

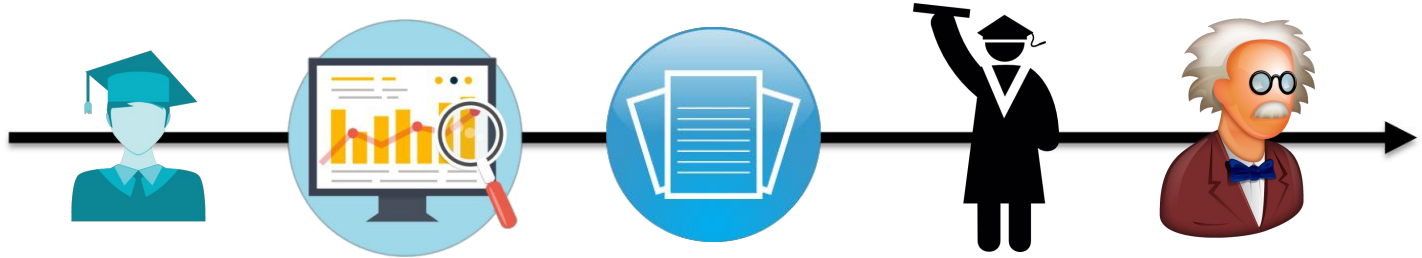
# EventThread: Visual Summarization and Stage Analysis of Event Sequence Data

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



# Event Sequence Data

Academic  
Record



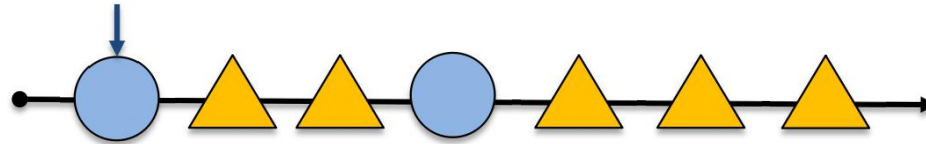
Electronic  
Health  
Record



# Electronic Health Record

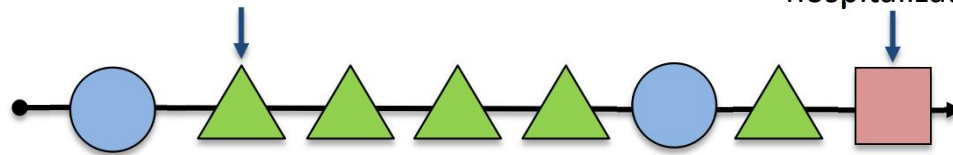


See a doctor

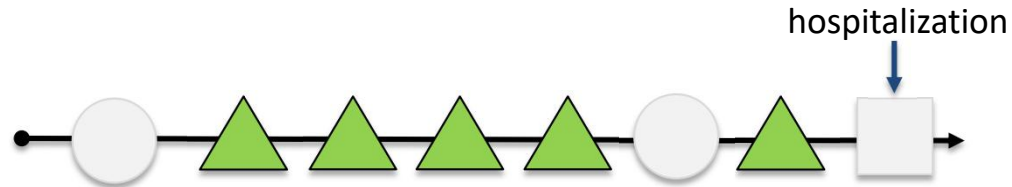
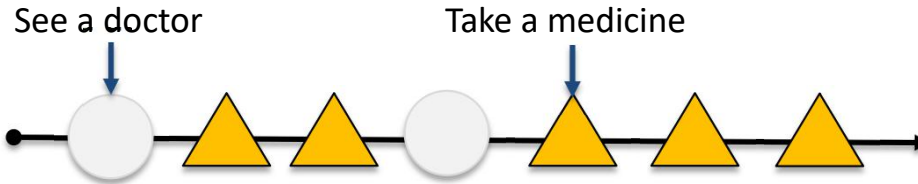


