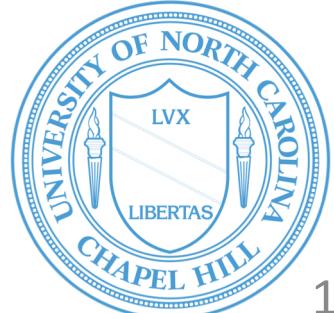


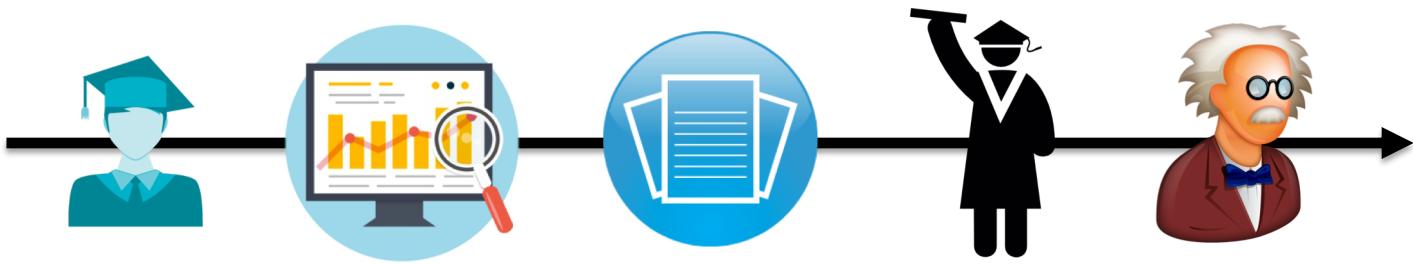
EventThread: Visual Summarization and Stage Analysis of Event Sequence Data

**Shunan Guo, Ke Xu, Rongwen Zhao,
David Gotz, Hongyuan Zha, Nan Cao**



Event Sequence Data

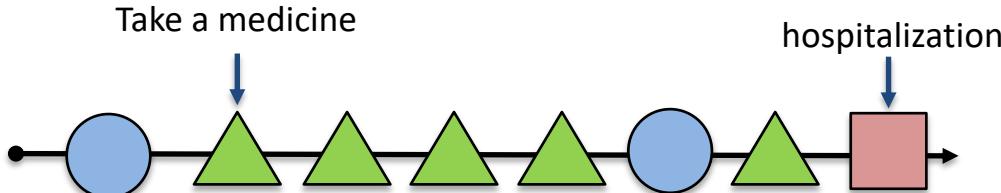
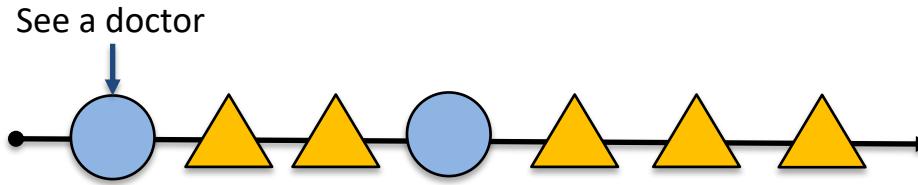
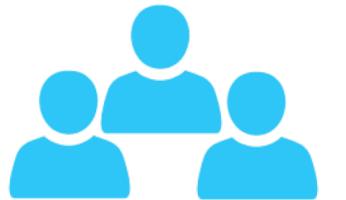
Academic
Record



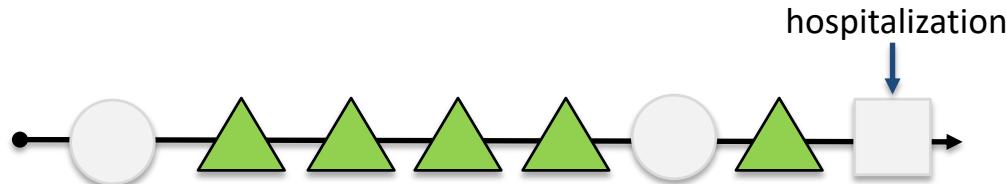
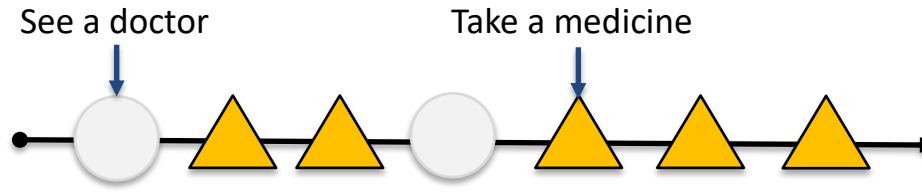
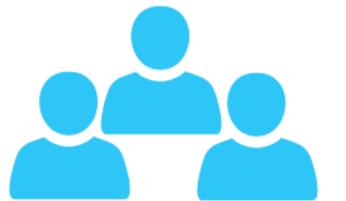
Electronic
Health
Record



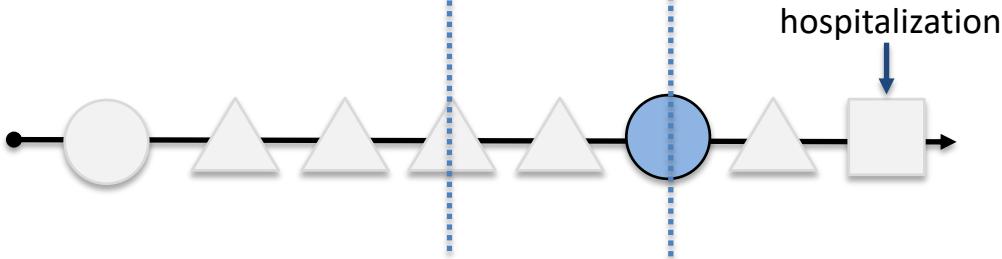
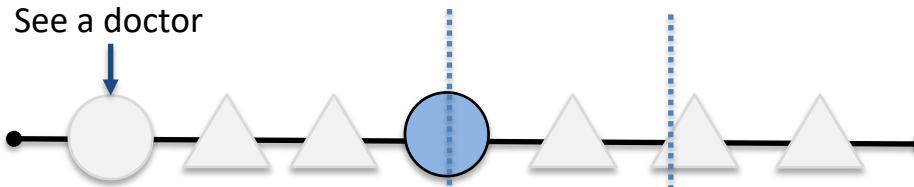
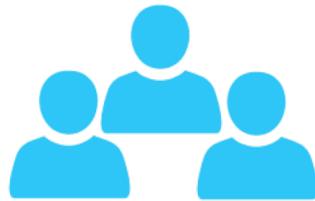
Electronic Health Record



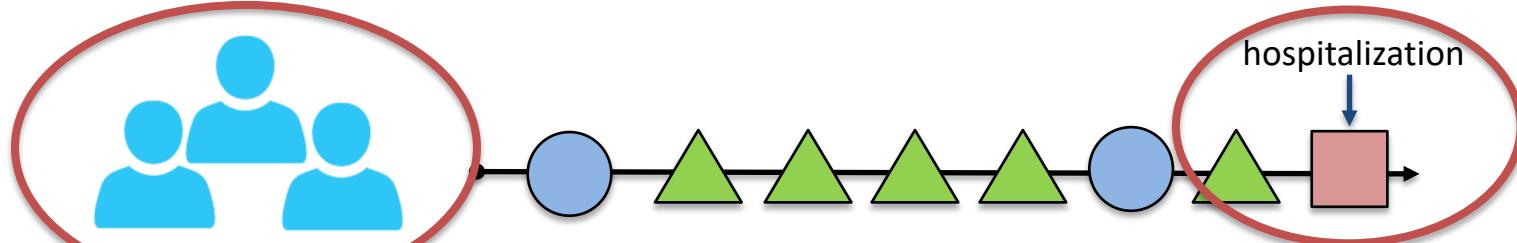
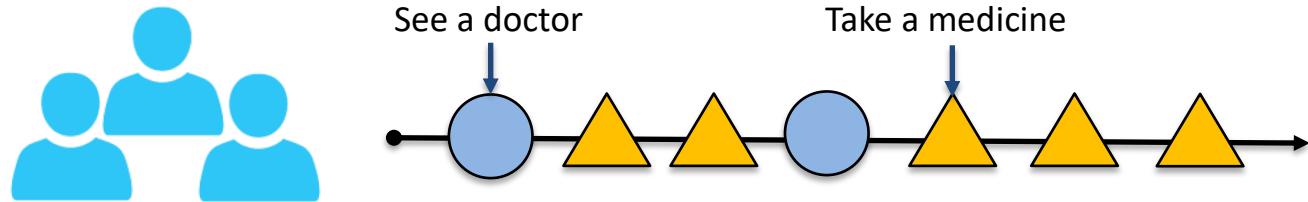
Electronic Health Record



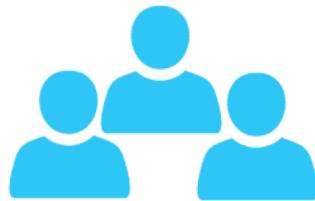
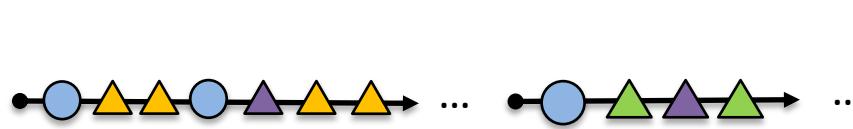
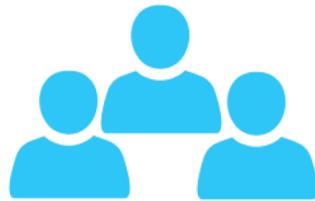
Electronic Health Record



