

Filter and Compare: Exploring Differences in Event Log Subsets

Liliia Aliakberova, Francesca Zerbato, Dirk Fahland

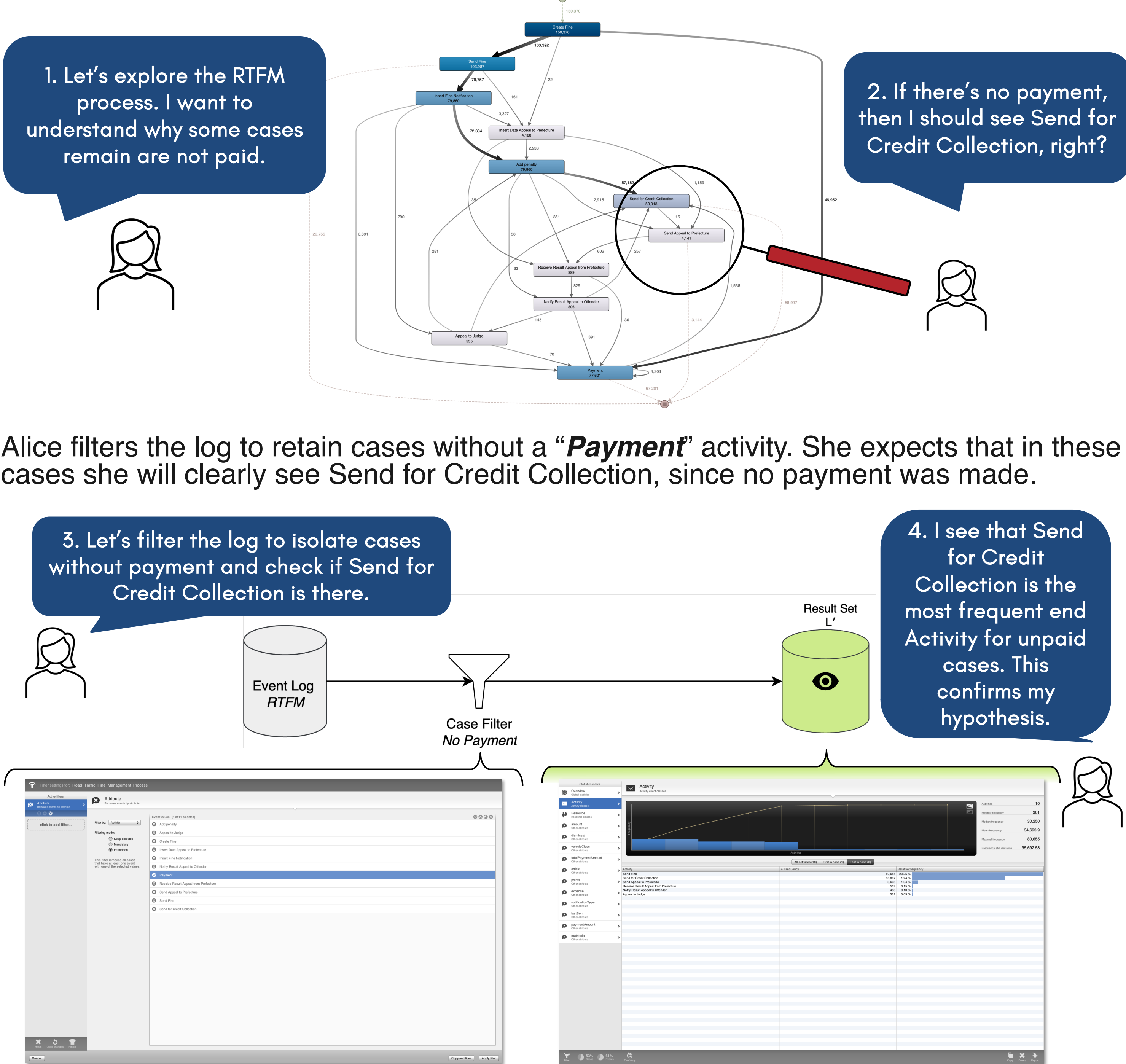
Department of Mathematics and Computer Science, Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands

