# EventThread: Visual Summarization and Stage Analysis of Event Sequence Data

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



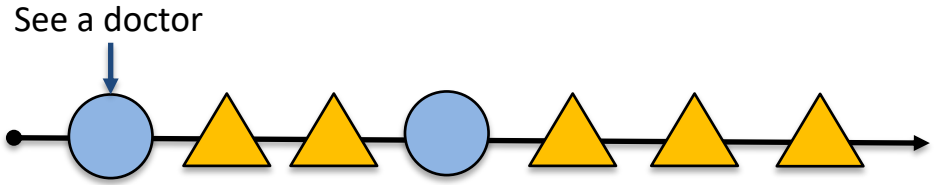


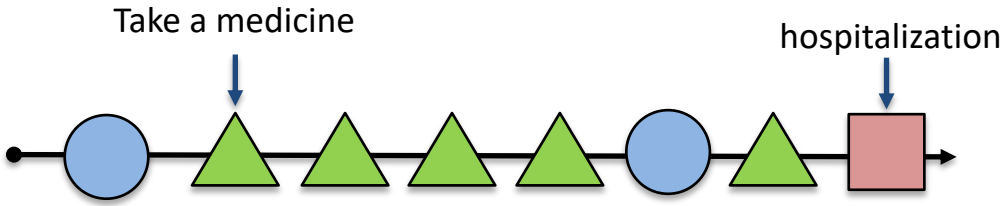
# Event Sequence Data

**Academic Record**

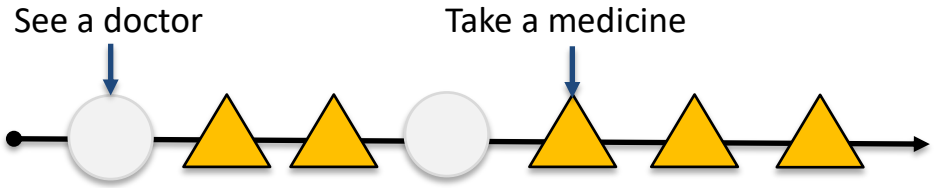
**Electronic Health Record**

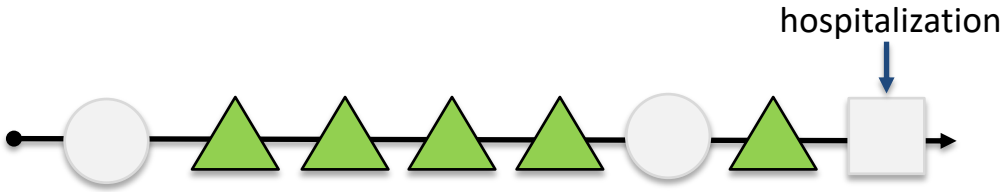
# Electronic Health Record



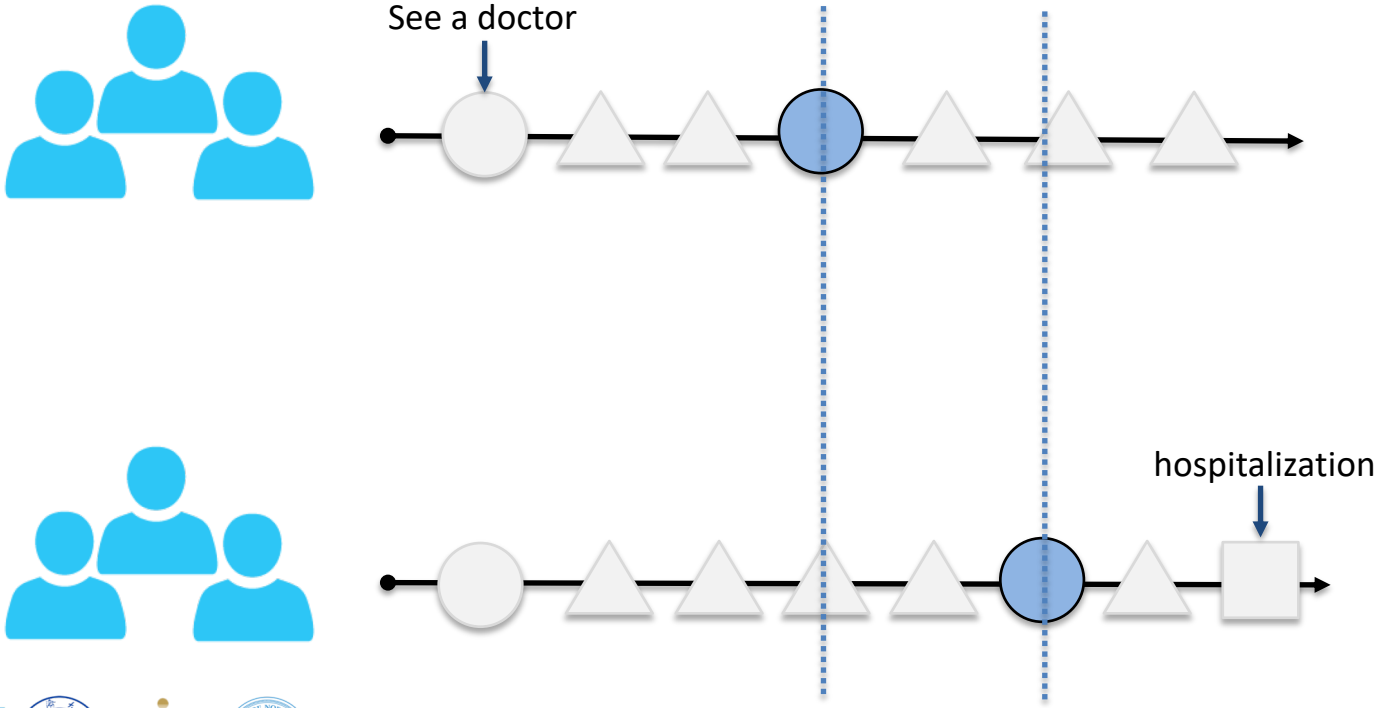



# Electronic Health Record

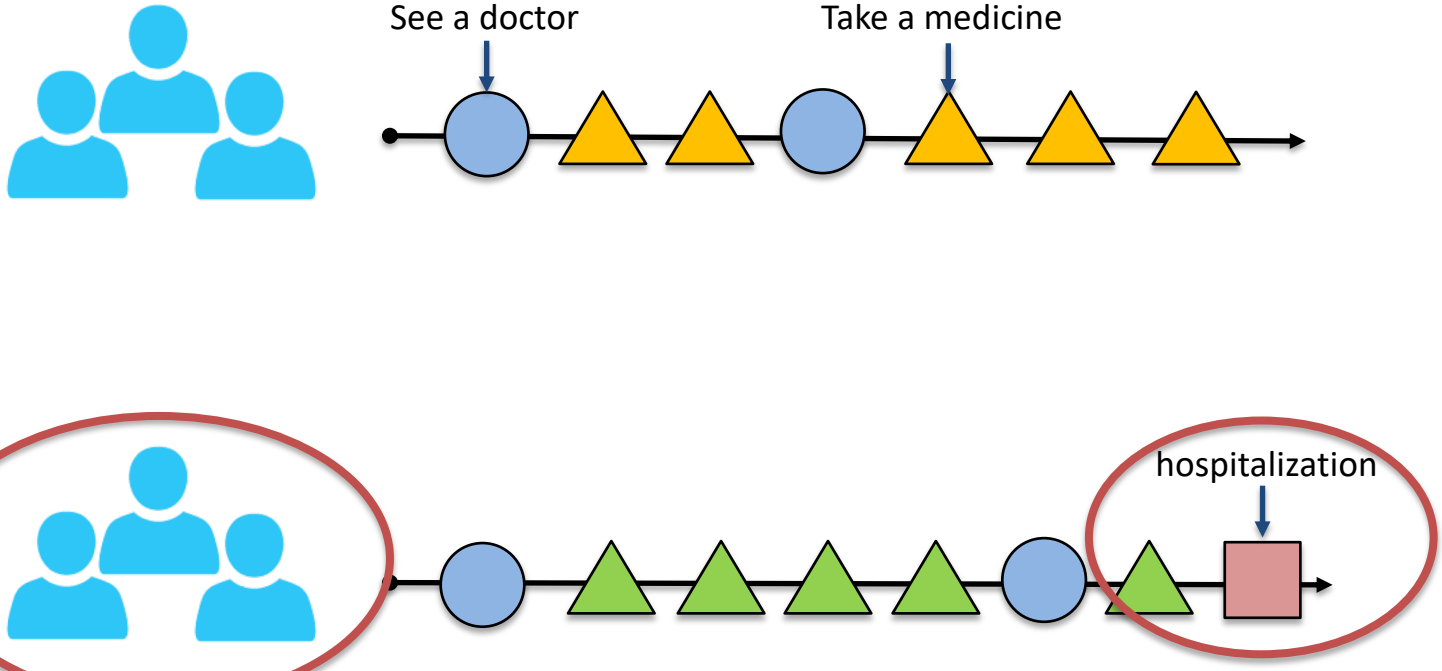


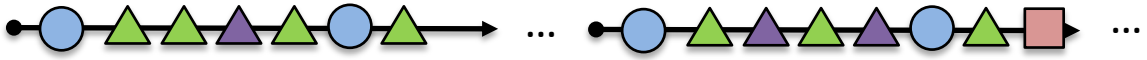
# Electronic Health Record



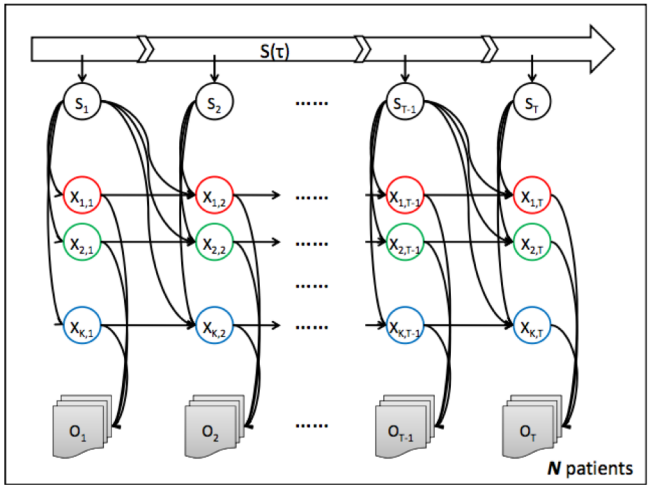
# Electronic Health Record



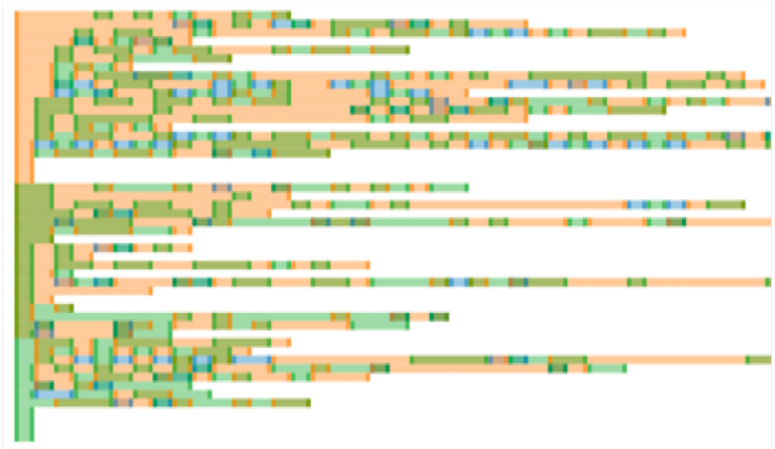