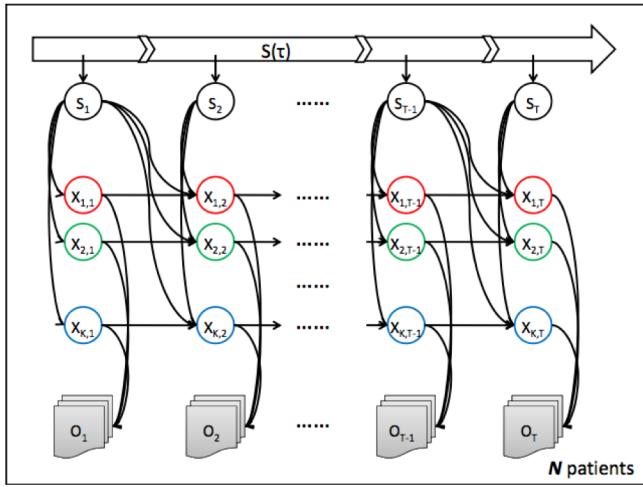
Electronic Health Record



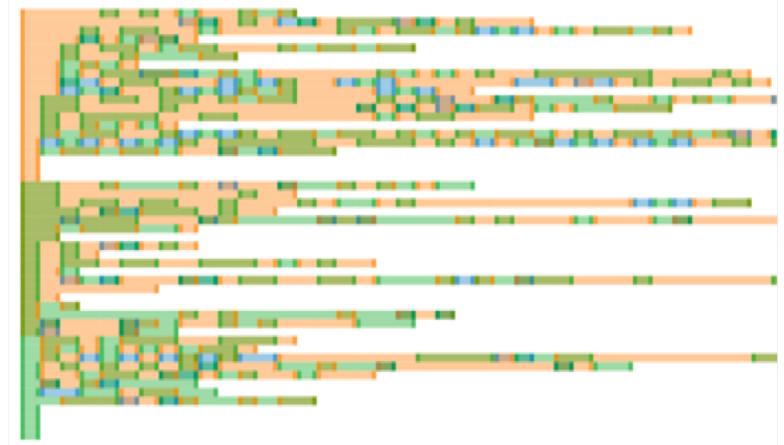
Electronic Health Record



Research Goal

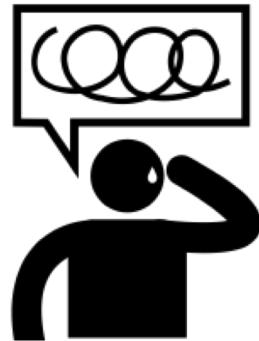


Gap

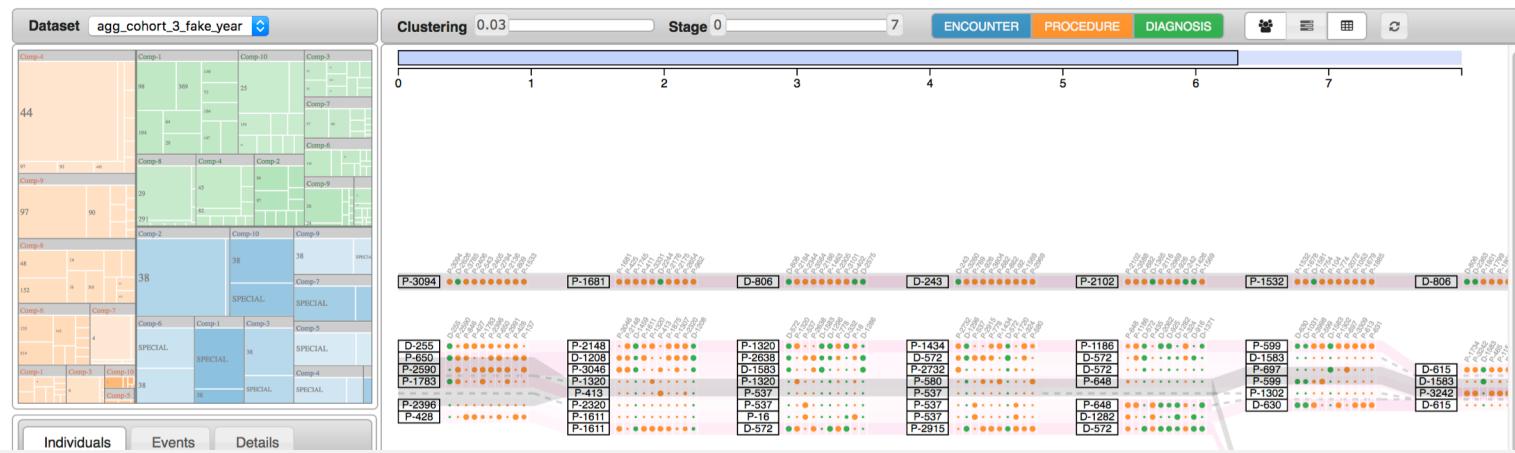


- Discover latent high-level structures of large-scale event sequence data.
- Provide users with information about the low-level events and sub-sequences of events.

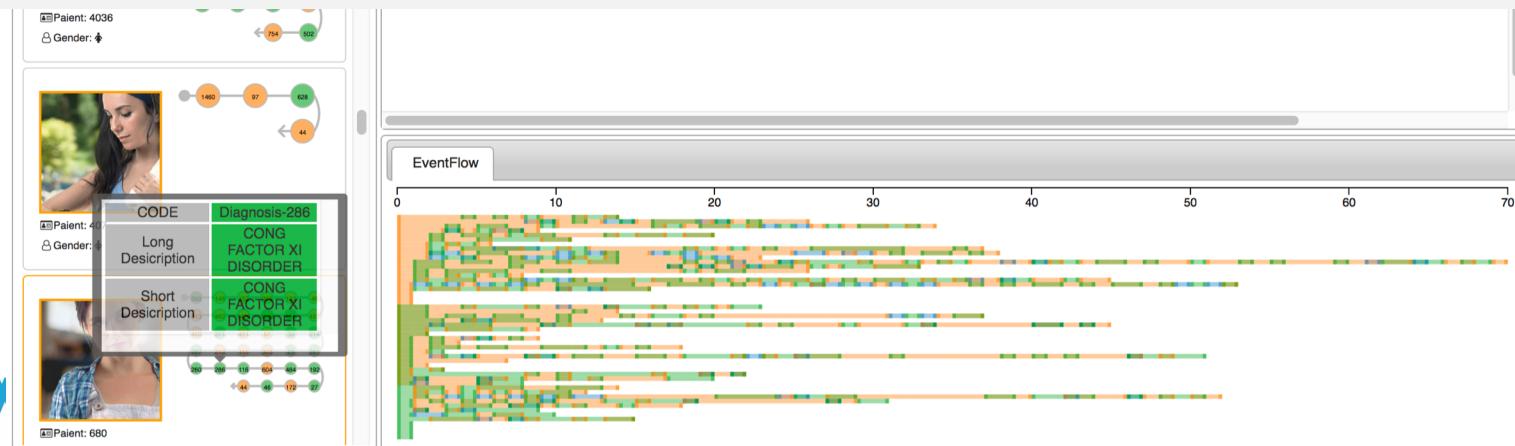
Key Challenges



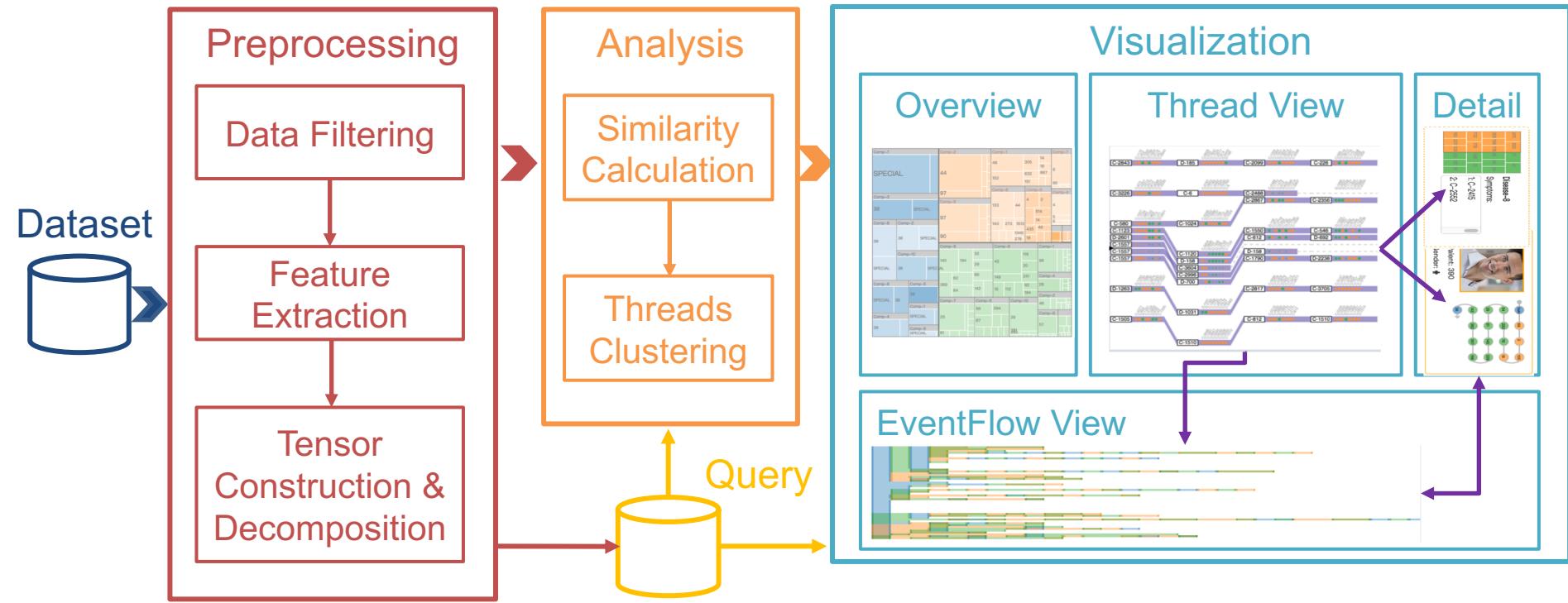
- It is difficult to transform large-scale heterogeneous event sequence data into an uniform data model without losing detailed information
- The method designed to detect high-level structures should also include sufficient relevant context to enable low-level semantic interpretation of what those structures represent
- Unavailable of ground truth to help users validate the result



