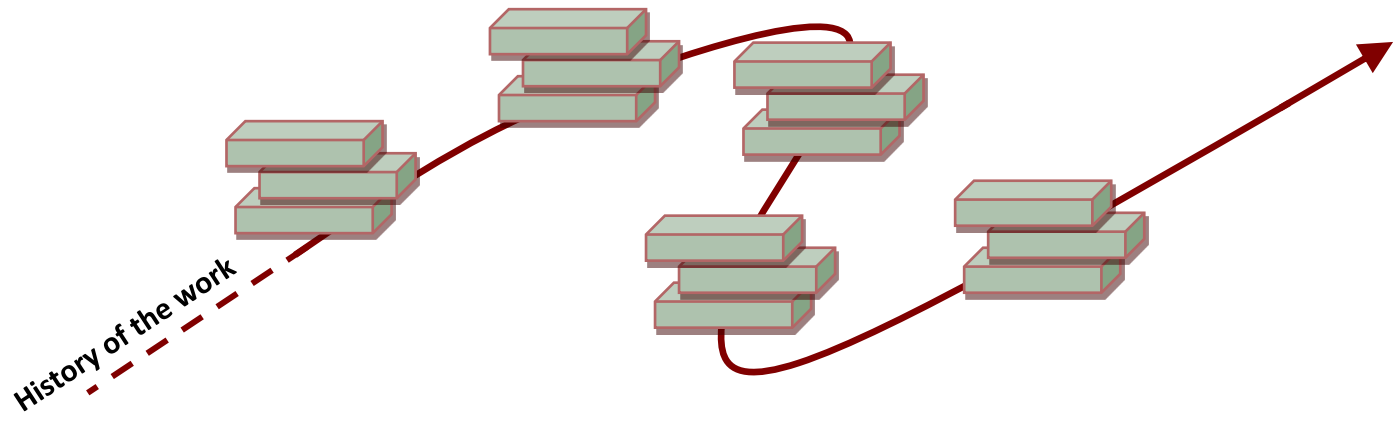


ProvThreads: Analytic Provenance Visualization and Segmentation

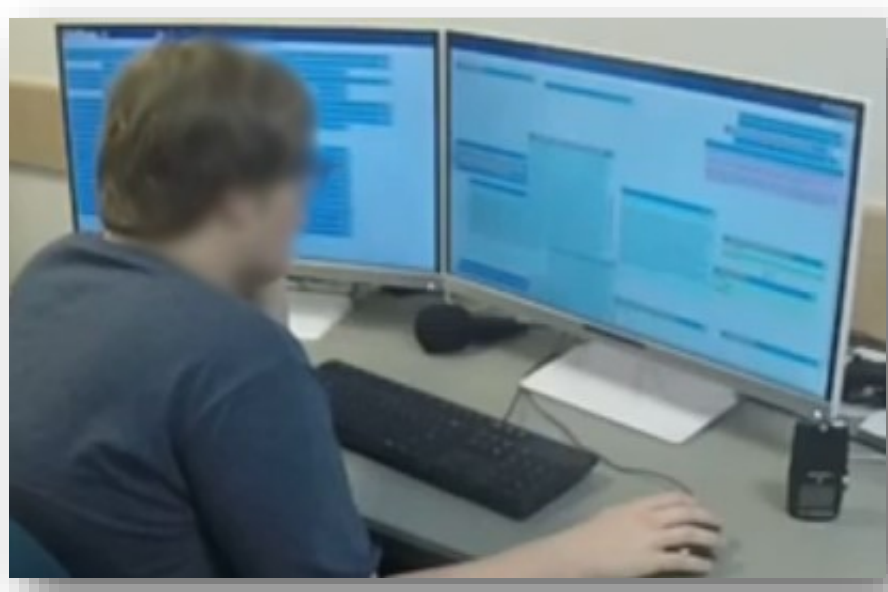
Sina Mohseni*, Alyssa Pena[†], Eric Ragan*[†]

*Dept. of Computer Science and Engineering, [†]Dept. of Visualization, Texas A&M University.



ANALYTIC PROVENANCE

- ⇒ Analytic provenance visualization provides an overview of analysts' actions and insights during the exploratory data analysis.
- ⇒ Our work aims to visually summarize interaction behaviors and show how different interactions are used to explore different data topics.