# Electronic Health Record



# Research Goal

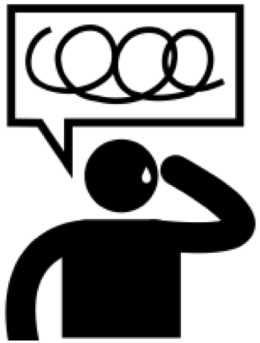



Gap

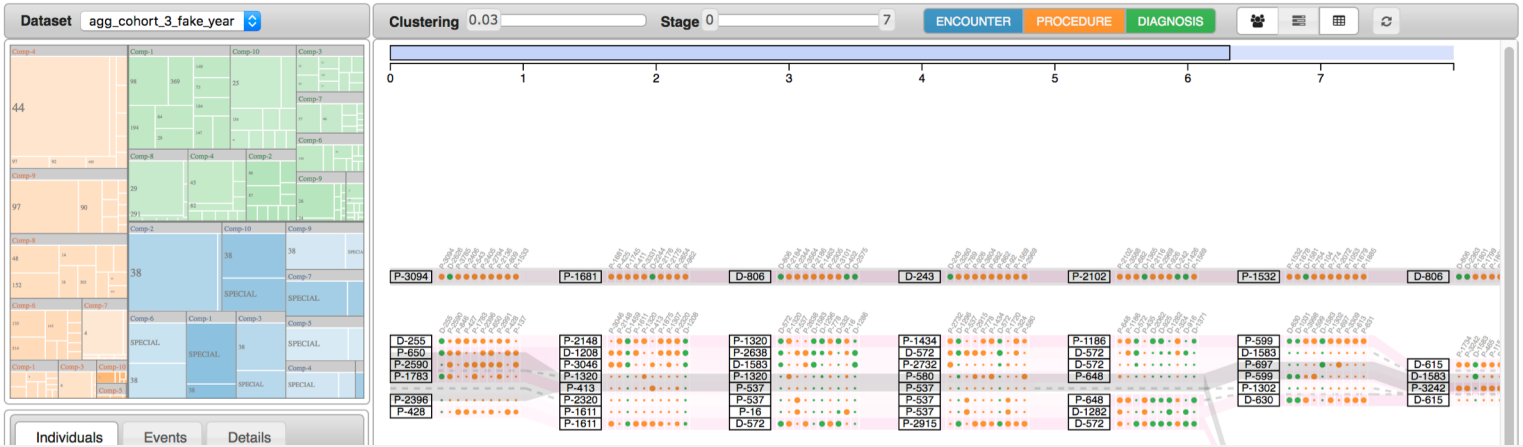


- Discover latent high-level structures of large-scale event sequence data.
- Provide users with information about the low-level events and sub-sequences of events.

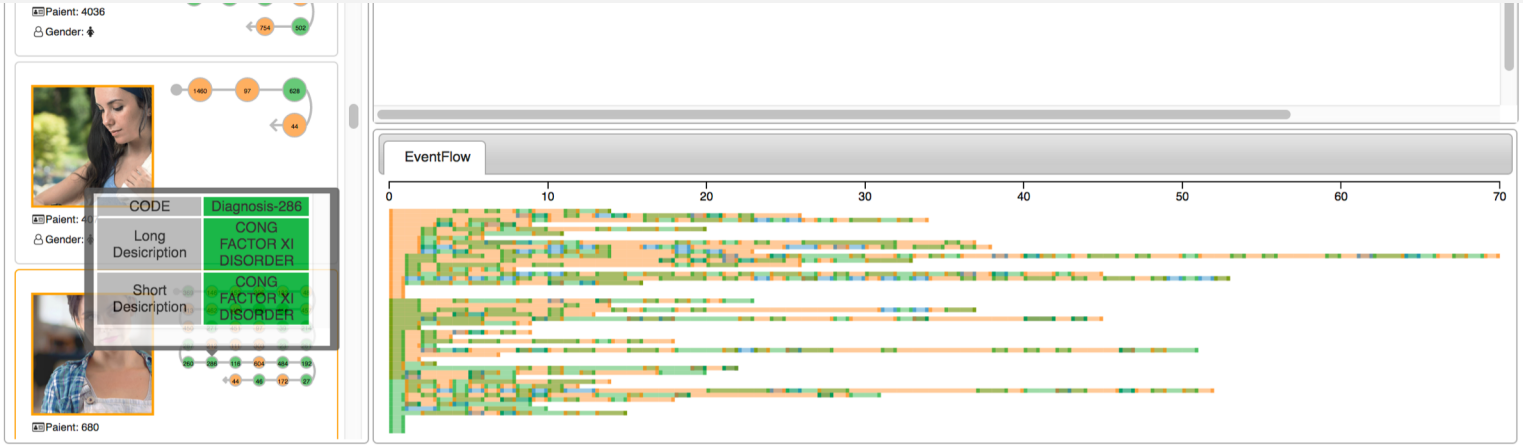
# Key Challenges



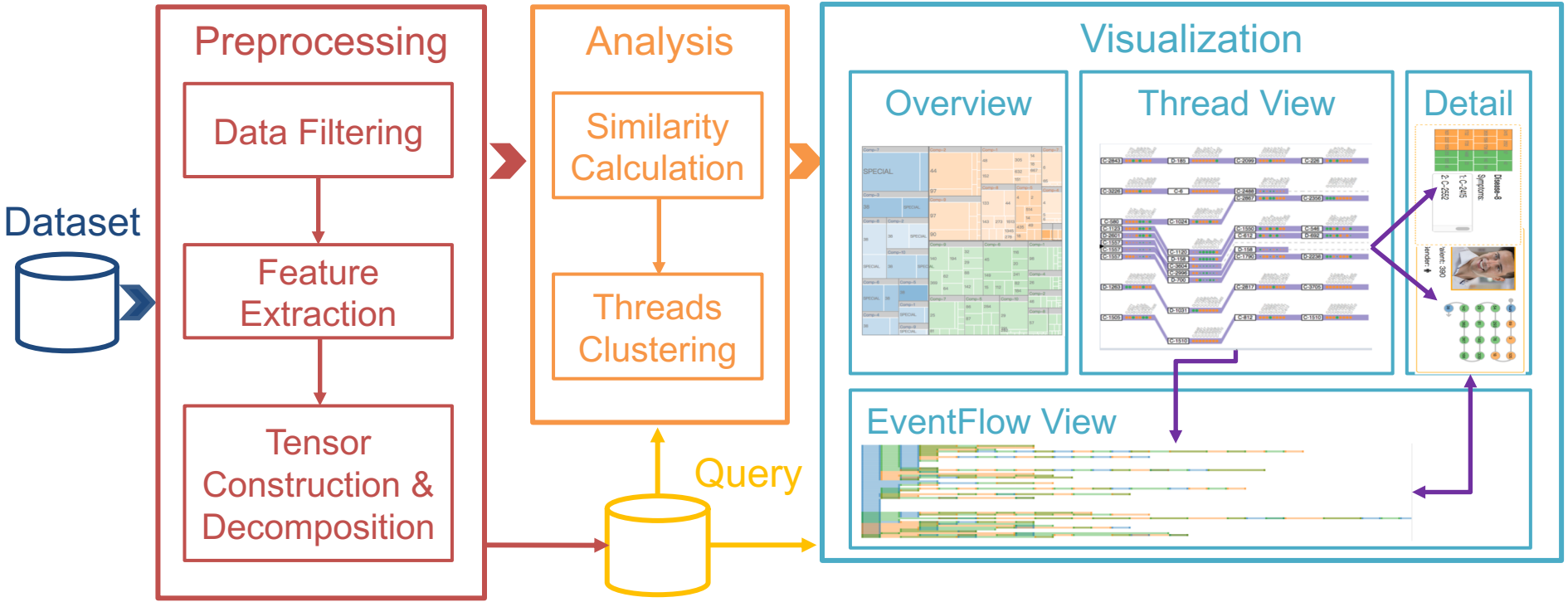
- It is difficult to transform large-scale heterogeneous event sequence data into an uniform data model without losing detailed information
- The method designed to detect high-level structures should also include sufficient relevant context to enable low-level semantic interpretation of what those structures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