Take a medicine

hospitalization



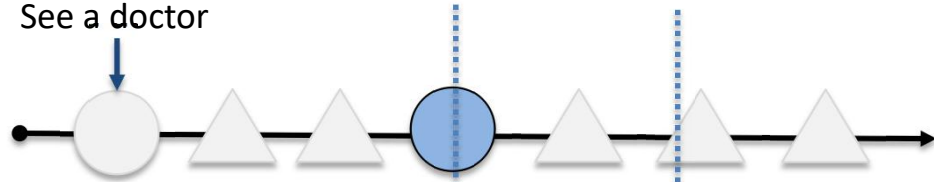
# Electronic Health Record



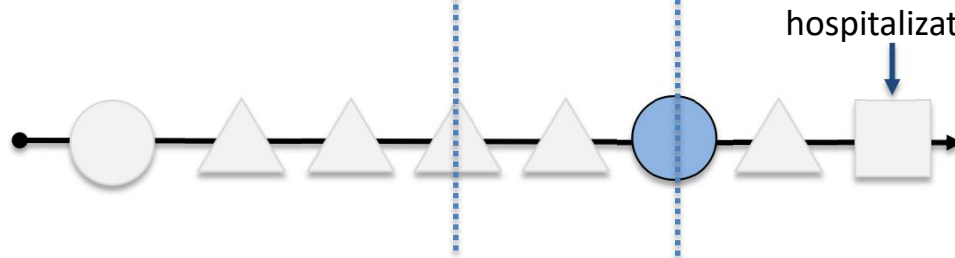
# Electronic Health Record



See a doctor



hospitalization

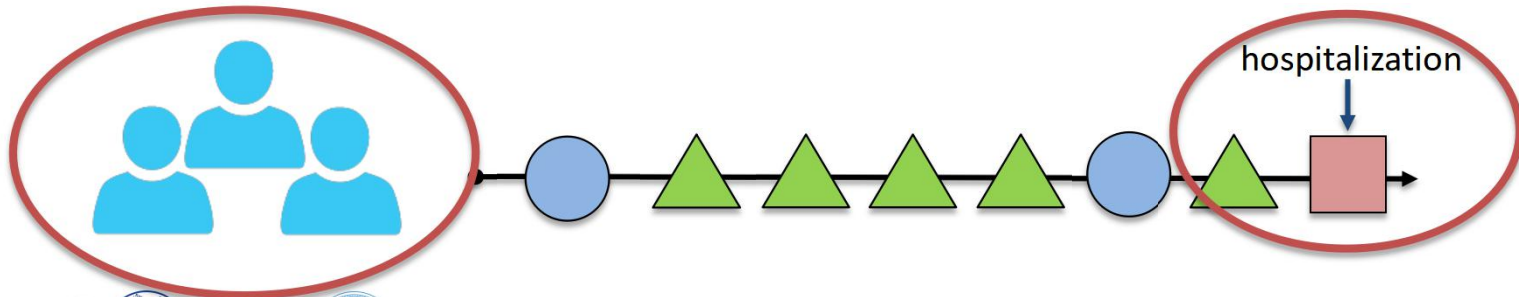
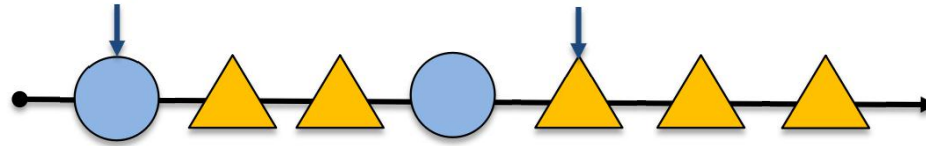


# Electronic Health Record



See a doctor

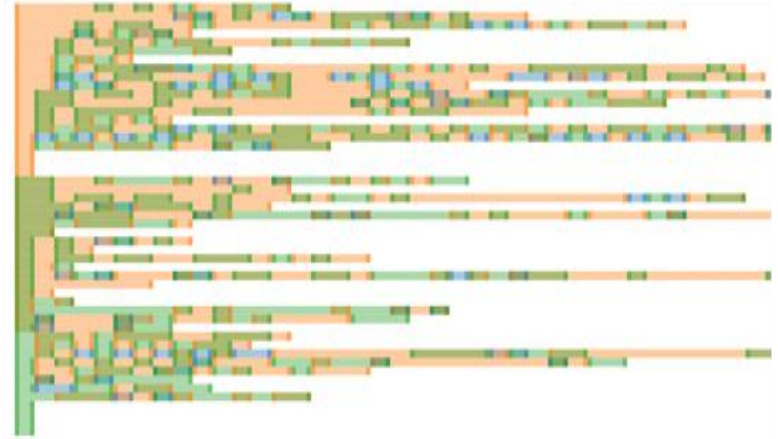
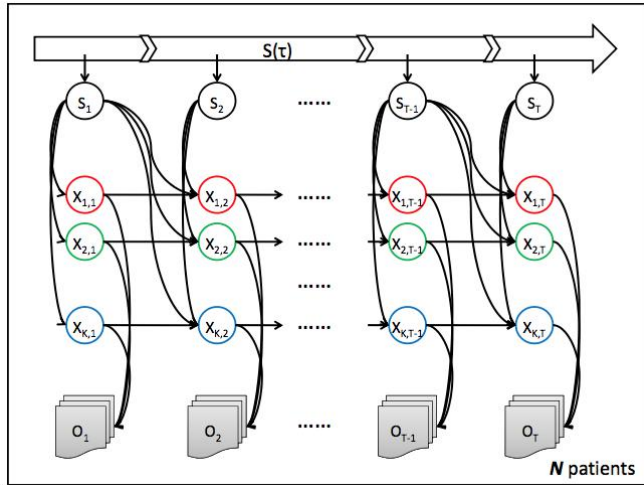
Take a medicine



# Electronic Health Record



# Research Goal

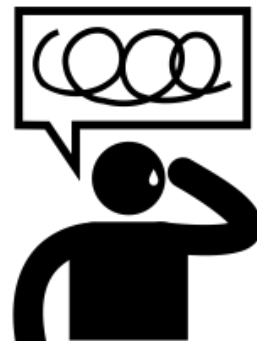


- Discover latent high-level structures of large-scale event sequence data.
- Provide users with information about the low-level events and sub-sequences of events.