MOTIVATION

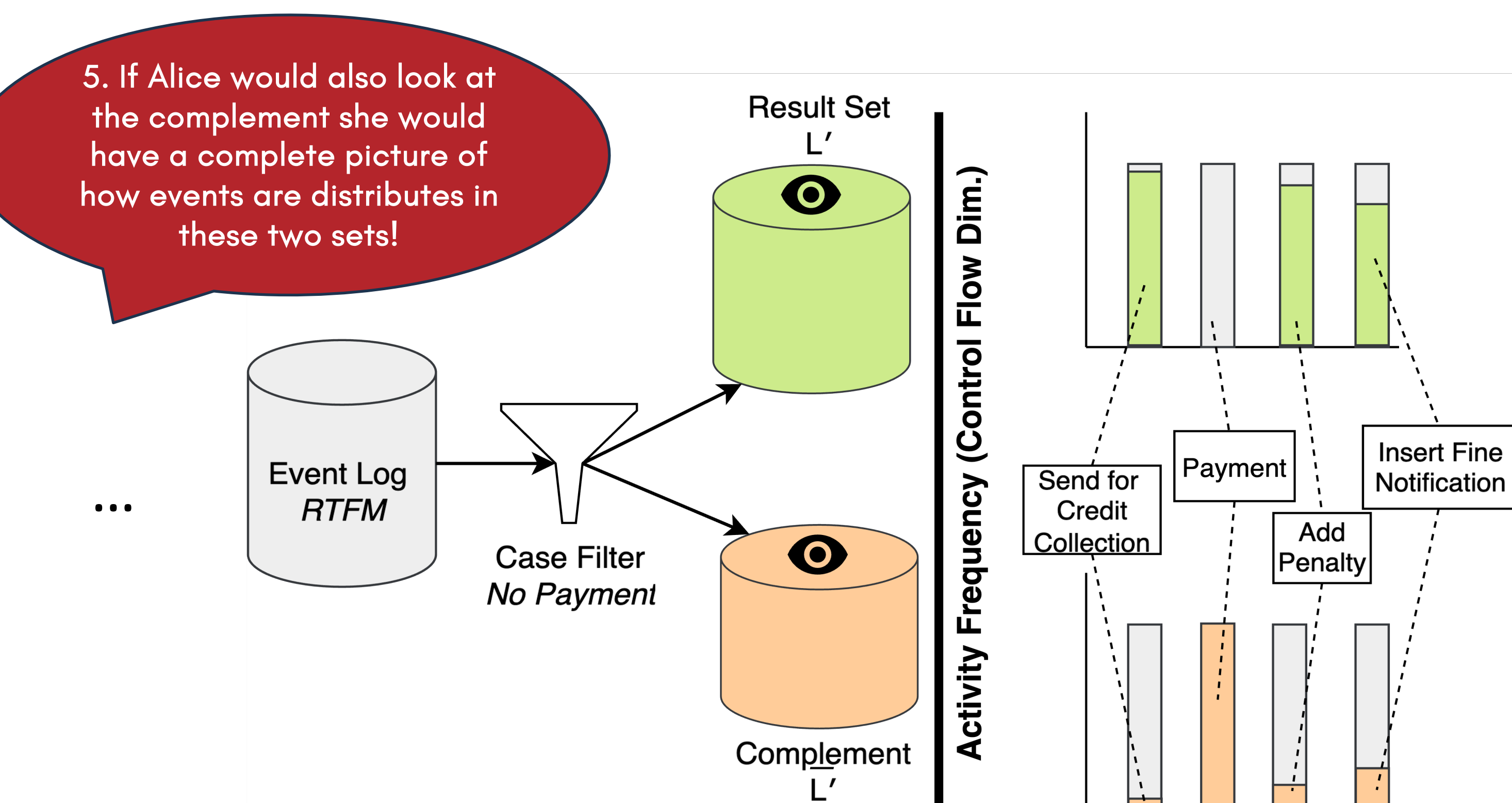
In **exploratory** settings, process analysts often start with limited knowledge of the event log. They use statistics, visualizations, and filters to explore the log and generate hypotheses. Often, testing hypotheses requires checking if conditions on the data are met using *filters*. Filters split the log into a **result set** (cases that meet the condition) and a **complement set** (cases that do not). In many tools, only the result set is shown to the analyst: this can lead to missing out on interesting patterns that manifest in the complement set. To address this problem, we propose comparing and visualizing both sets to uncover insights that go beyond the conditions expressed in the filter.

AN EXAMPLE PROCESS MINING ANALYSIS

Meet Alice, a process analyst exploring the Road Traffic Fine Management (RTFM) log. She wants to find scenarios and reasons why some fines are not paid. She notices the activity **“Send for credit collection”**, which likely marks unpaid cases.