# EventThread System

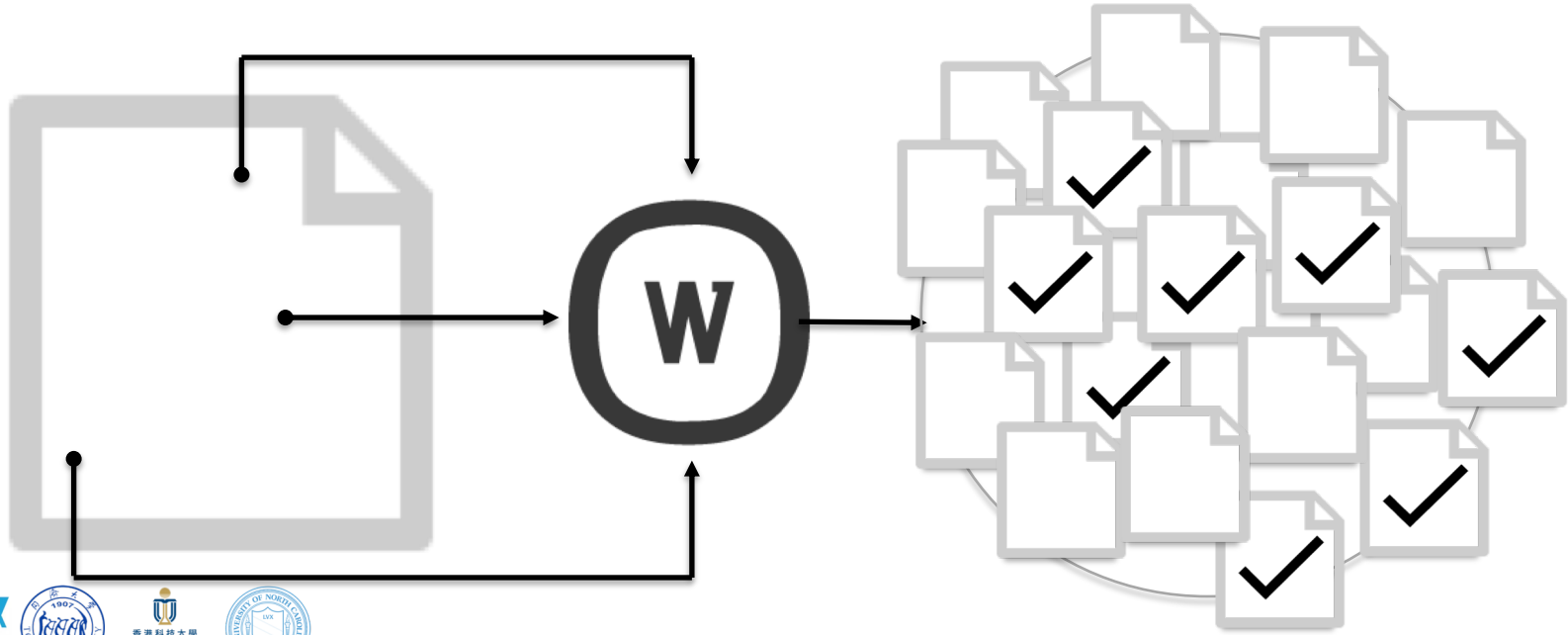


# System Overview



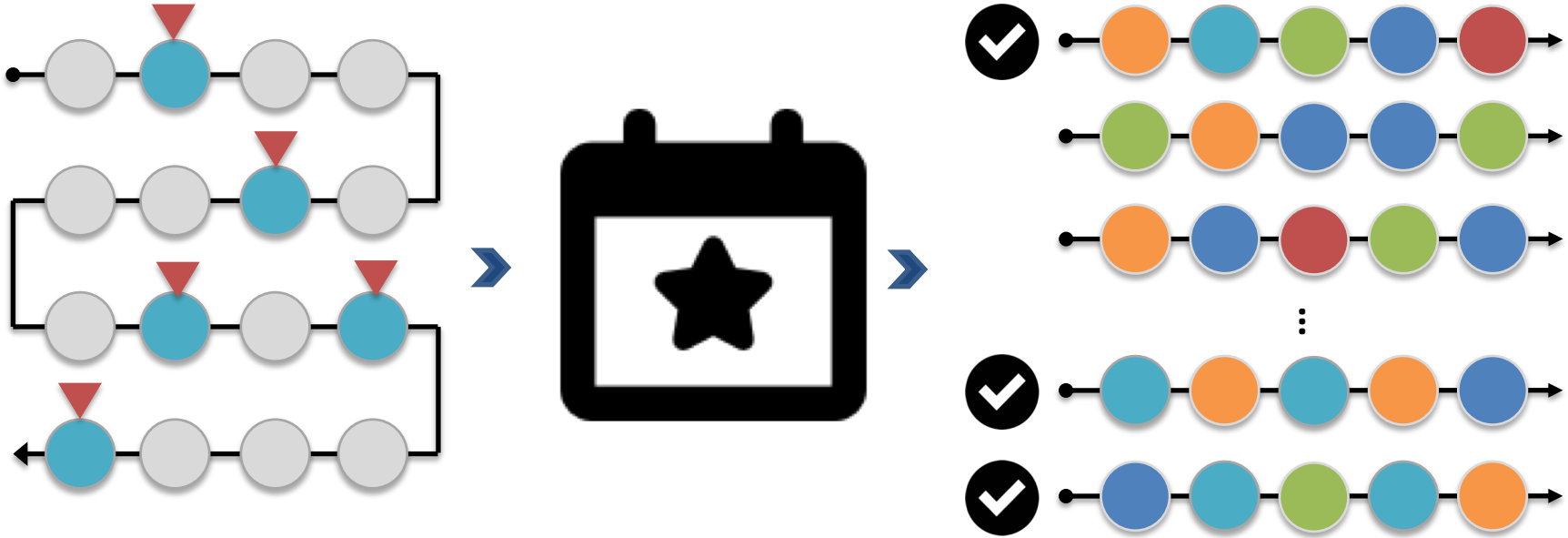
# Data Filtering

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

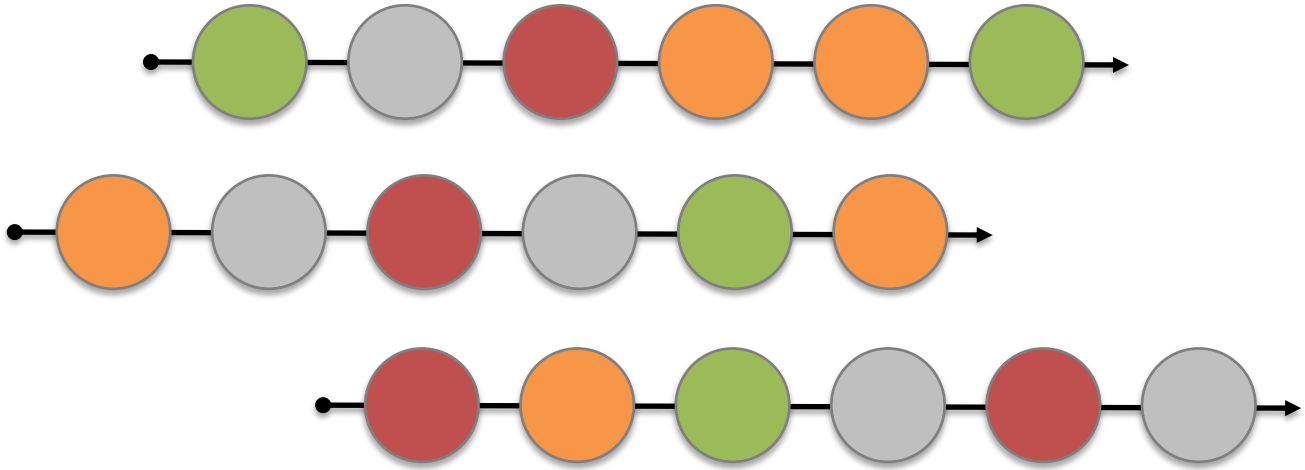


# Data Filtering

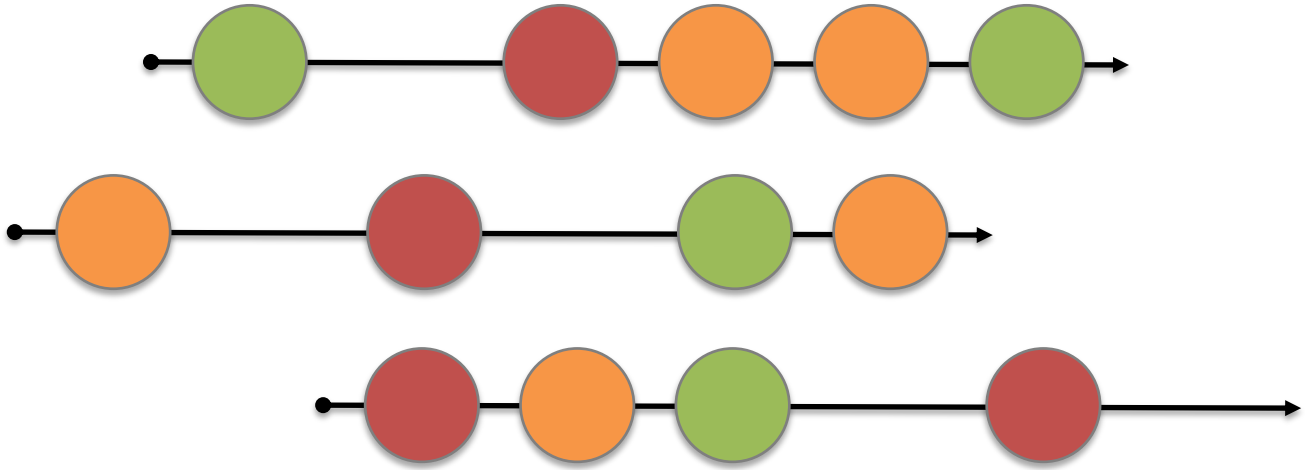
- *TF-IDF* - Event Sequences



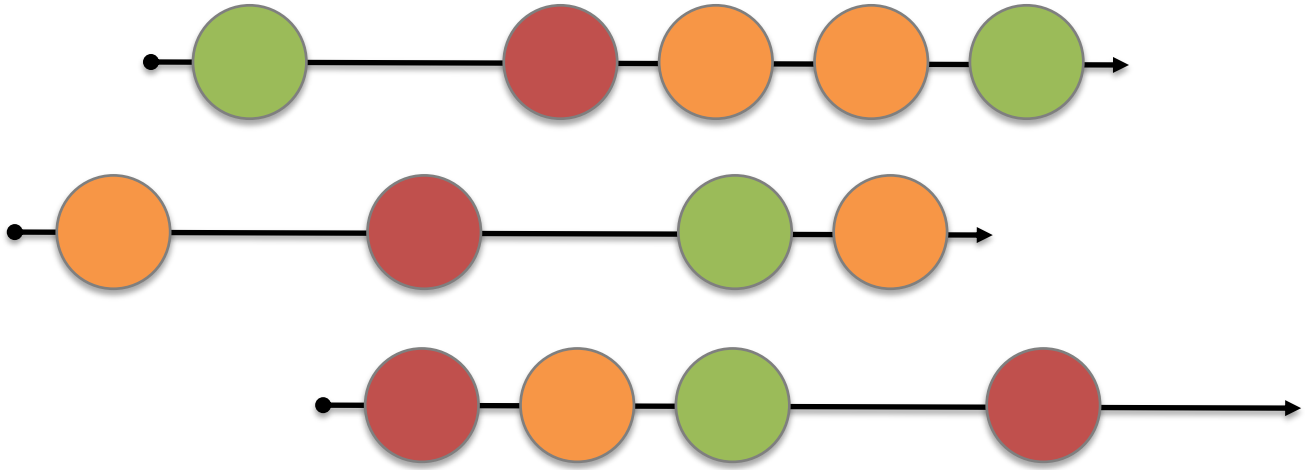
# Sequence Preprocessing



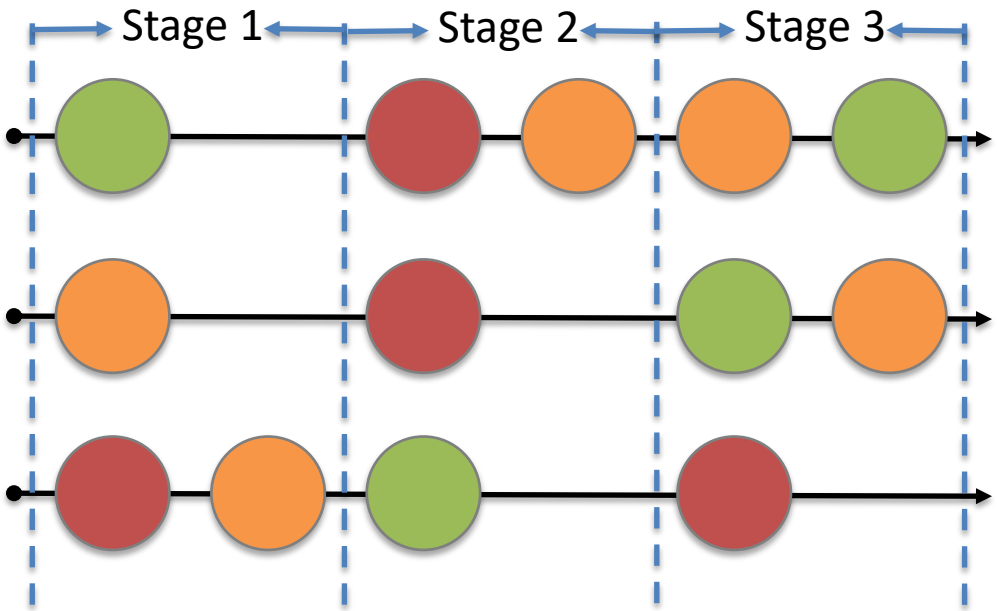