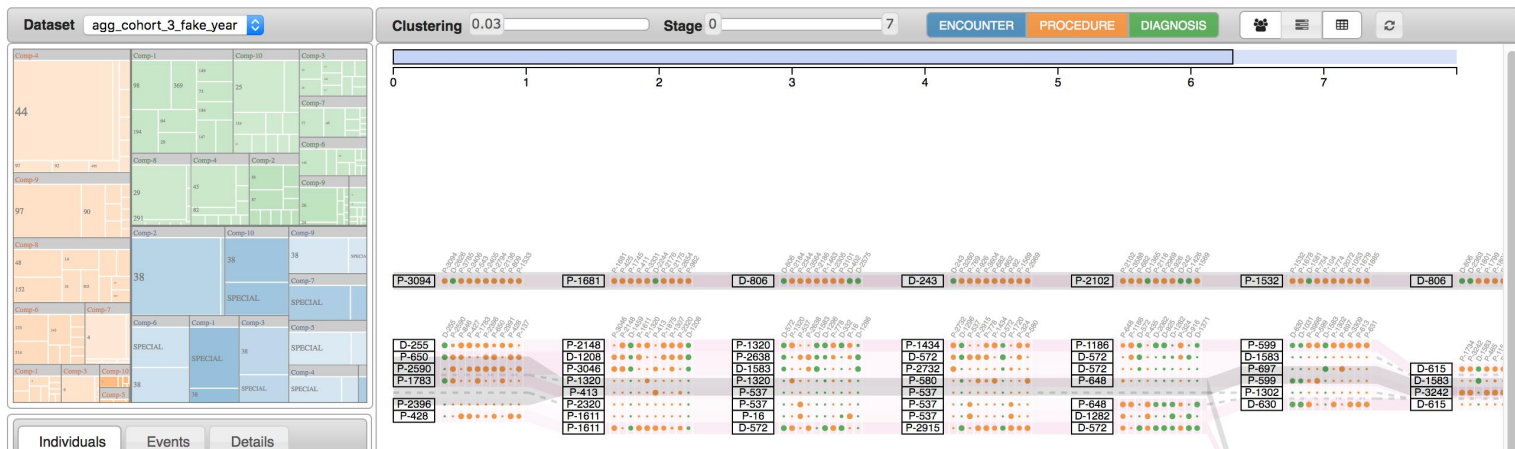




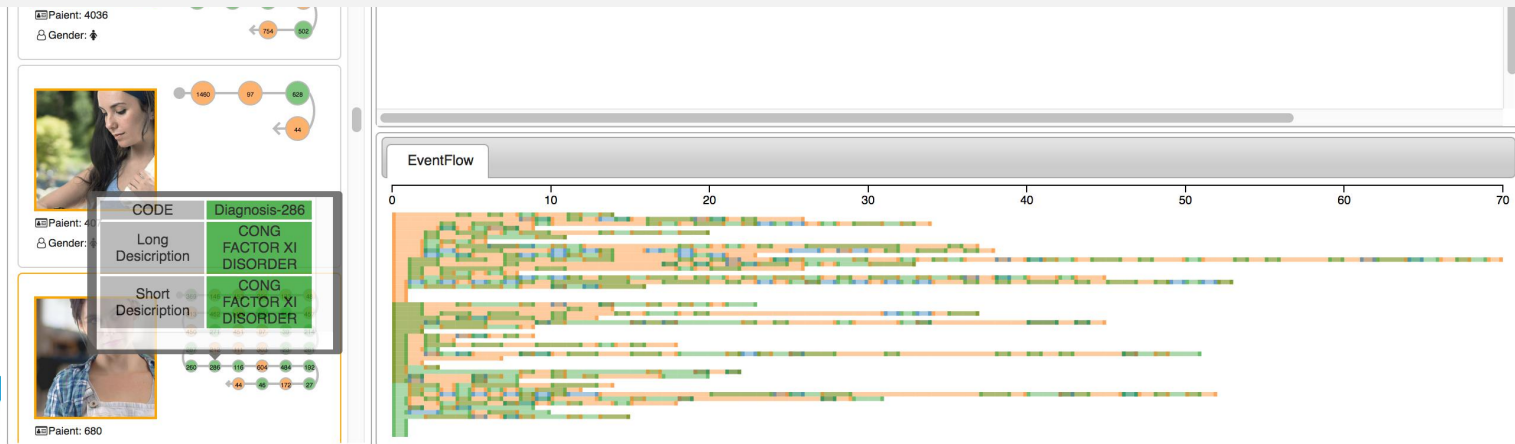
# Key Challenges



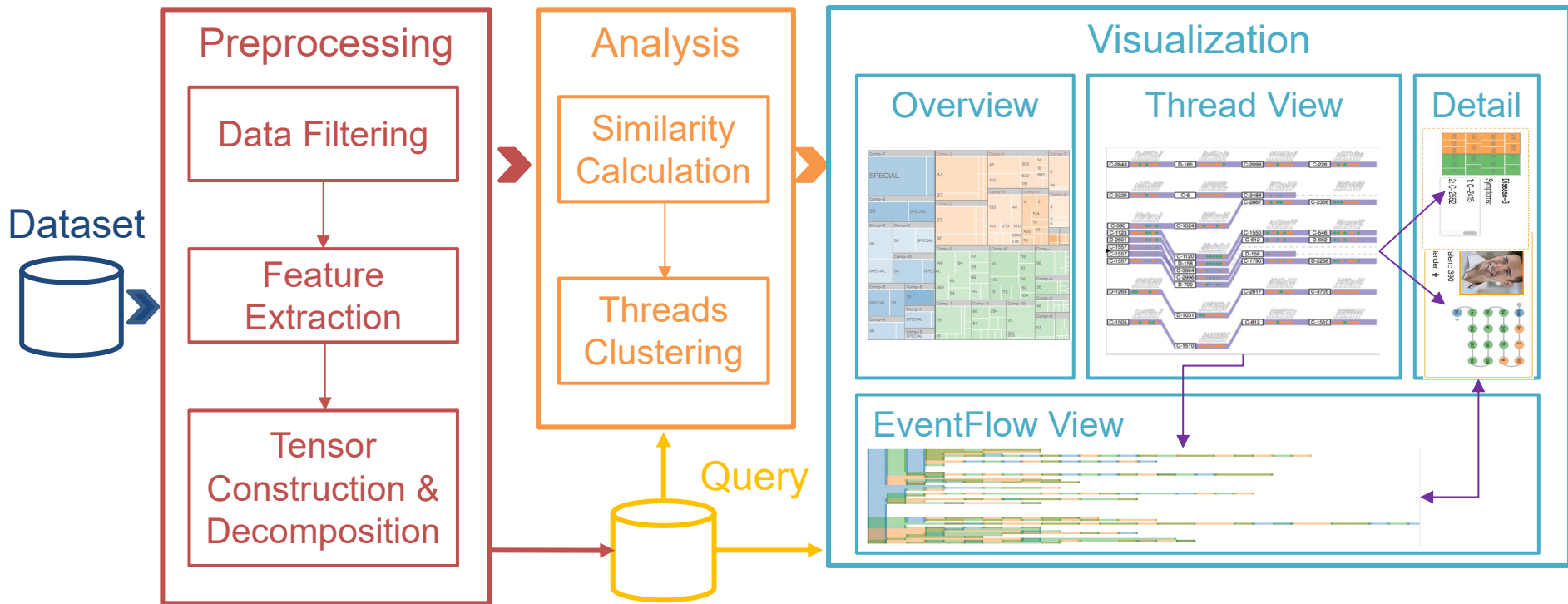
- It is difficult to transform large-scale heterogeneous event sequence data into an uniform data model without losing detailed information
- The method designed to detect high-level structures should also include sufficient relevant context to enable low-level semantic interpretation of what those structures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