EventThread System

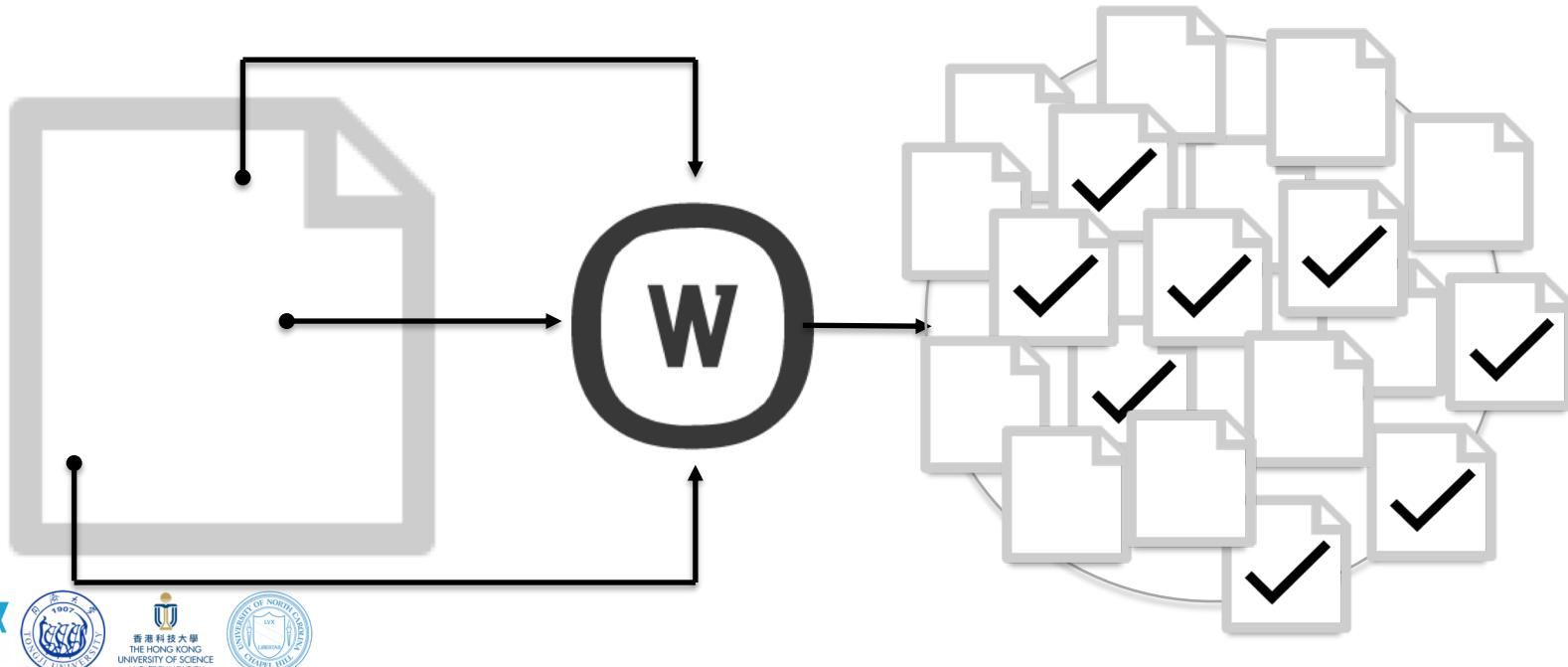


System Overview



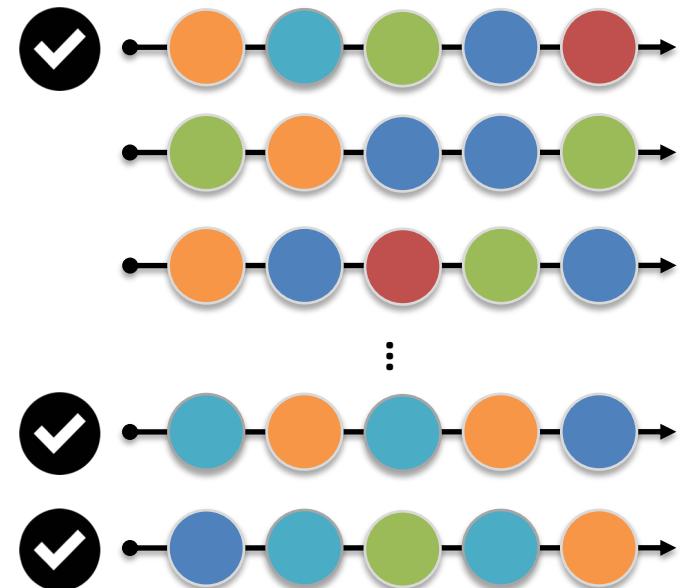
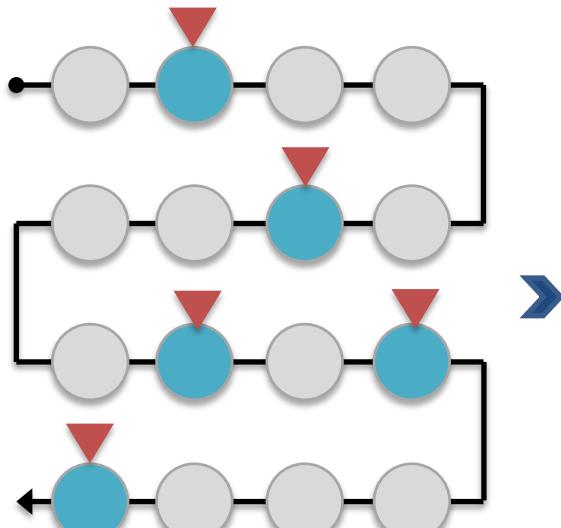
Data Filtering

- Term Frequency - Inverse Document Frequency ($TF-IDF$)