# Sequence Preprocessing



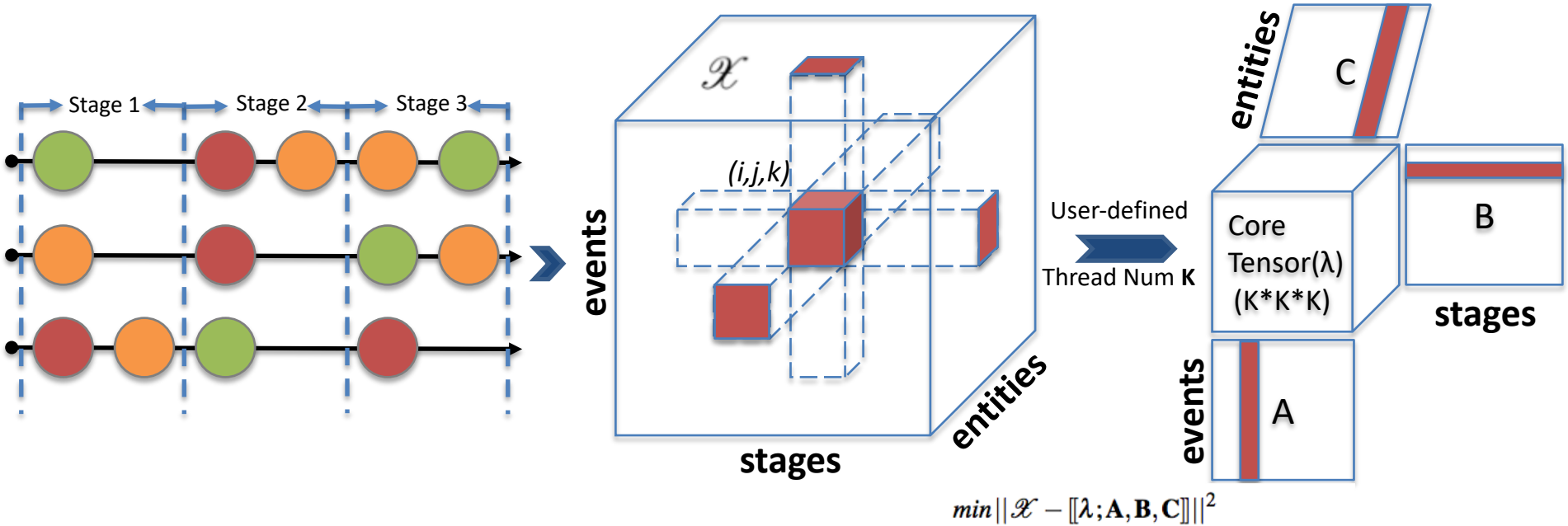
# Sequence Preprocessing



# Sequence Preprocessing



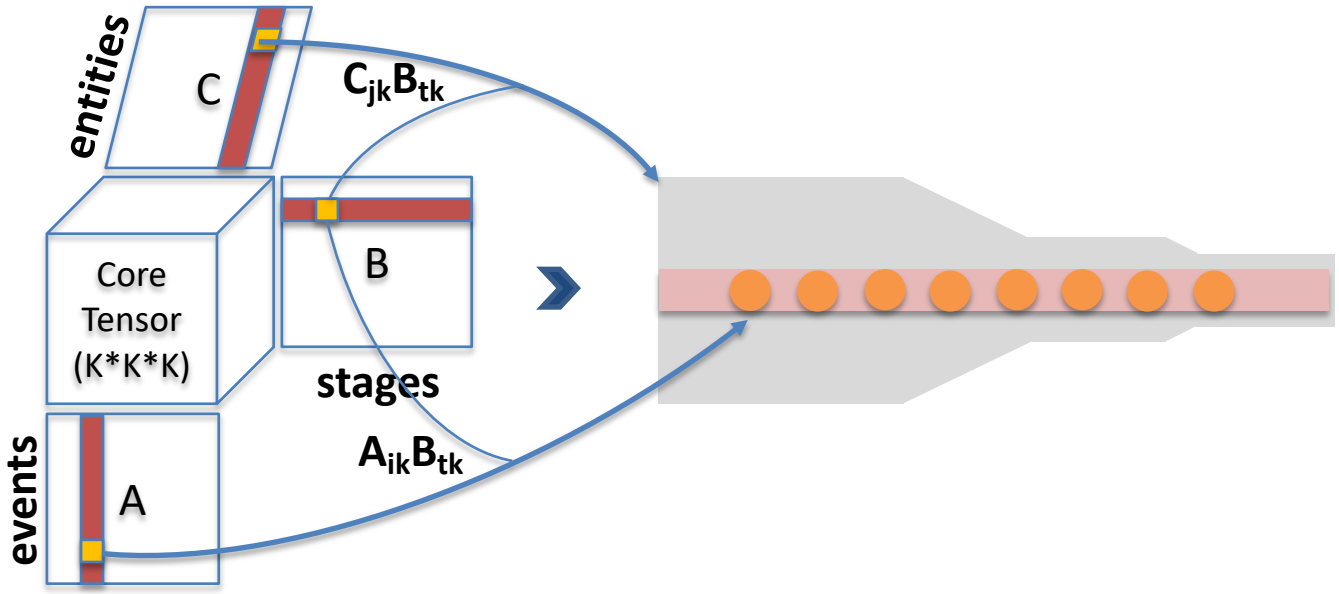
# Tensor Analysis




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

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

# Tensor Analysis

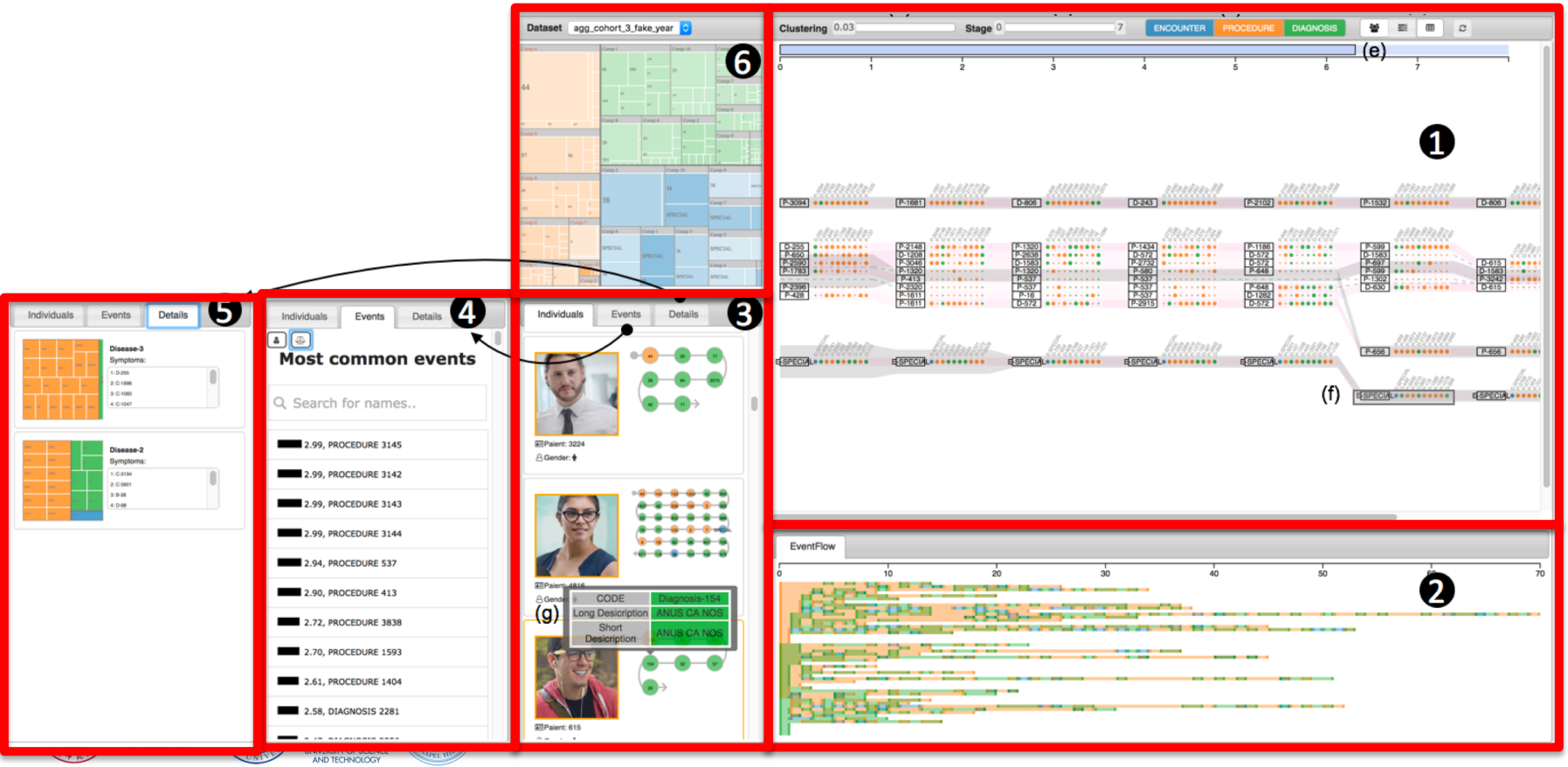


*Introduction*

***Visualization Design***

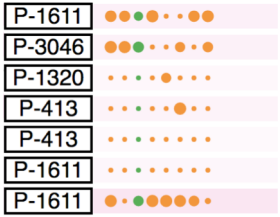
*Evaluation*