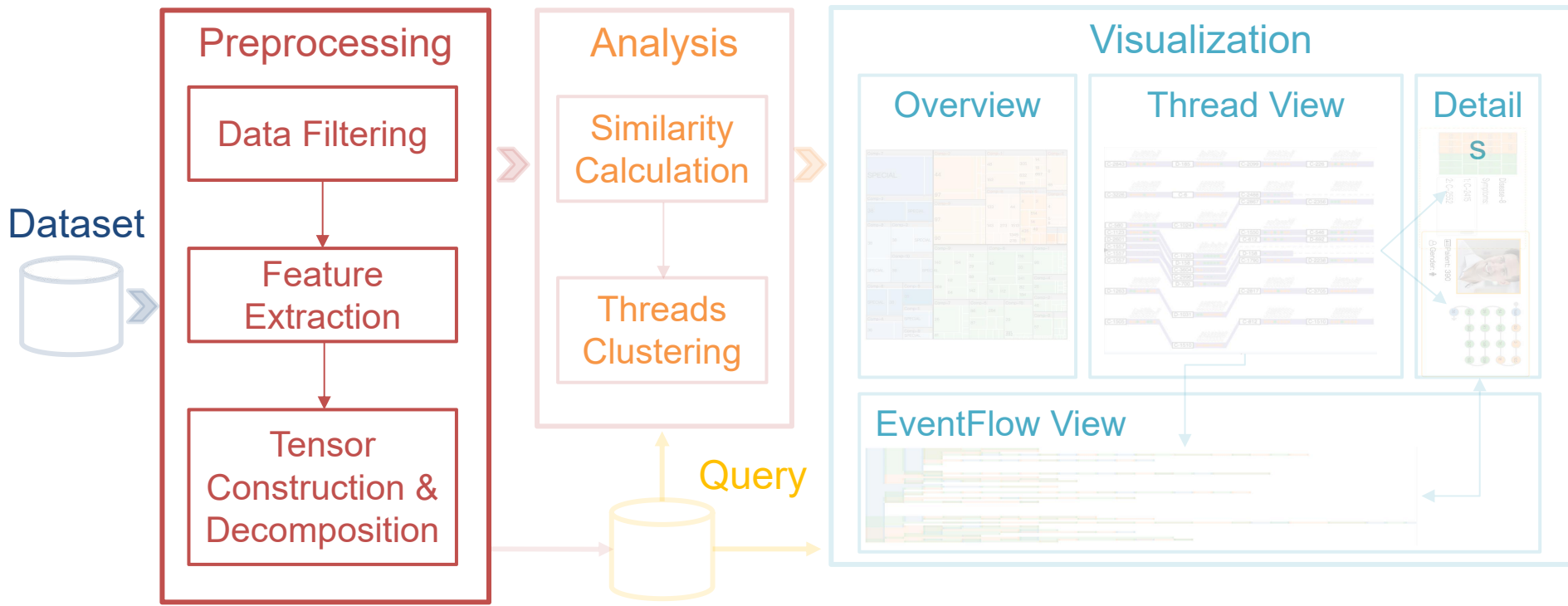
# EventThread System



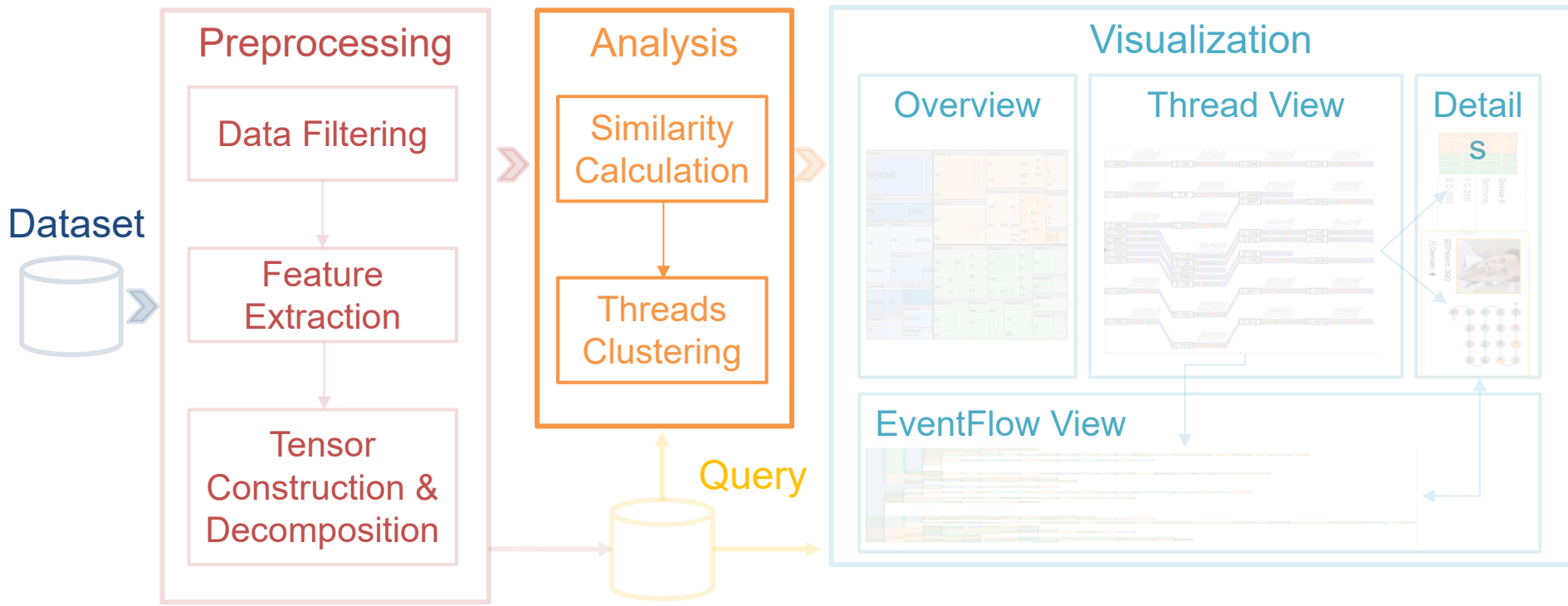
# System Overview



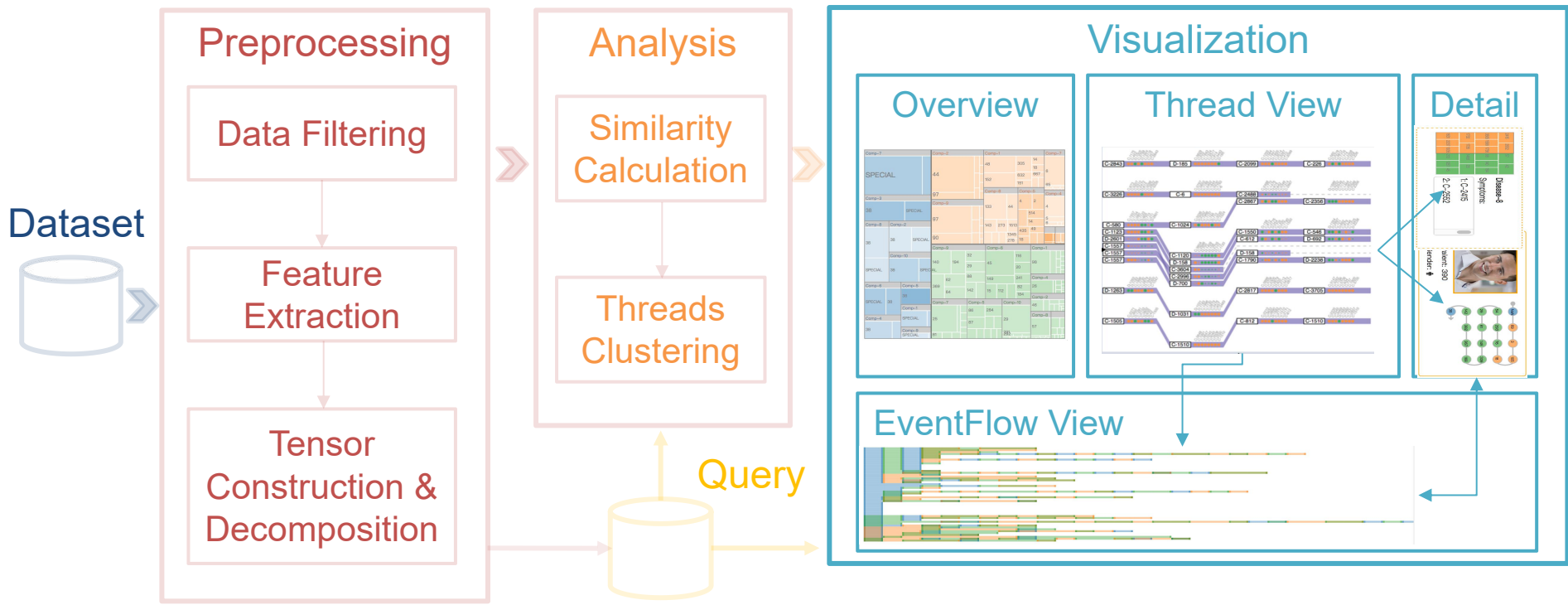
# System Overview



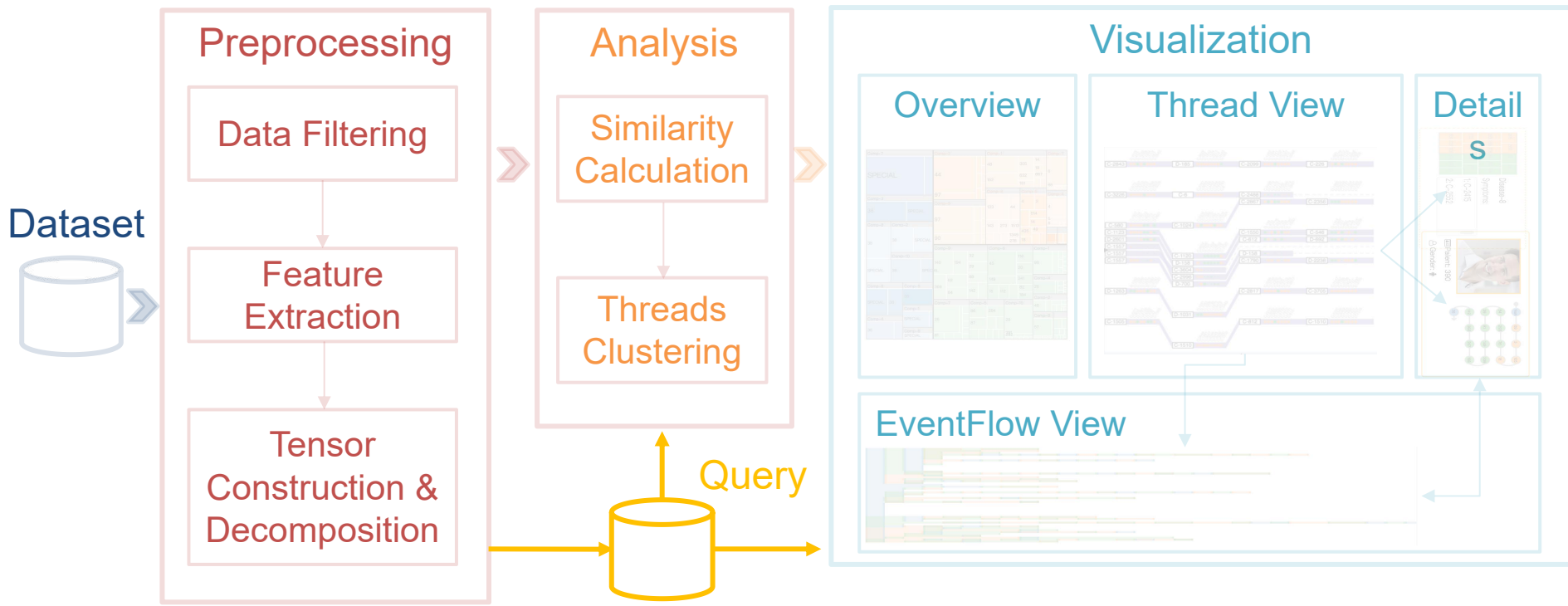