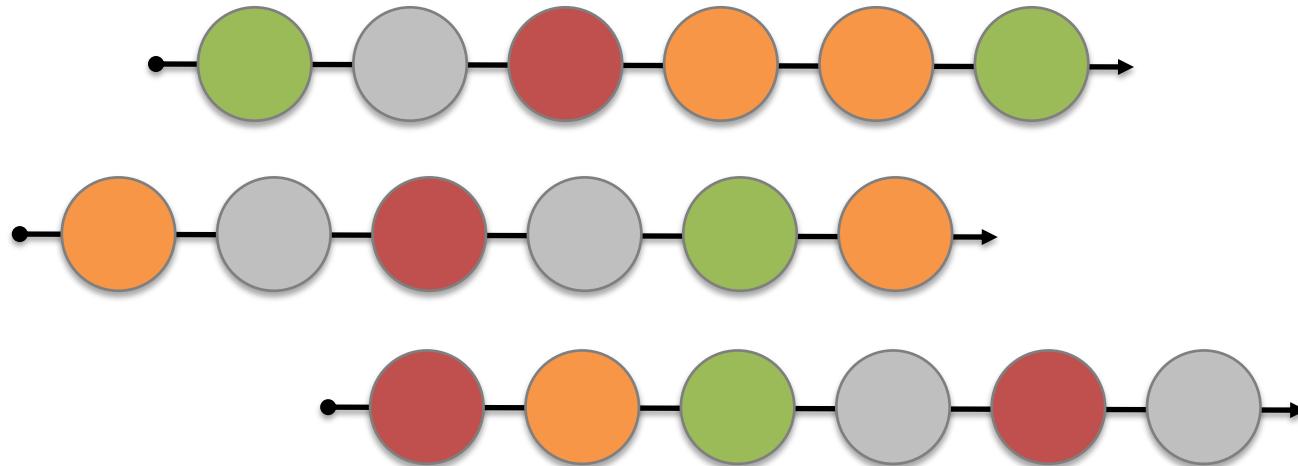


Data Filtering

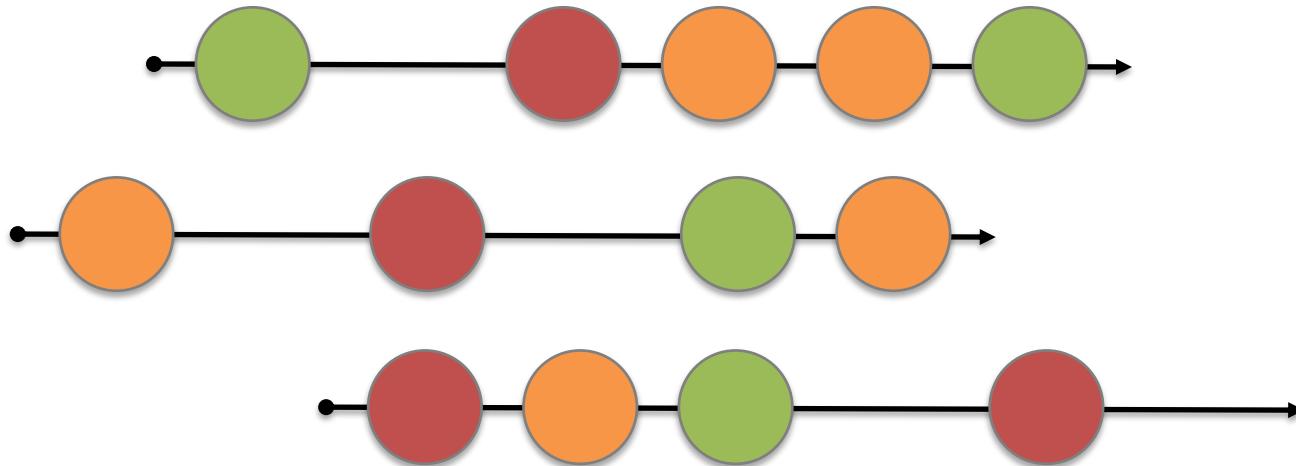
- *TF-IDF* - Event Sequences



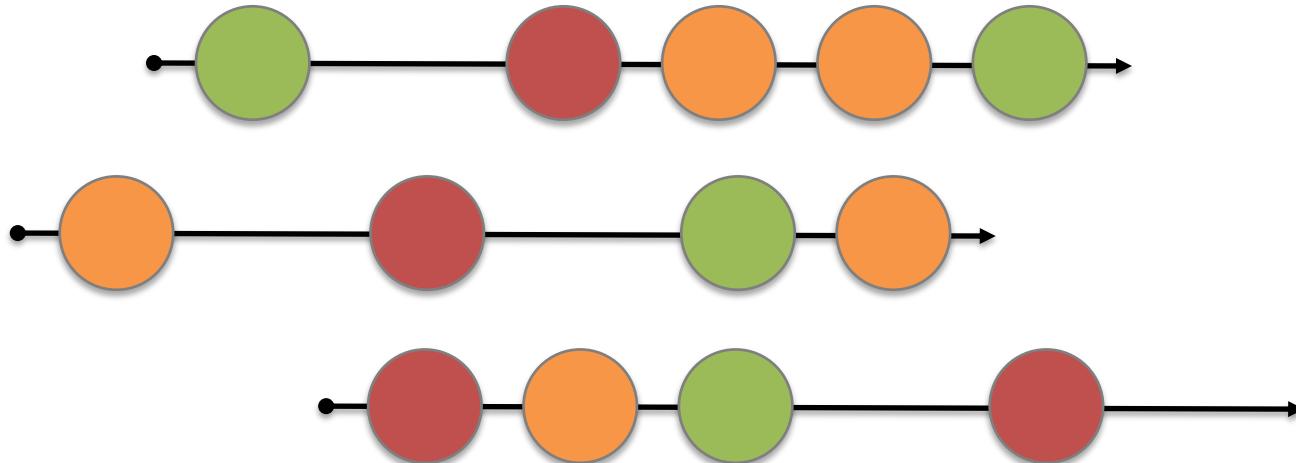
Sequence Preprocessing



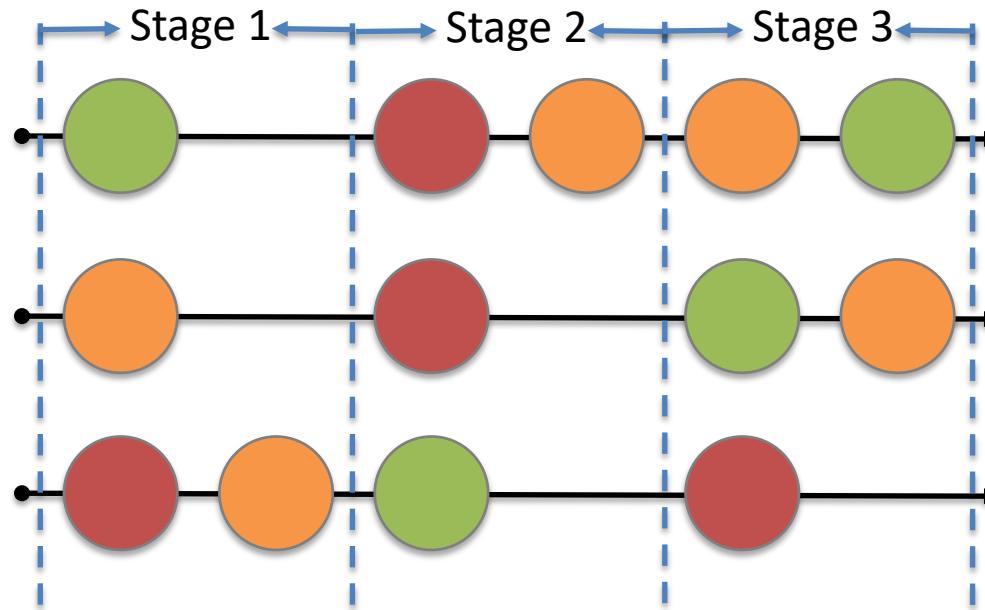
Sequence Preprocessing



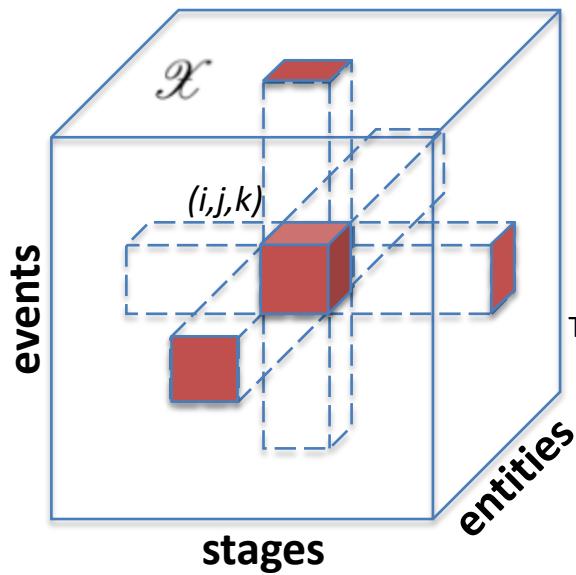
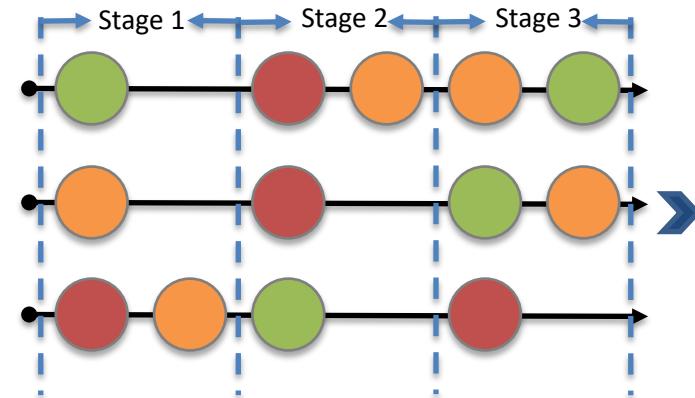
Sequence Preprocessing