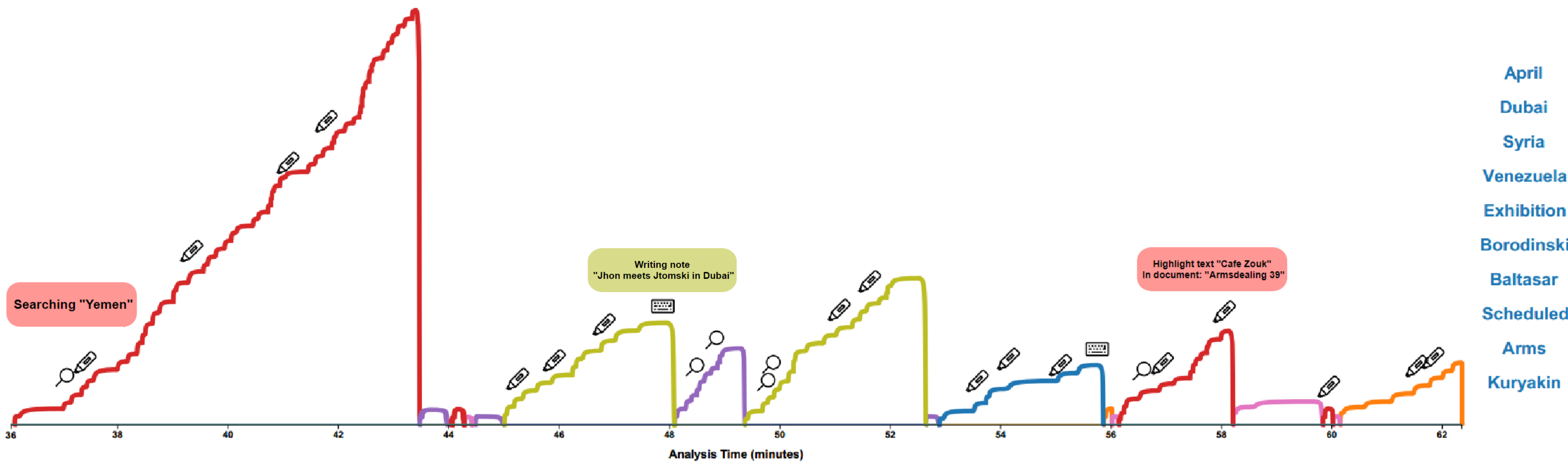
CAPTURE PROVENANCE DATA

- ⇒ We conducted a set of 24 user studies using text analysis scenarios to provide a provenance test data. Our provenance dataset is available online for research purposes.
- ⇒ A text explorer tool logged user interactions (*interaction provenance*) along with the sequence of text documents (*data*)



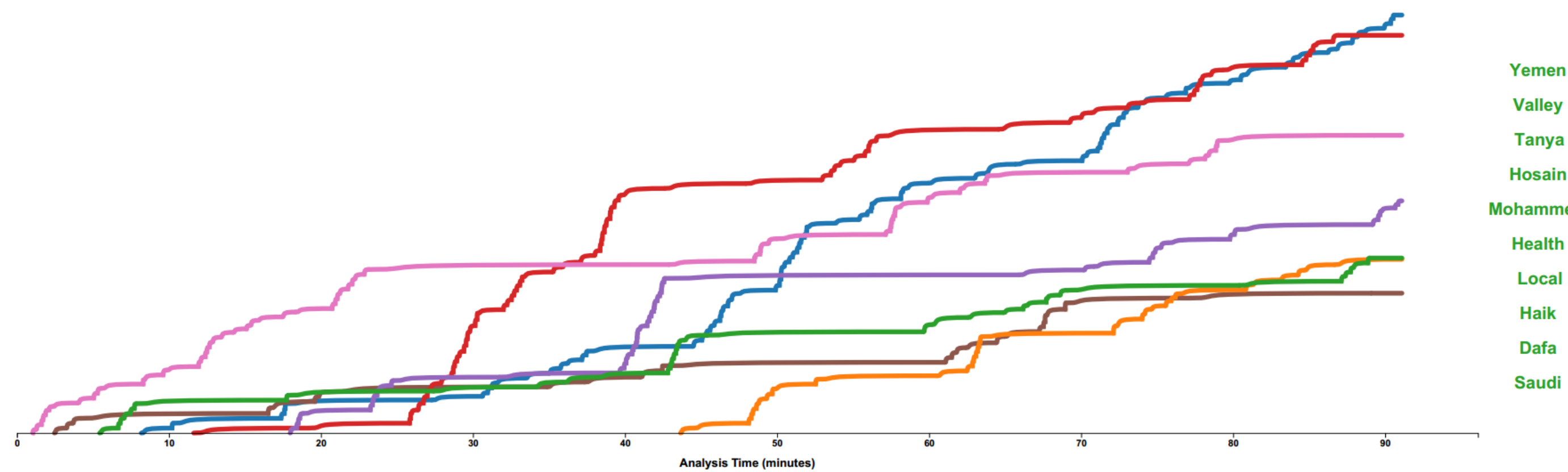
TOPIC MODELING

- ⇒ We use topic models (Latent Dirichlet Analysis modeler) to classify text documents and summarize topics.
- ⇒ In an interactive approach, user interactions can influence topic models based on analytical reasoning associated with each interaction.