# System Overview



# System Overview

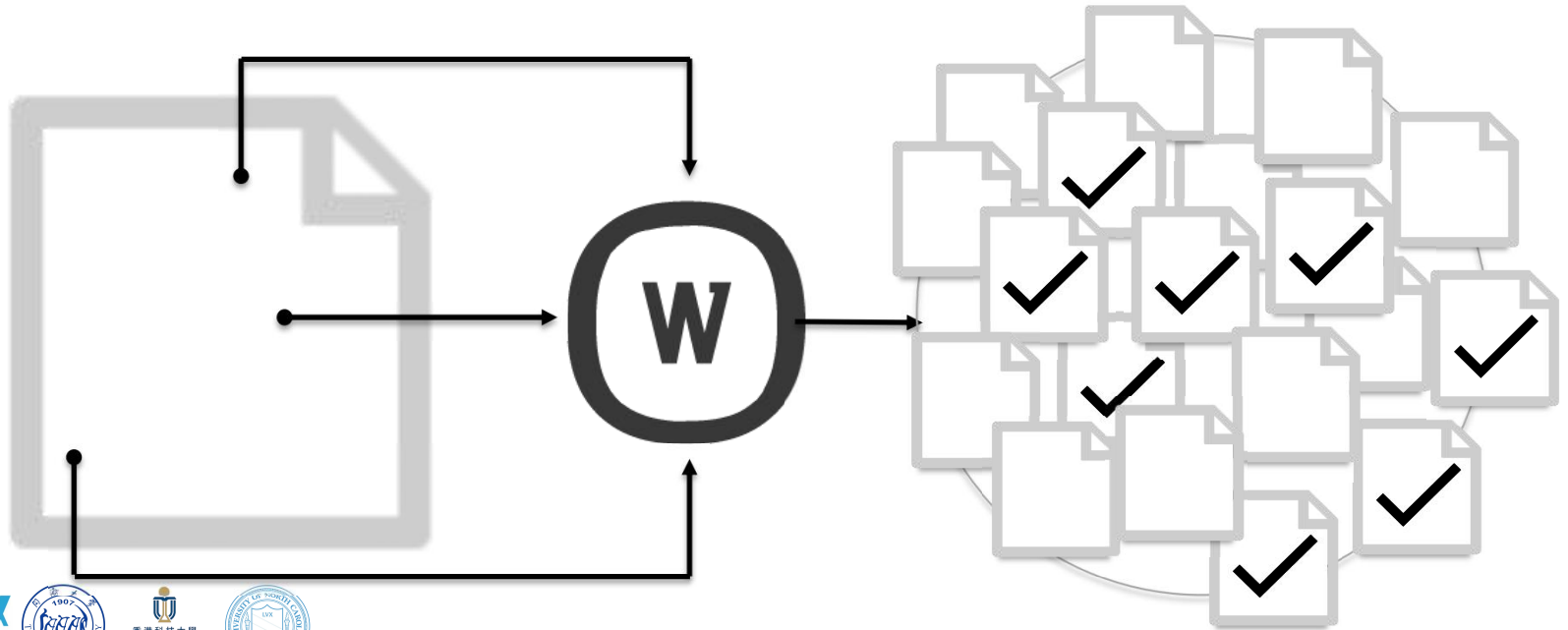


# System Overview



# Data Filtering

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