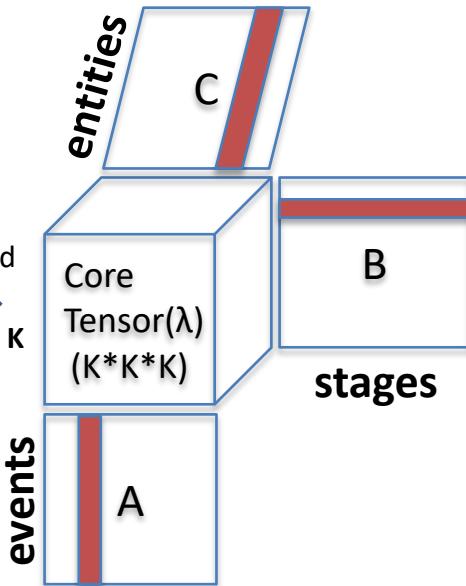
Sequence Preprocessing



Tensor Analysis



User-defined
Thread Num K

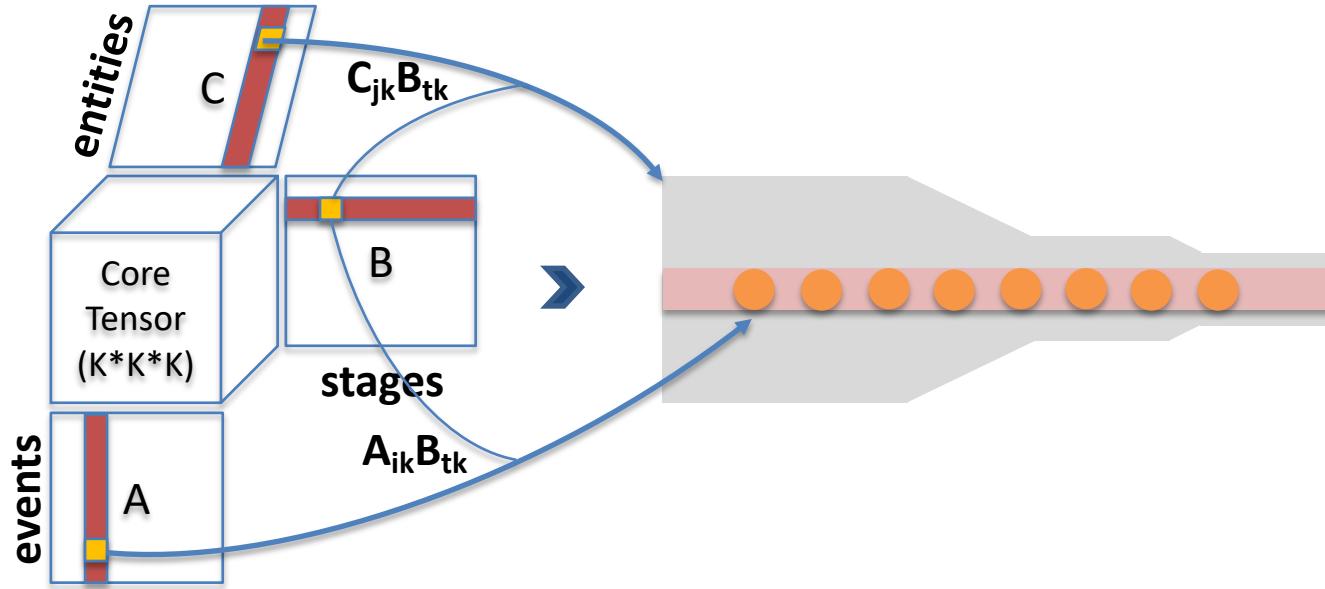


$$\min \|\mathcal{X} - [\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}] \|^2$$

Subject to: $\mathbf{B}^T \mathbf{B} = \mathbf{I}, \mathbf{A}, \mathbf{B}, \mathbf{C} \in R_+^{N, T, M} \times K$



Tensor Analysis



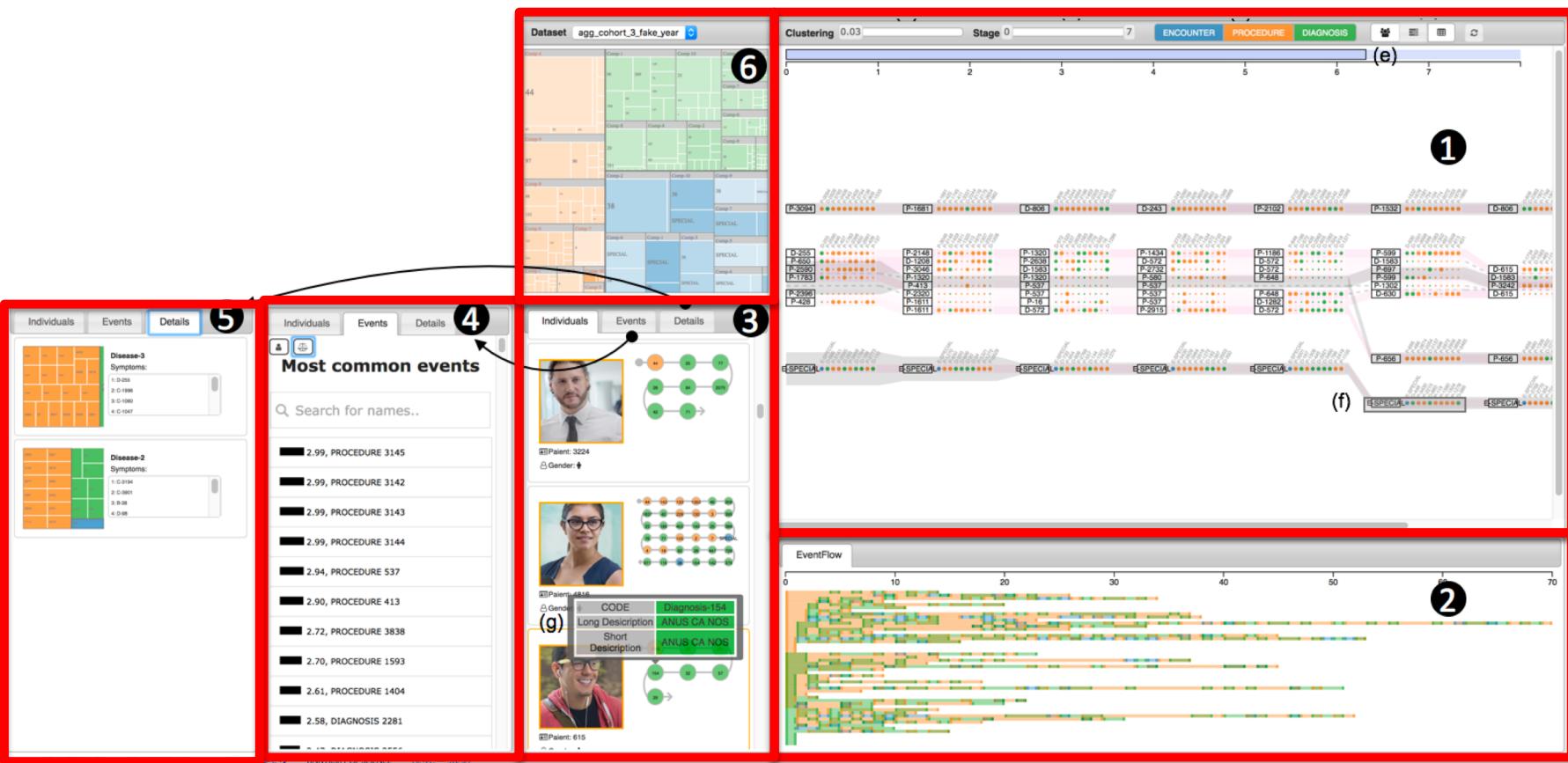
Introduction

Visualization Design

Evaluation

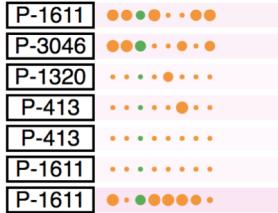


User Interface

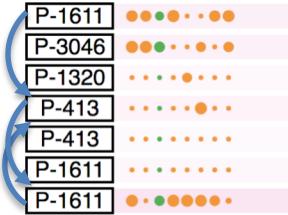


Creating the Thread View

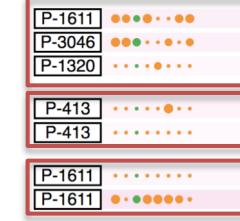
Visualize Threads



Layout Threads



Rearrange Threads



Creating the Thread View

Visualize Threads

P-1611	●●●●●●●●
P-3046	●●●●●●●●
P-1320	●●●●●●●●
P-413	●●●●●●●●
P-413	●●●●●●●●
P-1611	●●●●●●●●
P-1611	●●●●●●●●

Layout Threads

P-1611	●●●●●●●●
P-3046	●●●●●●●●
P-1320	●●●●●●●●
P-413	●●●●●●●●
P-413	●●●●●●●●
P-1611	●●●●●●●●
P-1611	●●●●●●●●

Rearrange Threads

P-1611	●●●●●●●●
P-3046	●●●●●●●●
P-1320	●●●●●●●●
P-413	●●●●●●●●
P-413	●●●●●●●●
P-1611	●●●●●●●●
P-1611	●●●●●●●●

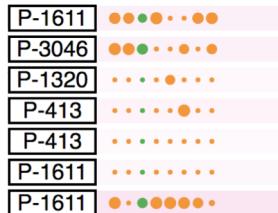
Thread →

	P-2148	P-1320	P-1434	P-1186	
P-1611	●●●●●●●●	P-2638	●●●●●●●●	D-572	●●●●●●●●
P-3046	●●●●●●●●	D-1583	●●●●●●●●	P-2732	●●●●●●●●
P-1320	●●●●●●●●	P-1320	●●●●●●●●	P-1434	●●●●●●●●
P-413	●●●●●●●●	P-537	●●●●●●●●	P-537	●●●●●●●●
P-413	●●●●●●●●	P-537	●●●●●●●●	P-537	●●●●●●●●
P-1611	●●●●●●●●	P-537	●●●●●●●●	P-648	●●●●●●●●
P-1611	●●●●●●●●	D-572	●●●●●●●●	P-648	●●●●●●●●
		D-572	●●●●●●●●	D-572	●●●●●●●●
			P-2915	D-1282	●●●●●●●●
				D-572	●●●●●●●●