Focusing only on unpaid cases, Alice links “Send for Credit Collection” to non-payment, **However, this interpretation is incomplete.** By ignoring the complement, she overlooks the presence of Send for Credit Collection also in paid cases.



This side-by-side comparison of the **result set** and its **complement** reveals these hidden differences. Through this combined view, Alice can better interpret the process and refine her hypotheses.

RESEARCH QUESTION

? How can we compare a result set, obtained after filtering the event log, with its complement to support process mining analysts in their exploration?

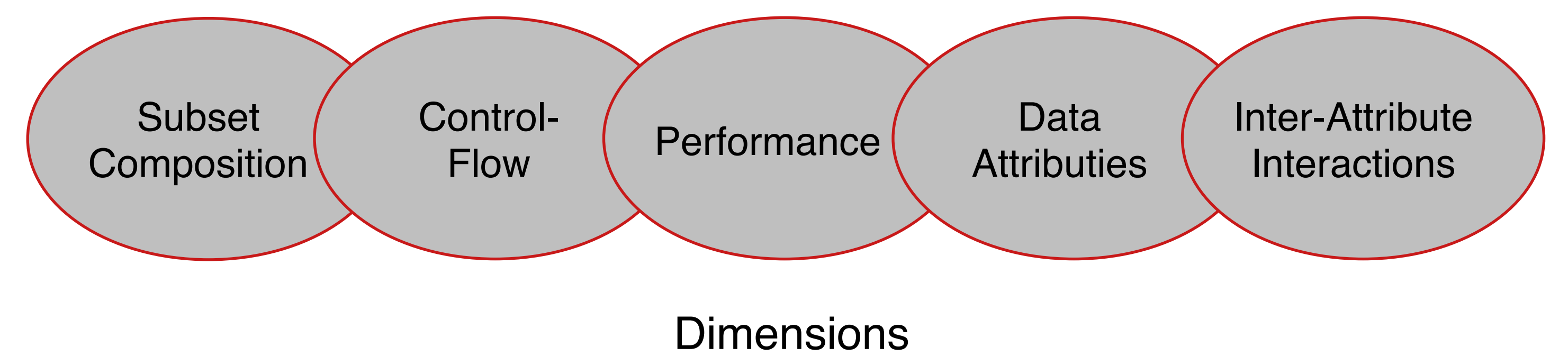
RELATED LITERATURE

- Zerbato, F., Soffer, P., & Weber, B. (2022, September). Process mining practices: Evidence from interviews. In International Conference on Business Process Management(pp. 268-285). Cham: Springer International Publishing.
- Klinkmüller, C., Müller, R., & Weber, I. (2019, July). Mining process mining practices: an exploratory characterization of information needs in process analytics. In International Conference on Business Process Management (pp. 322-337). Cham: Springer International Publishing.
- Zerbato, F., Franceschetti, M., & Weber, B. (2024, August). A Framework to Support the Validation of Process Mining Inquiries. In International Conference on Business Process Management (pp. 249-266). Cham: Springer Nature Switzerland.

HOW TO COMPARE?