# User Interface

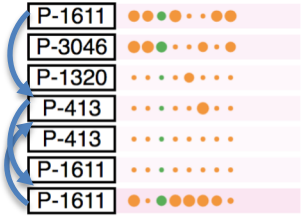


# Creating the Thread View

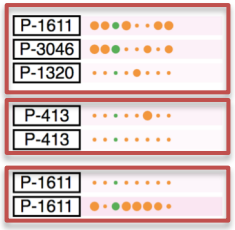
## Visualize Threads



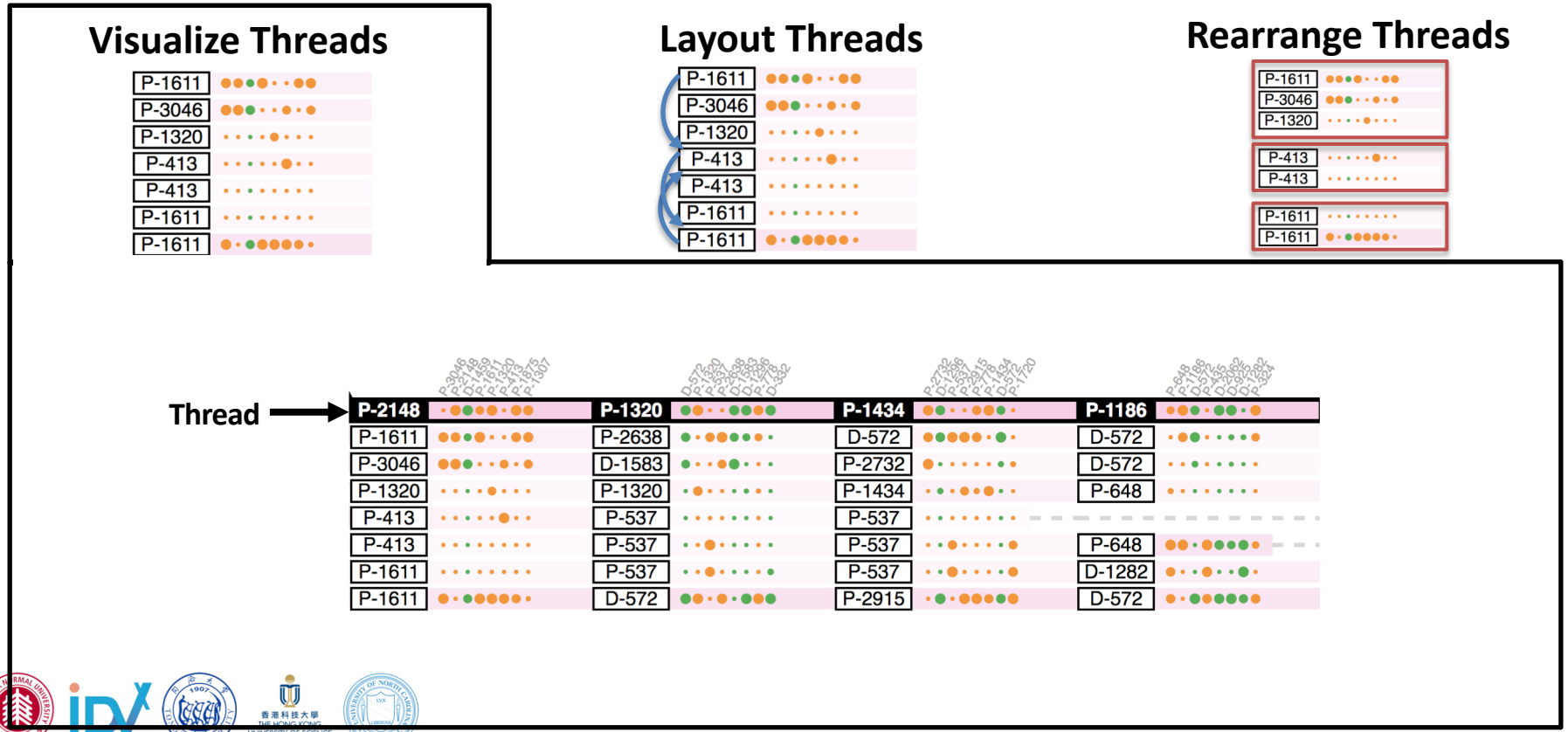

## Layout Threads



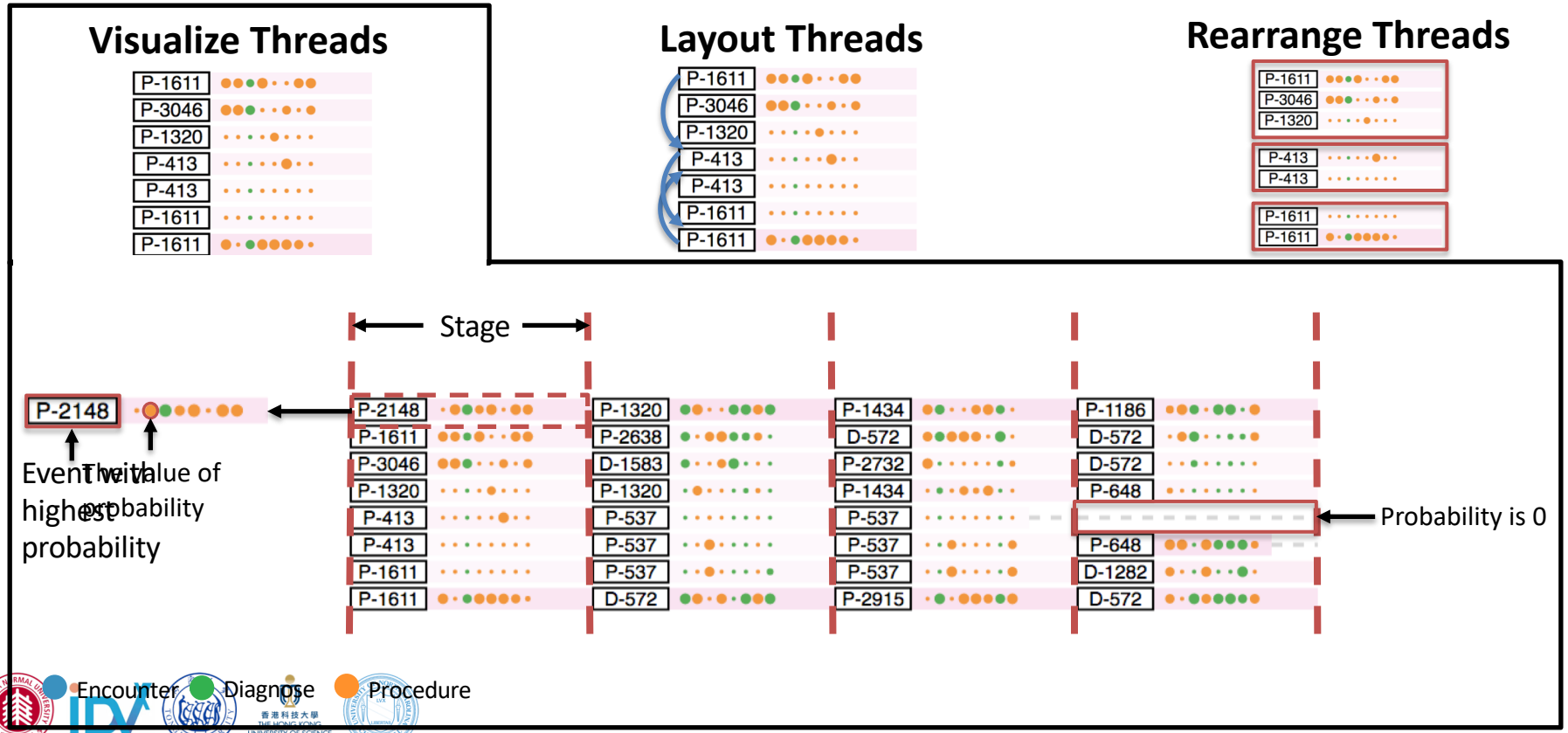

## Rearrange Threads



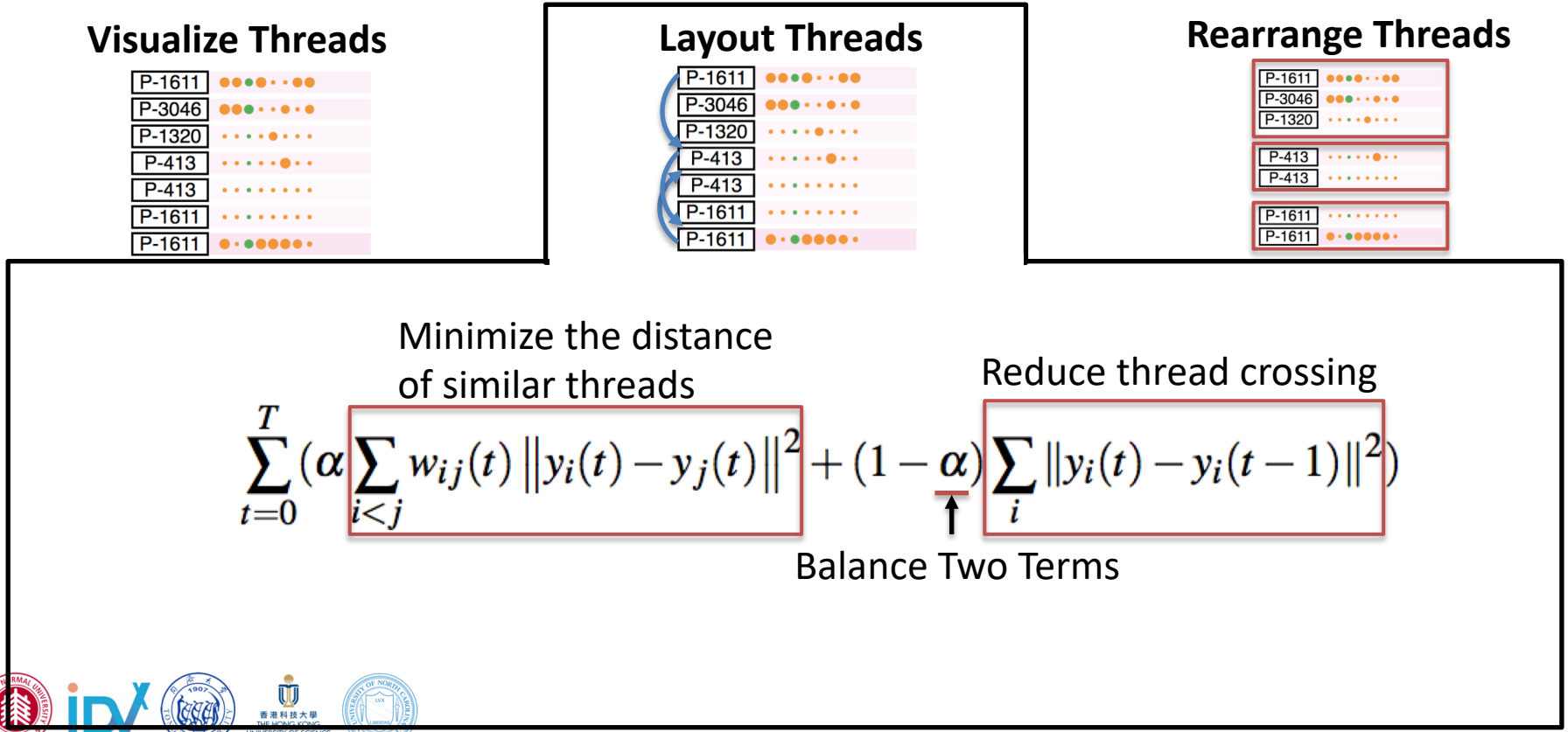
# Creating the Thread View



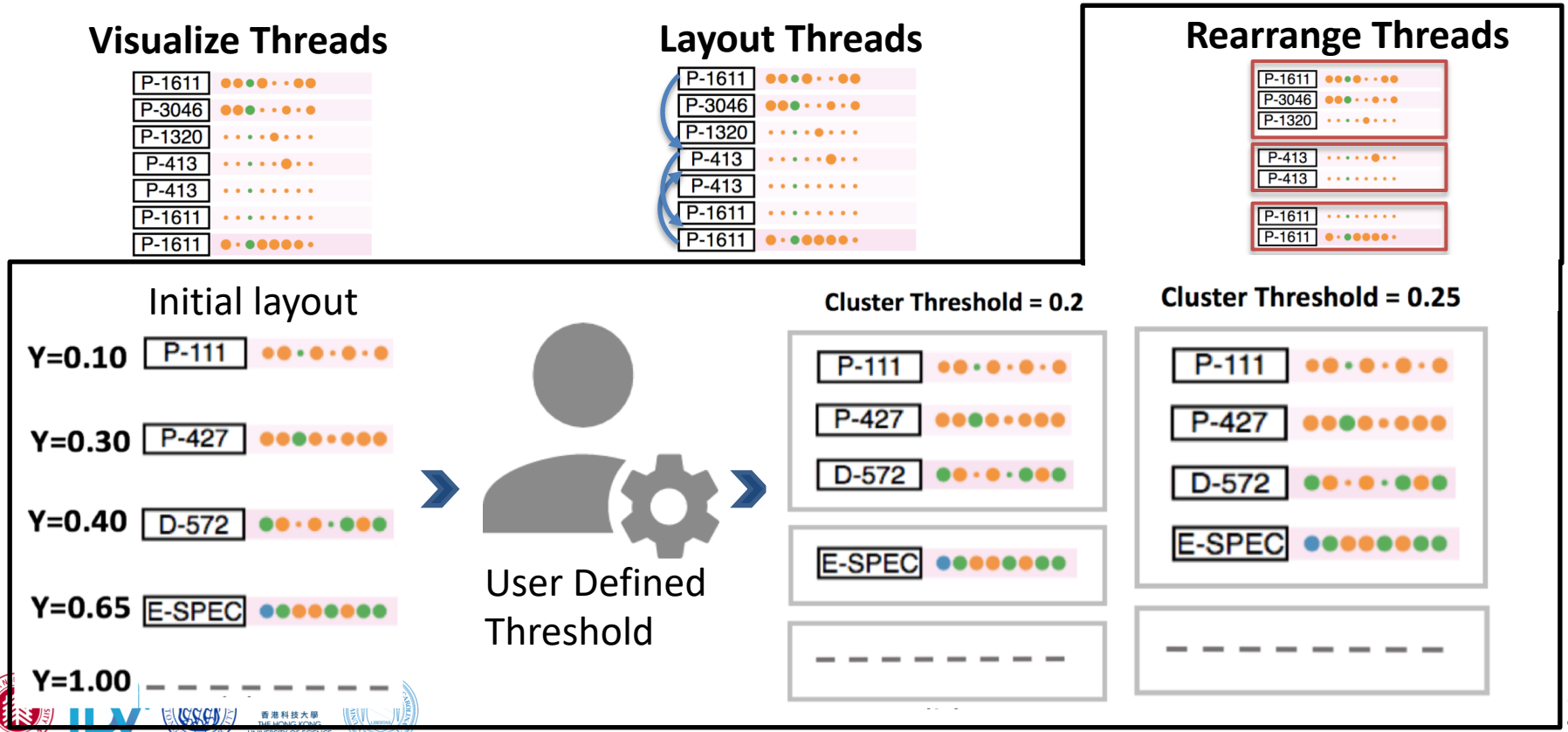