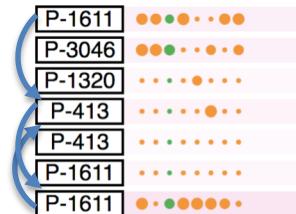


Creating the Thread View

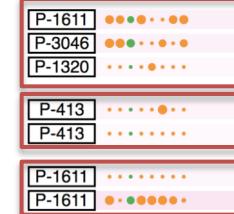
Visualize Threads



Layout Threads



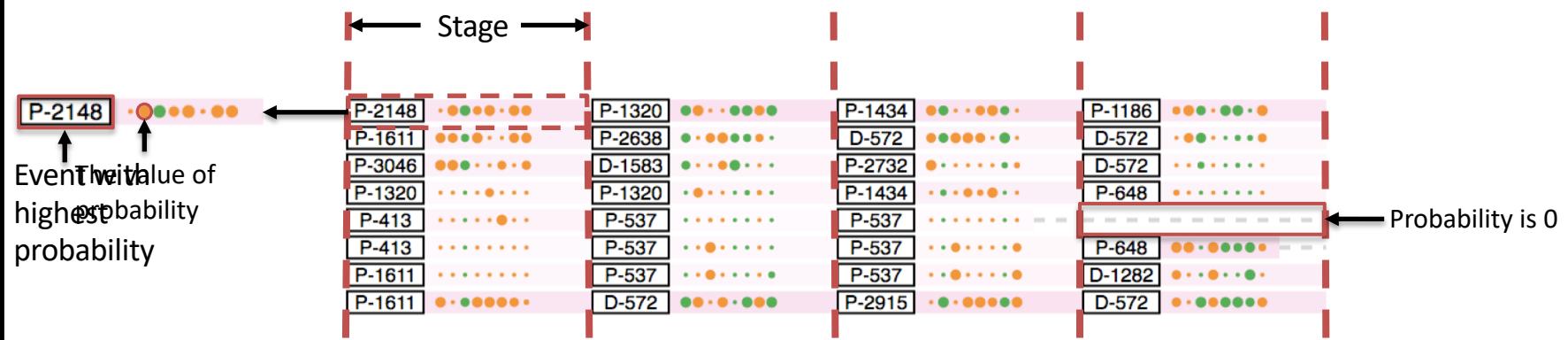
Rearrange Threads



← Stage →

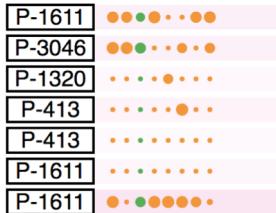
Event with the value of highest probability

Probability is 0

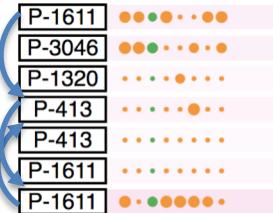


Creating the Thread View

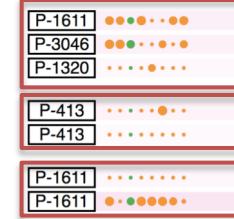
Visualize Threads



Layout Threads



Rearrange Threads



Minimize the distance
of similar threads

$$\sum_{t=0}^T \left(\alpha \sum_{i < j} w_{ij}(t) \|y_i(t) - y_j(t)\|^2 + (1 - \alpha) \sum_i \|y_i(t) - y_i(t-1)\|^2 \right)$$

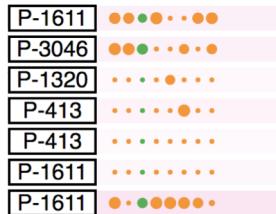
Reduce thread crossing

$$\sum_i \|y_i(t) - y_i(t-1)\|^2$$

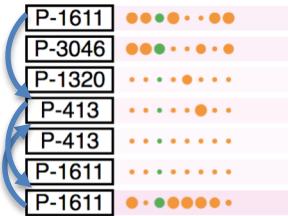
Balance Two Terms

Creating the Thread View

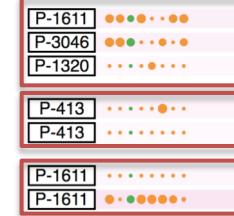
Visualize Threads



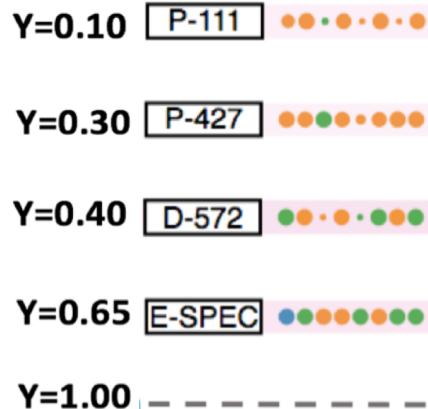
Layout Threads



Rearrange Threads

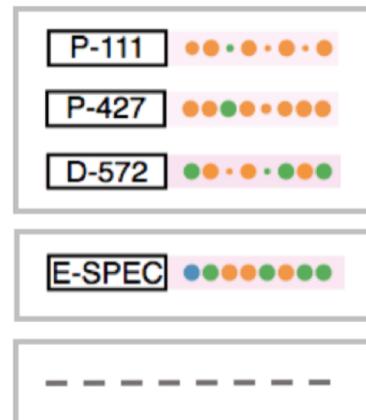


Initial layout

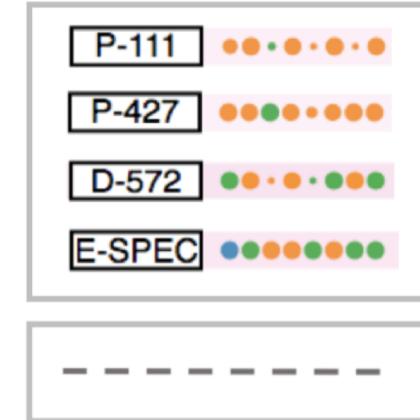


User Defined
Threshold

Cluster Threshold = 0.2

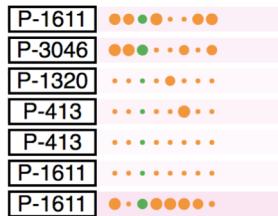


Cluster Threshold = 0.25

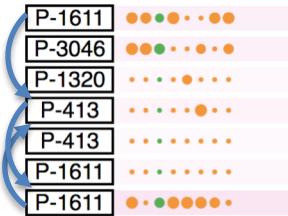


Creating the Thread View

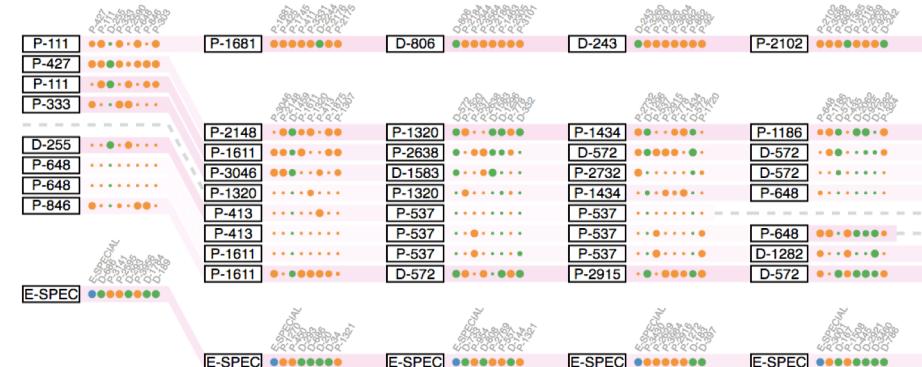
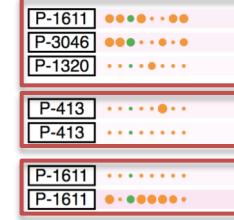
Visualize Threads