# Data Filtering

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

$$w_{x,y} = \text{tf}_{x,y} \times \log \left( \frac{N}{\text{df}_x} \right)$$

**TF-IDF**

Term  $x$  within document  $y$

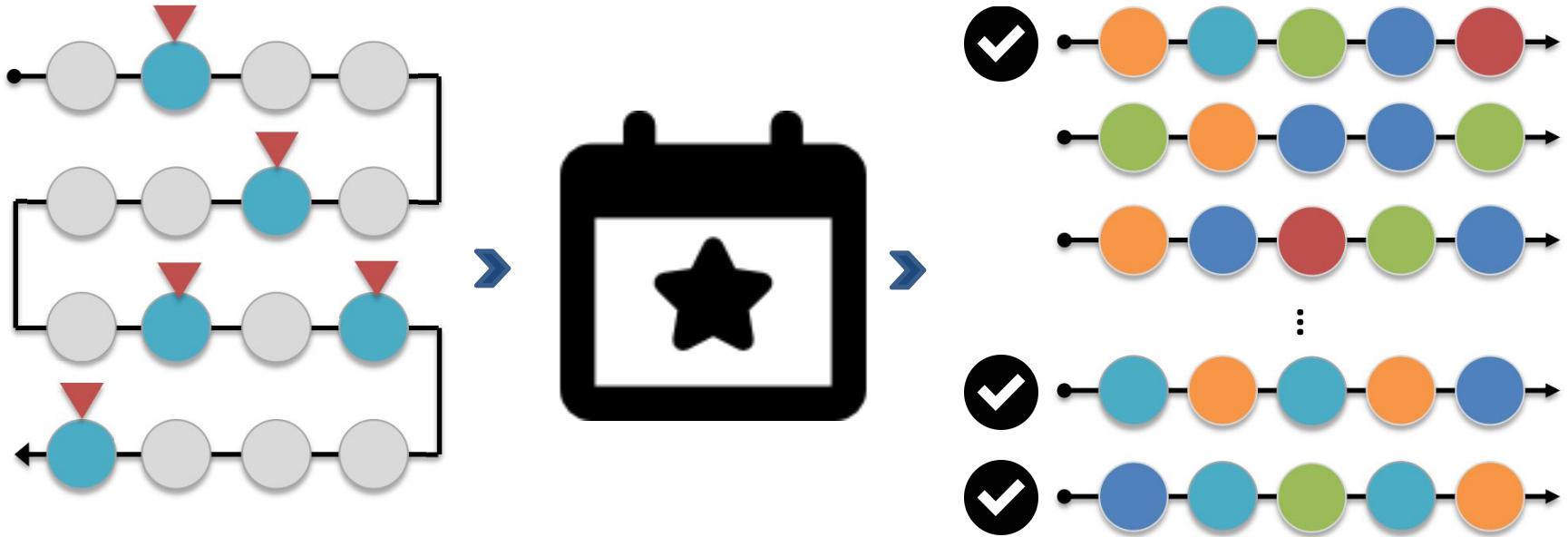