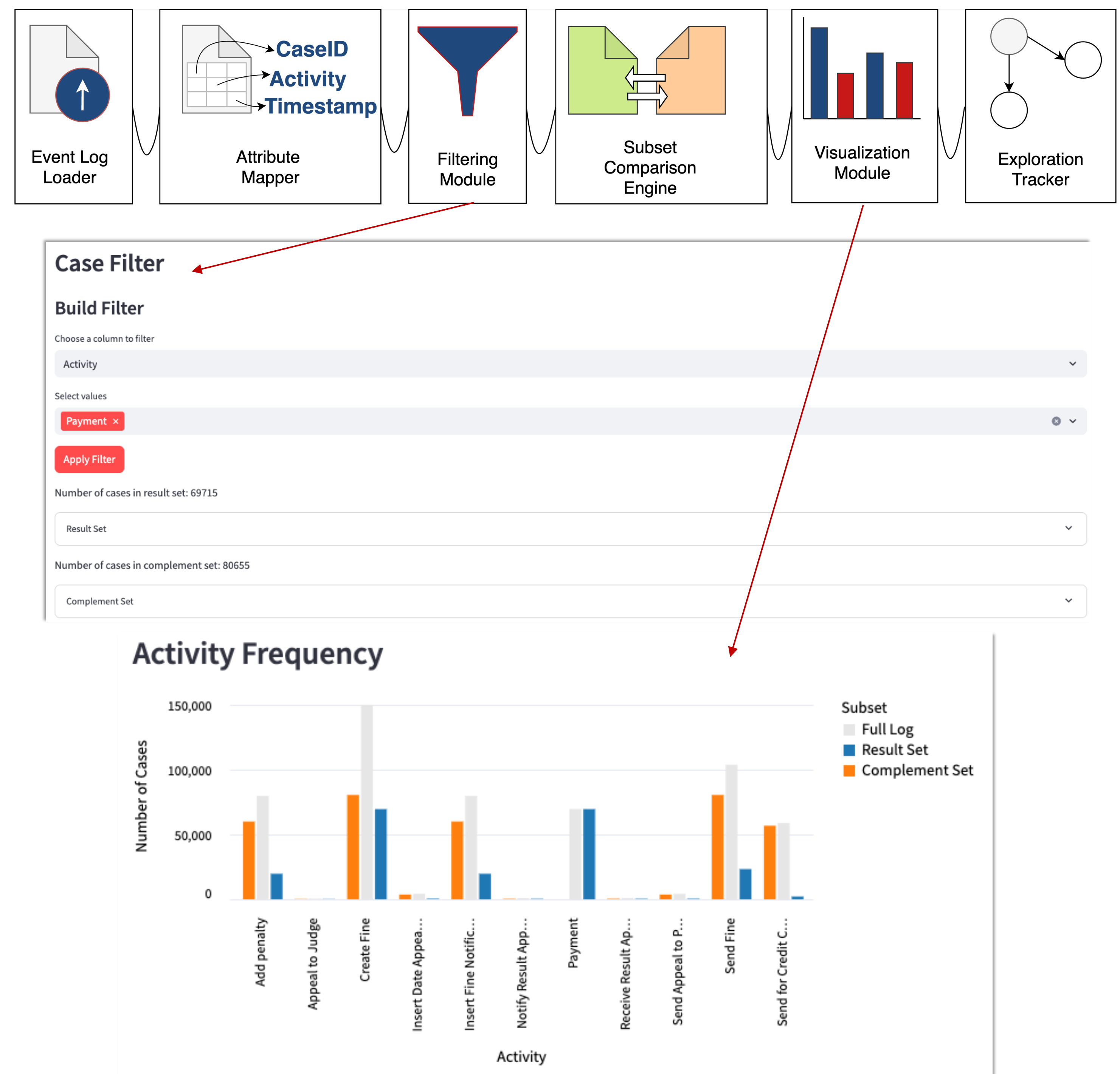
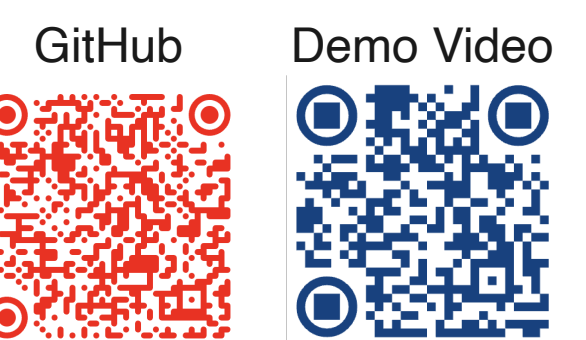
To derive dimensions for comparing a result set and its complement, we reflected on aspects that analysts typically examine when exploring event logs. Based on literature, empirical insights, and hands-on experience with filtering in PM tools, we defined key requirements for **comparison dimensions**, which must be:

- Easy to interpret
- Efficient to compute
- Relevant to the exploration goal



PROOF OF CONCEPT IMPLEMENTATION

We have implemented a baseline tool in Python to compare subsets of an event log. The tool includes several core components organized into a modular architecture.



CONCLUSION

- We introduced a challenge in exploratory process mining concerning the use of filters: analysts often focus only on the filtered result set, overlooking the characteristics of its complement.
- To address this, we proposed a subset comparison approach that considers both result and complement sets across multiple dimensions.
- We developed a prototype tool that enables interactive comparison of subsets, serving as a first step toward multi-perspective subset analysis.

FUTURE WORK

- Automated difference measures (detect and highlight interesting subset differences).
- Comparison visualization (richer, multi-dimensional views for better insight).
- Extend our approach to multi-stage comparison (support comparisons across subsets from different filtering stages).
- Evaluation (apply our approach on real event logs and conduct user studies with PM analysts to assess effectiveness).