Layout Threads

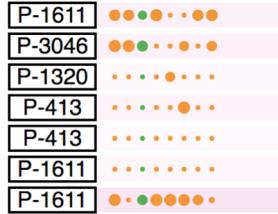


Rearrange Threads

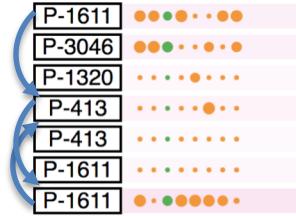


Creating the Thread View

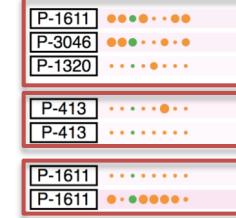
Visualize Threads



Layout Threads

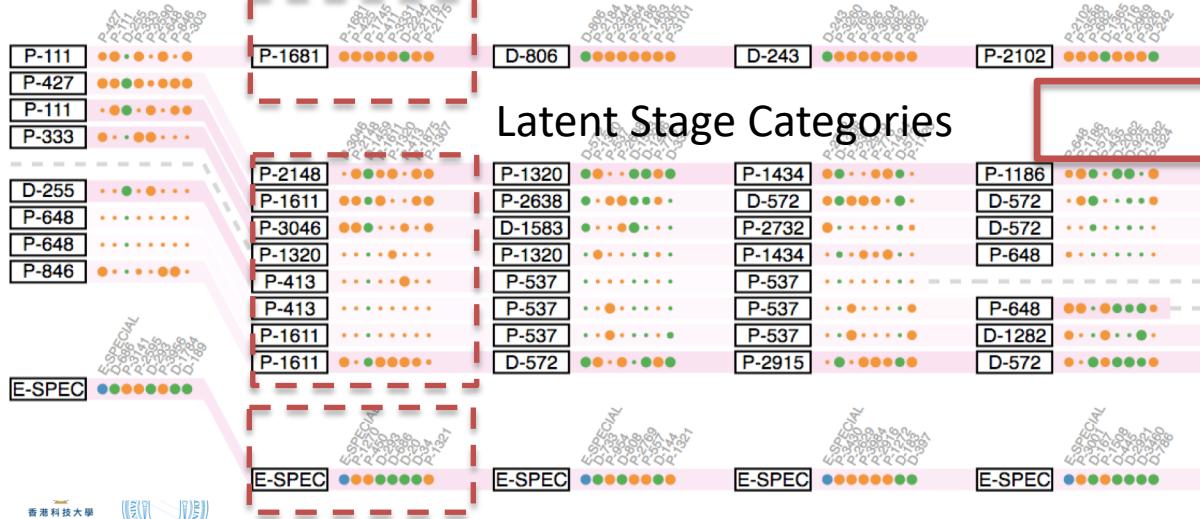


Rearrange Threads



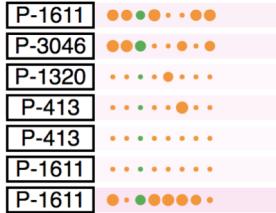
Latent Stage Categories

Category Labels

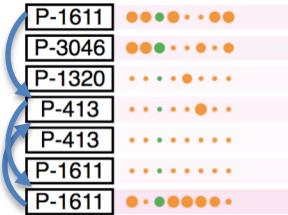


Creating the Thread View

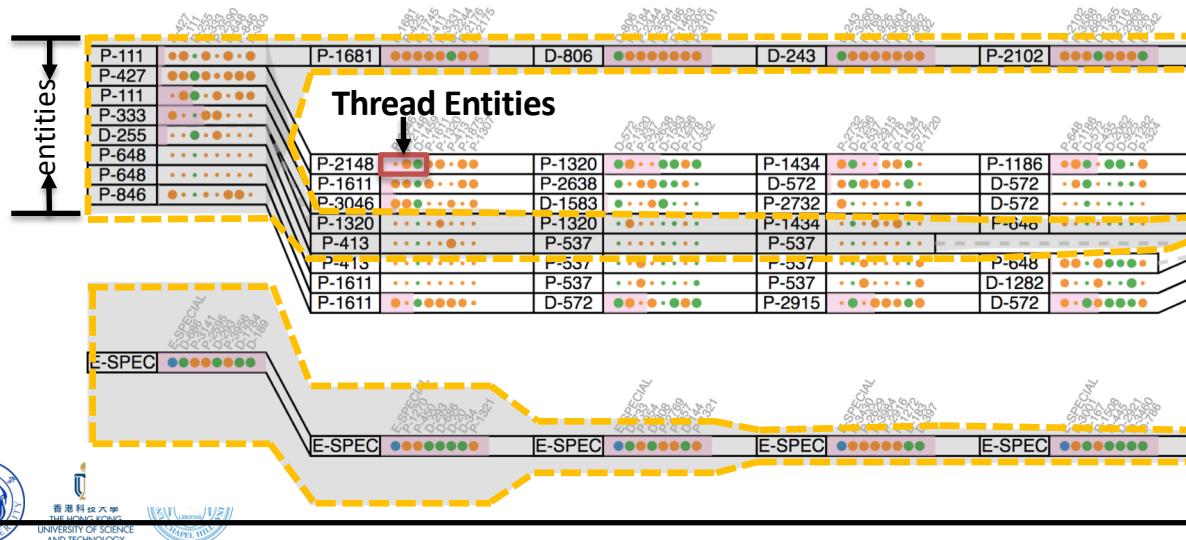
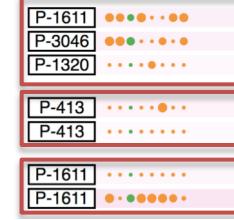
Visualize Threads



Layout Threads



Rearrange Threads



The figure shows a user interface for data analysis, likely a machine learning pipeline visualization tool.

Dataset: u2_professor_year

Clustering: 0 Stage 0 22

Education: Publications Title

Individuals: Events Details

Event Flow:

The main area displays a timeline from 0 to 22. The timeline is divided into several horizontal tracks representing different entities or events. Each track consists of a pink rectangular box labeled with a code identifier (e.g., E-0, T-2, P-1) followed by a series of small colored dots and arrows indicating sequence or state transitions. The tracks are arranged vertically, with some tracks appearing multiple times across the timeline.

Introduction

Visualization Design

Evaluation

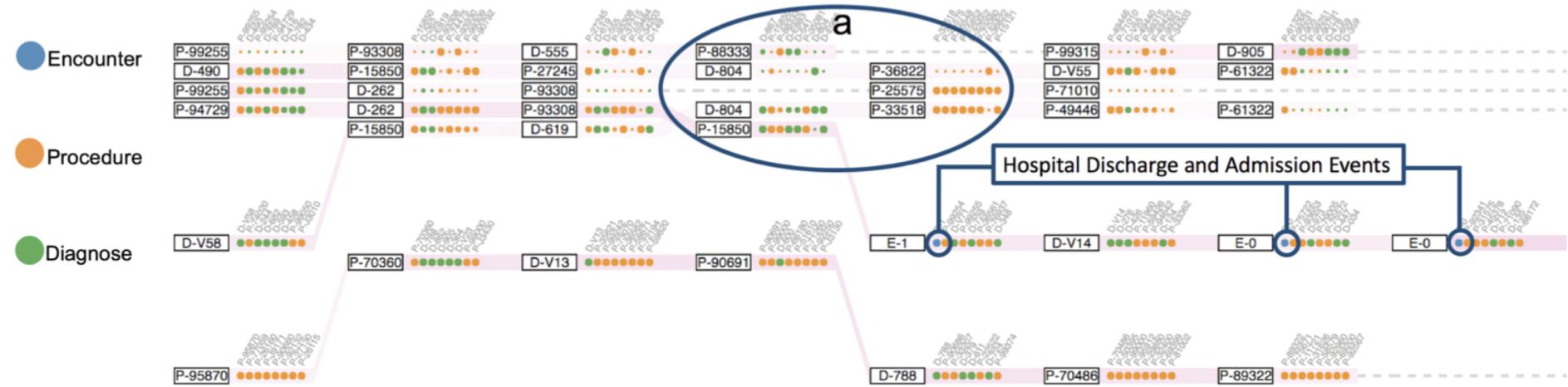


Usage Scenario: COPD Cohort



- 5084 COPD patients
- Timestamped events: diagnosis, procedure, encounter
- From 2008-2014

Usage Scenario: COPD Cohort



Evaluation: Expert Interview



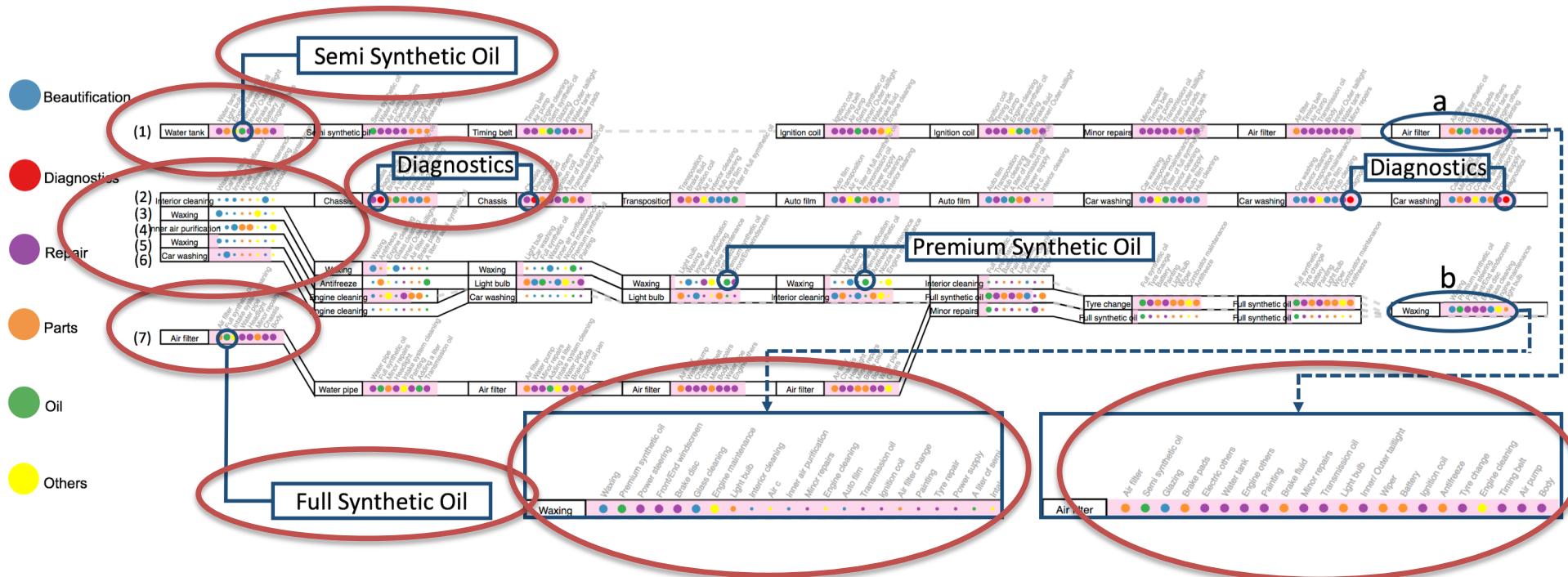
Assistant Professor of Medicine at
the University of North Carolina
School of Medicine

Usage Scenario: Car Maintenance



- 5000 maintenance record
- 1112 cars
- Maintenance type, specific maintenance item, description of the item