# Creating the Thread View



# Creating the Thread View

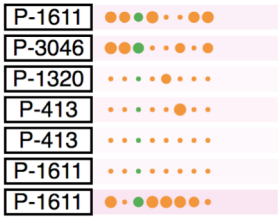


# Creating the Thread View

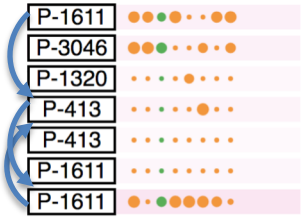


# Creating the Thread View

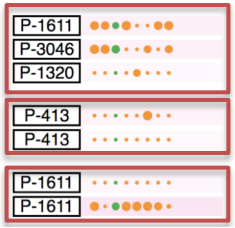
## Visualize Threads

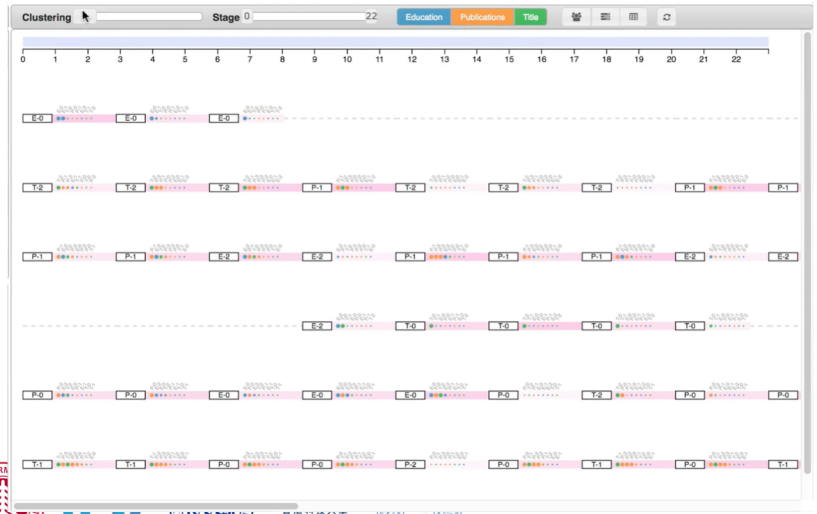


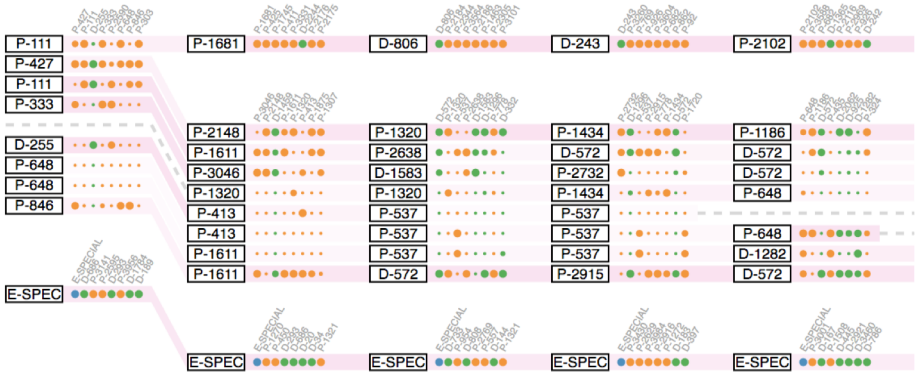

## Layout Threads



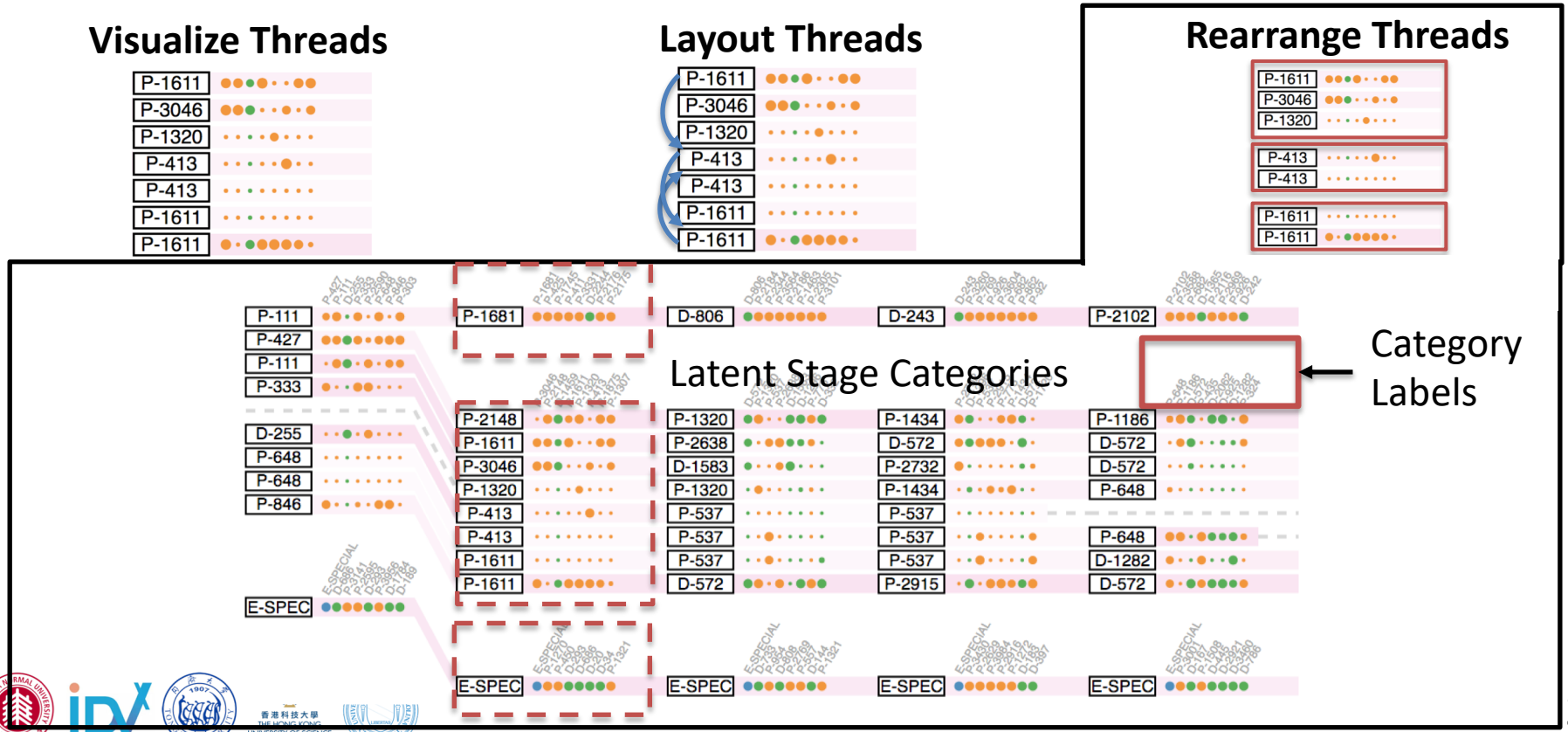

## Rearrange Threads



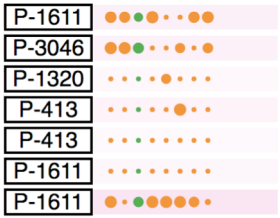