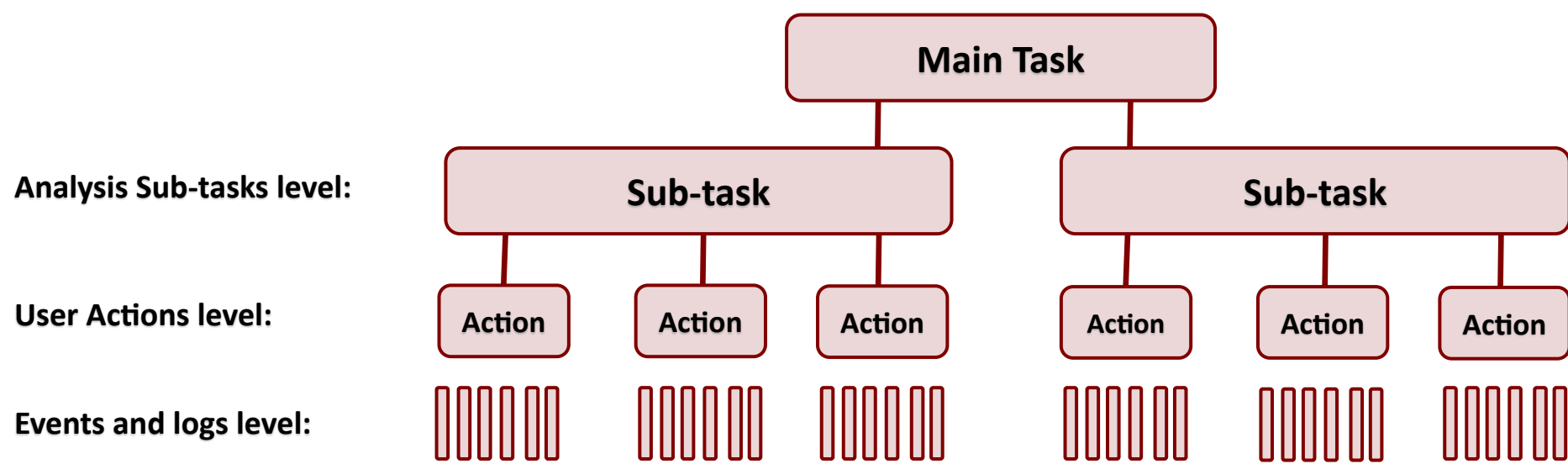
PROVENANCE SEGMENTS

- ⇒ This visualization is designed to help meta-analysts understand topic transitions in analysis history, such as when a user leaves a topic and later returns to the same topic at a later time. The tool's multiple configurations and interactive features aid in the understanding of analysis strategies.
- ⇒ In our design, we focused on simplifying complex high-dimensional temporal event sequence data. The approach segments provenance records based on user interaction with different topics in data to identify topic changes during the analysis.
- ⇒ This visualization embeds user interaction history into the vertical increases in topic threads, meaning that each type of interaction increases threads height with a predefined value. User action icons that encode interaction types and mouse over threads gives details about users actions in a tooltip.



TOPIC COVERAGE

- ⇒ ProvThreads projects a series of analysis paths to demonstrate both topic coverage and the progression of an investigation over time. Each colored thread represents a topic and user interactions increase the topic threads height at the time of events.
- ⇒ We use word lists to visualize a summary of topics. Users are able to merge and split topics in this system. Topic merging allows users to find the preferred number of topics for the text documents.



CONCLUSION

- ⇒ This work is an effort to determine user sub-tasks from captured user interactions and under process data.
- ⇒ We ran a series of user studies to generate set of provenance test data from textual intelligence analysis scenarios. (datasets publicly available now!)
- ⇒ We present a new visual analytics tool and segmentation method for analytic provenance records.
- ⇒ Our goal is to enable meaningful meta-analyses of analytic processes.