Usage Scenario: Car Maintenance

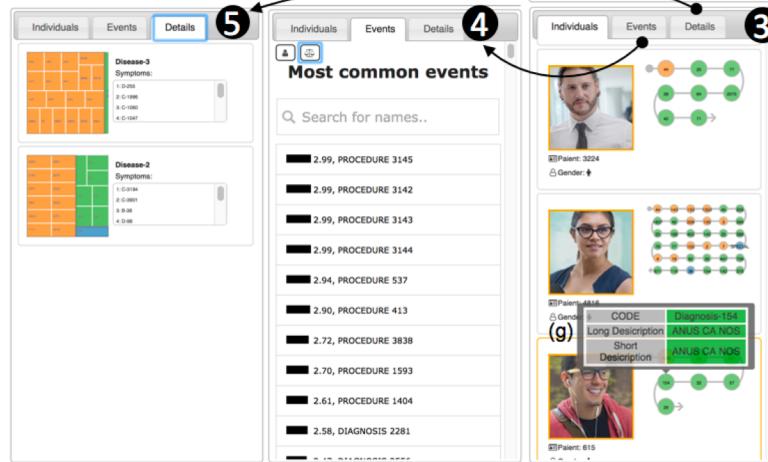


Conclusion

- We have presented EventThread, a technique designed to support visual summarization and latent stage analysis of large scale and high-dimensional event sequence data
- We evaluated our system via real-world event sequence datasets, and conducted an interview with an expert from the health-care domain



Thank You



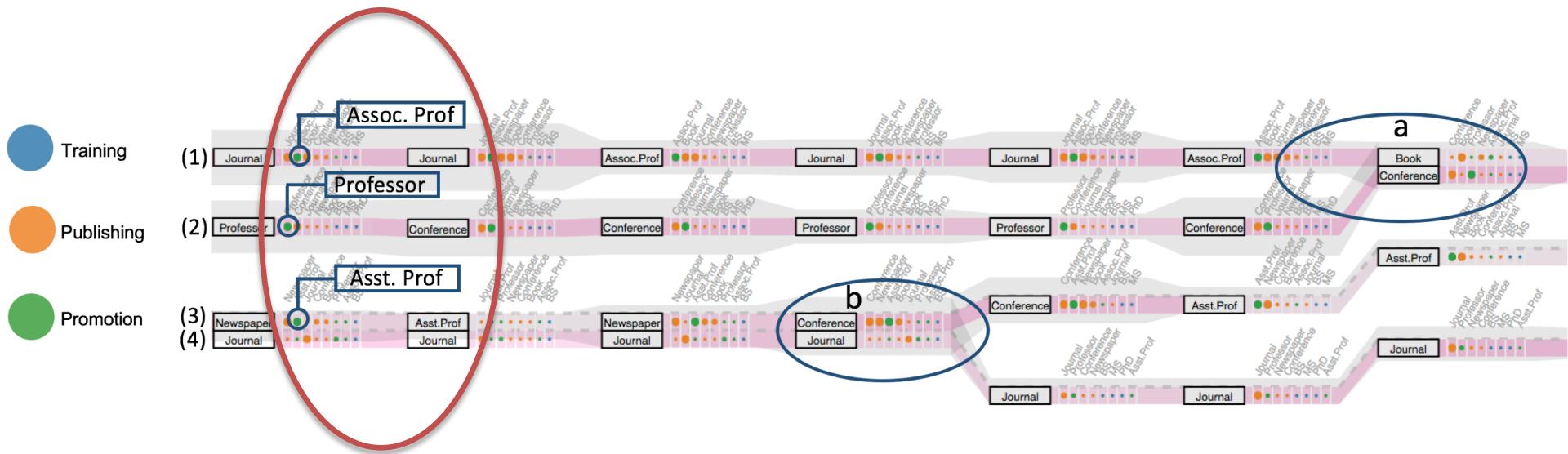
EventThread: Visual Summarization and Stage Analysis of Event Sequence Data

Usage Scenario: Academic Behaviors

- 40 individuals
- 23 years
- 10 event types, classified into 3 high-level categories: training, publishing, promotion

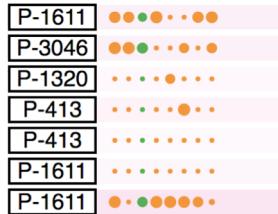


Usage Scenario: Academic Behaviors

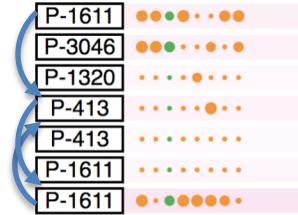


Creating the Thread View

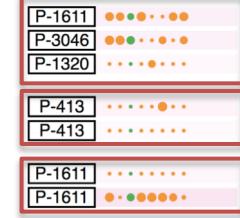
Visualize Threads



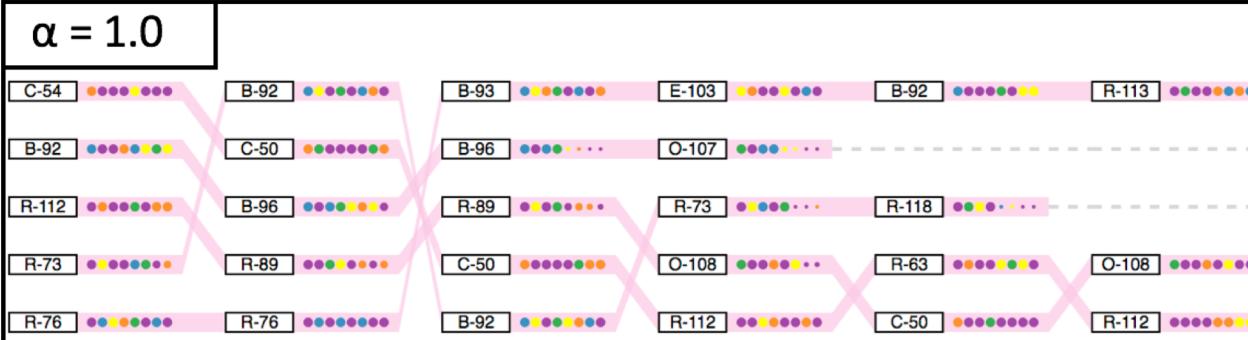
Layout Threads



Rearrange Threads

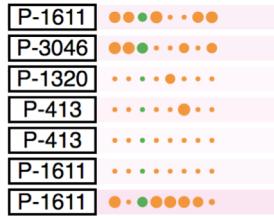


$$\sum_{t=0}^T \left(\alpha \sum_{i < j} w_{ij}(t) \|y_i(t) - y_j(t)\|^2 + (1 - \alpha) \sum_i \|y_i(t) - y_i(t-1)\|^2 \right)$$

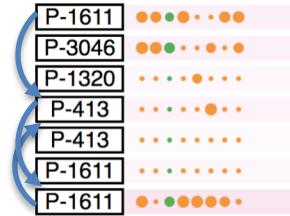


Creating the Thread View

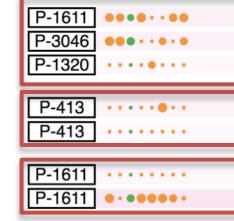
Visualize Threads



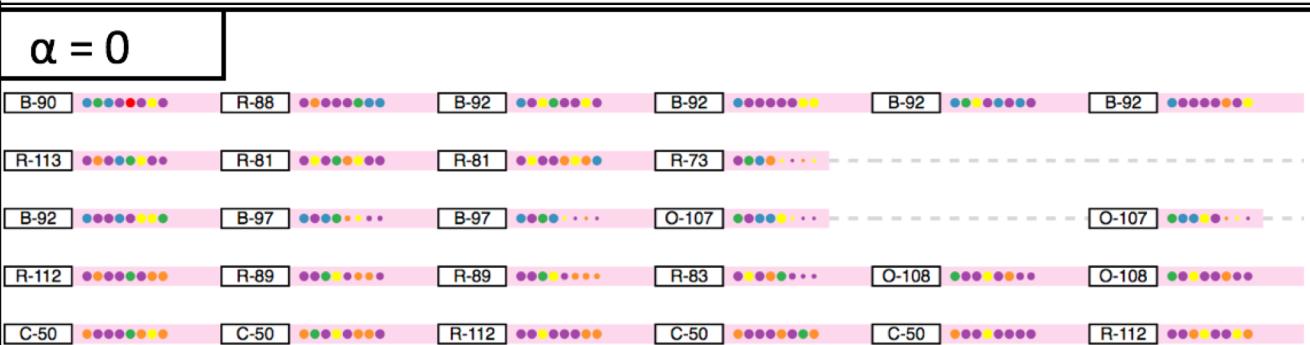
Layout Threads



Rearrange Threads

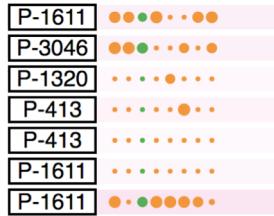


$$\sum_{t=0}^T \left(\alpha \sum_{i < j} w_{ij}(t) \|y_i(t) - y_j(t)\|^2 + (1 - \alpha) \sum_i \|y_i(t) - y_i(t-1)\|^2 \right)$$

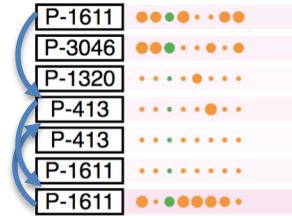


Creating the Thread View

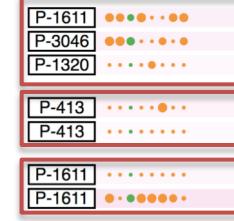
Visualize Threads



Layout Threads



Rearrange Threads



$$\sum_{t=0}^T \left(\alpha \sum_{i < j} w_{ij}(t) \|y_i(t) - y_j(t)\|^2 + (1 - \alpha) \sum_i \|y_i(t) - y_i(t-1)\|^2 \right)$$

$\alpha = 0.5$