# Creating the Thread View

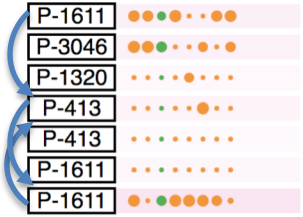


# Creating the Thread View

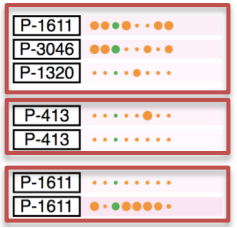
## Visualize Threads

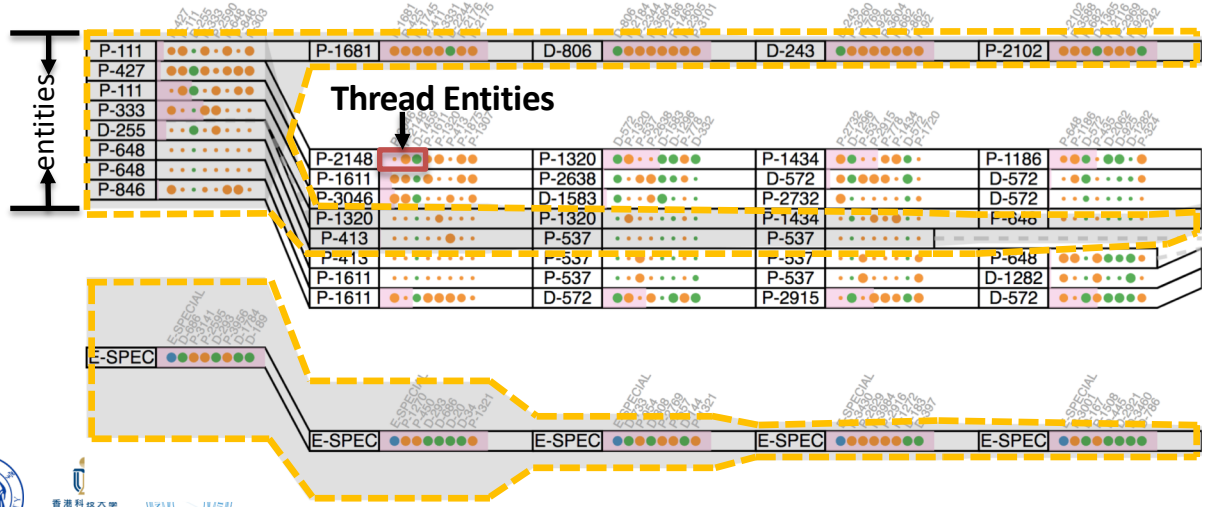


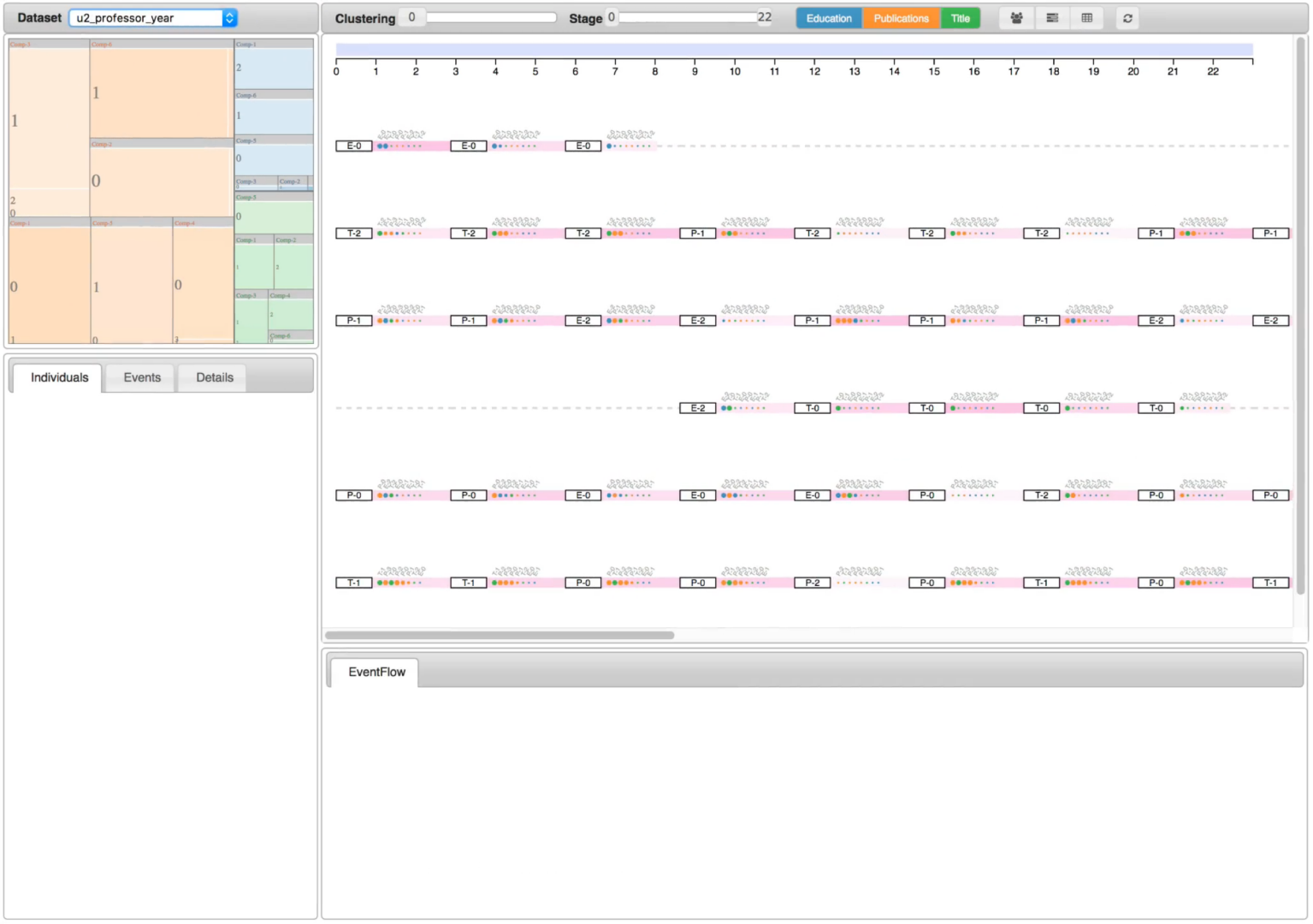

## Layout Threads




## Rearrange Threads




Dataset
u2_professor_year
Clustering
0
Stage
0
22
Education
Publications
Title
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Individuals
Events
Details
EventFlow

*Introduction*

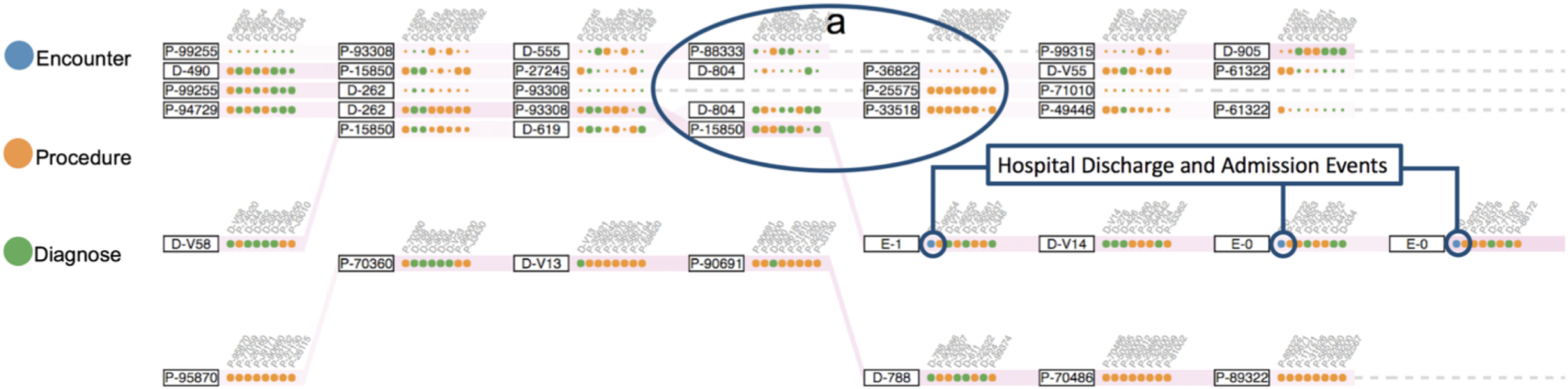
*Visualization Design*

***Evaluation***

# Usage Scenario: COPD Cohort

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

# Usage Scenario: COPD Cohort



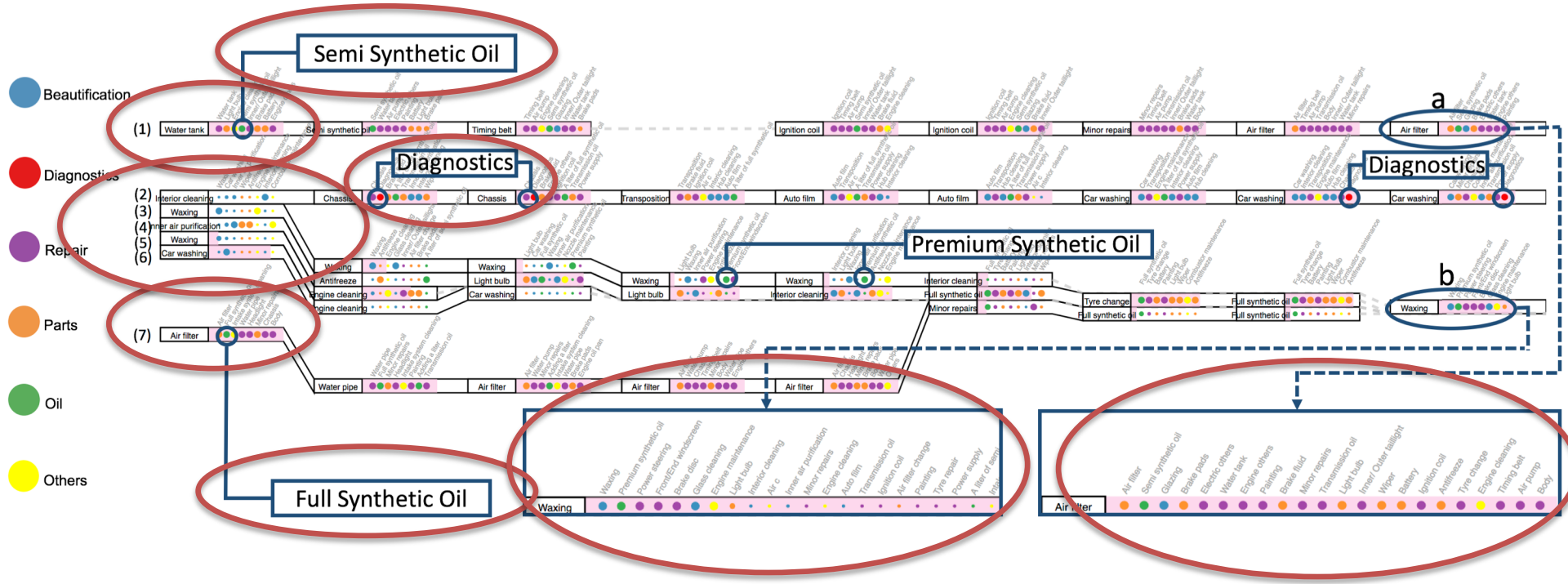
# Evaluation: Expert Interview

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

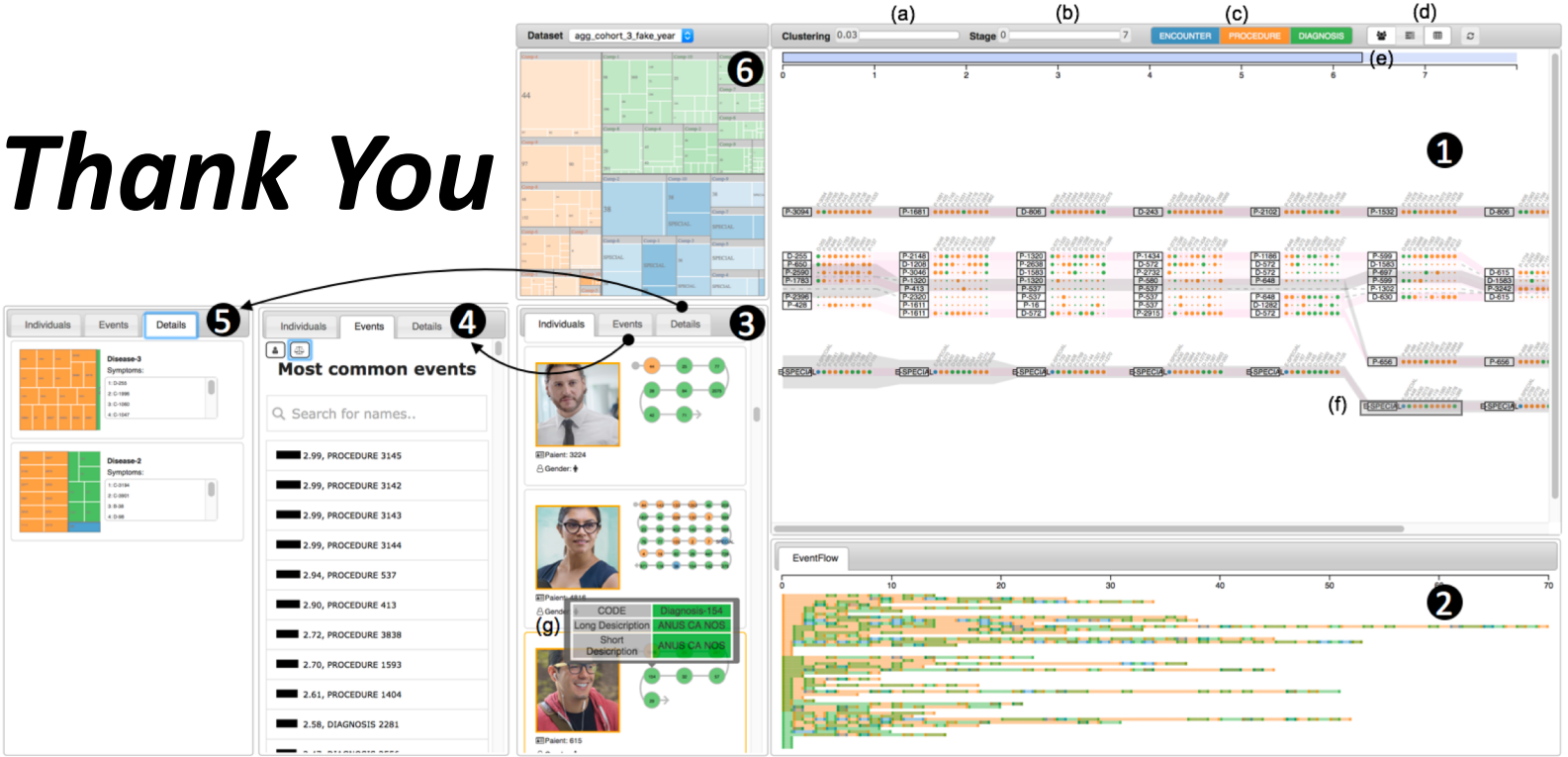


# Conclusion

- 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*

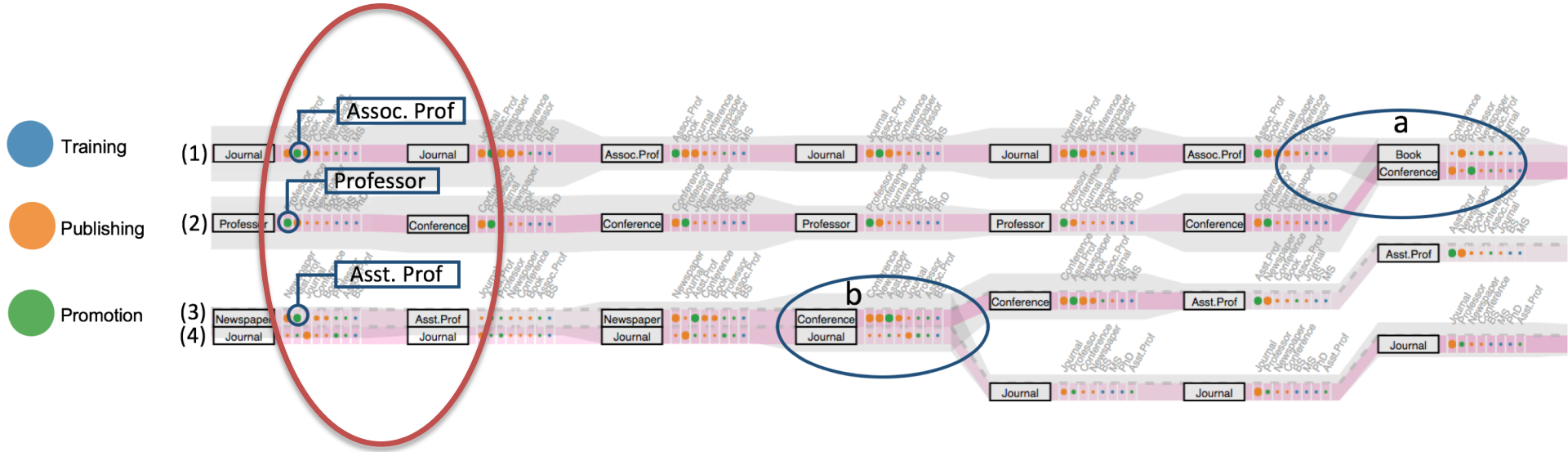



***EventThread: Visual Summarization and Stage Analysis of Event Sequence Data***

# Usage Scenario: Academic Behaviors

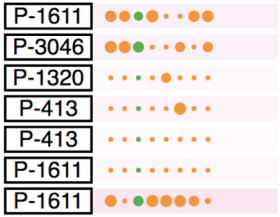
- 40 individuals
- 23 years
- 10 event types, classified into 3 high-level categories: training, publishing, promotion

# Usage Scenario: Academic Behaviors

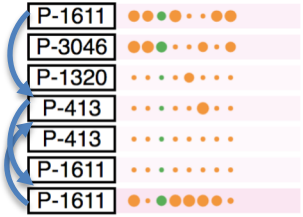


# Creating the Thread View

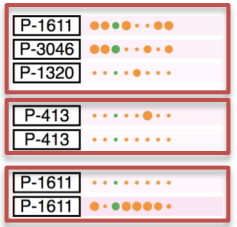
## Visualize Threads



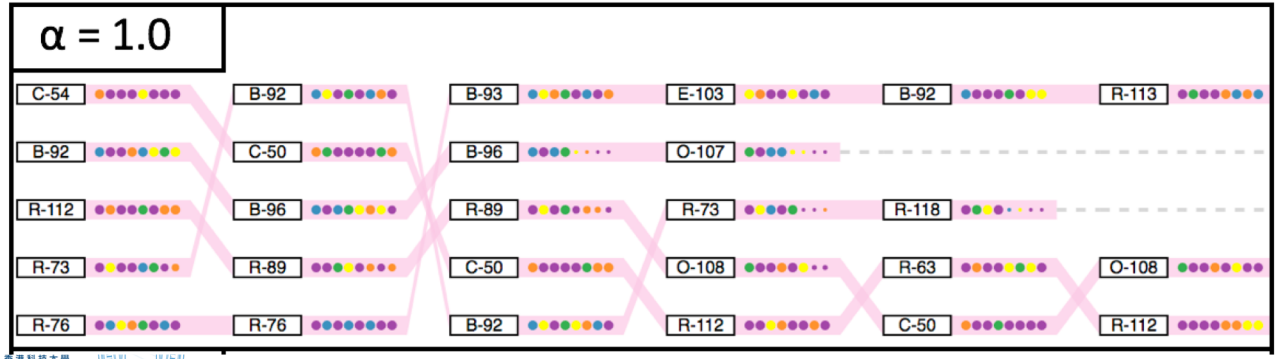

## Layout Threads




## Rearrange Threads

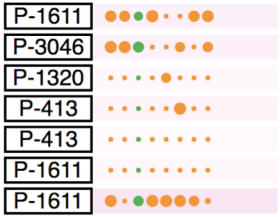



$$\sum_{t=0}^{T}(\alpha\sum_{i<j}w_{ij}(t)\left\|y_i(t)-y_j(t)\right\|^2+(1-\alpha)\sum_i\|y_i(t)-y_i(t-1)\|^2)$$

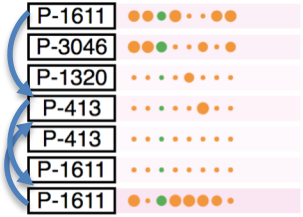


# Creating the Thread View

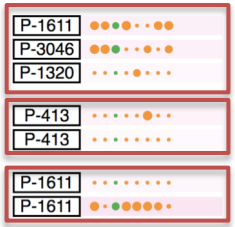
## Visualize Threads



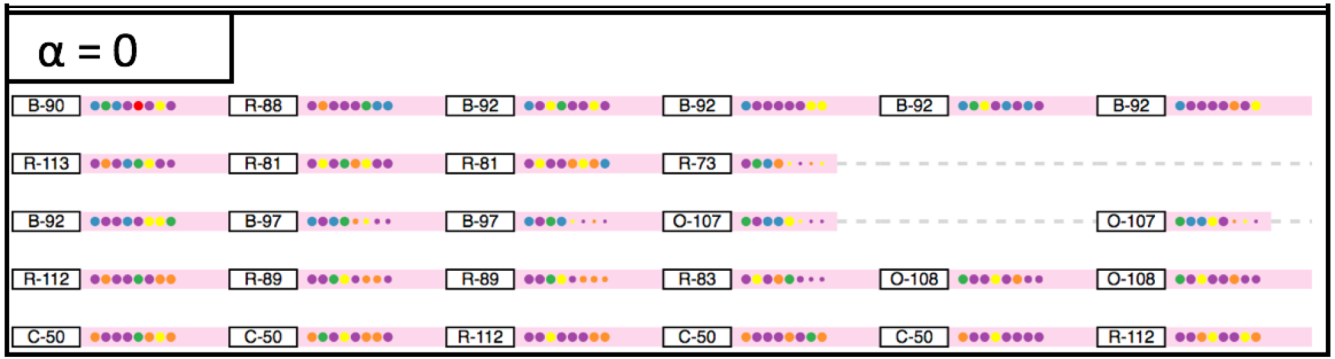

## Layout Threads




## Rearrange Threads

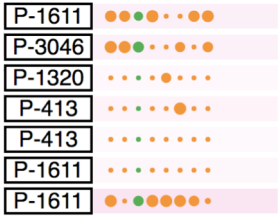



$$\sum_{t=0}^{T}(\alpha\sum_{i<j}w_{ij}(t)\left\|y_i(t)-y_j(t)\right\|^2+(1-\alpha)\sum_i\|y_i(t)-y_i(t-1)\|^2)$$

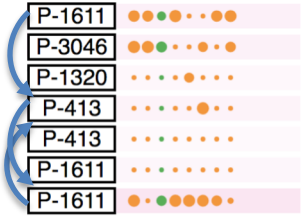


# Creating the Thread View

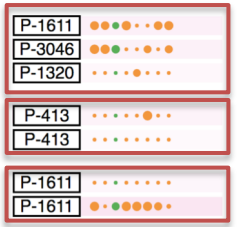
## Visualize Threads




## Layout Threads




## Rearrange Threads




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

α = 0.5