$\text{tf}_{x,y}$  = frequency of  $x$  in  $y$

$\text{df}_x$  = number of documents containing  $x$

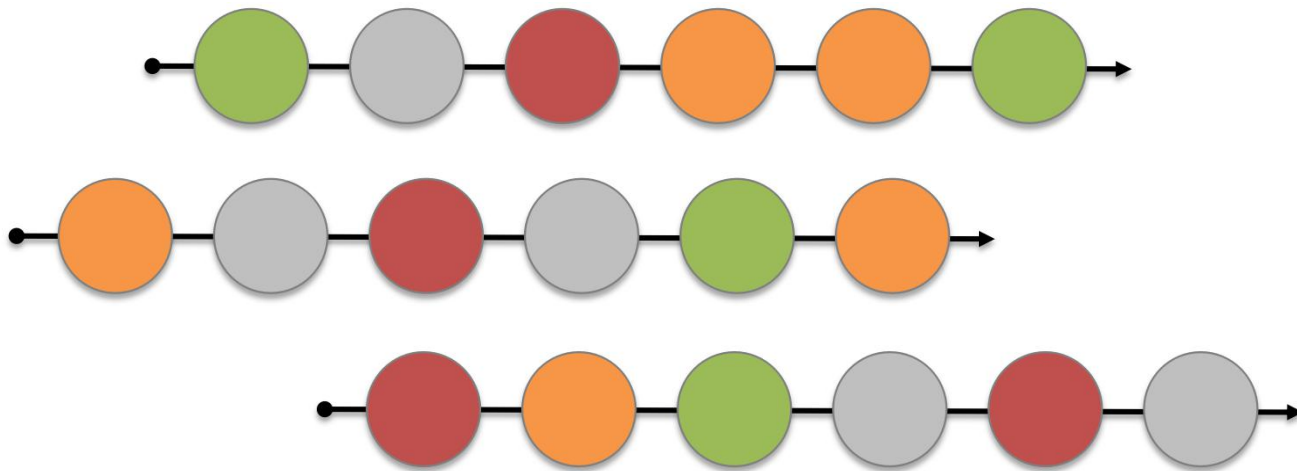
$N$  = total number of documents

# Data Filtering

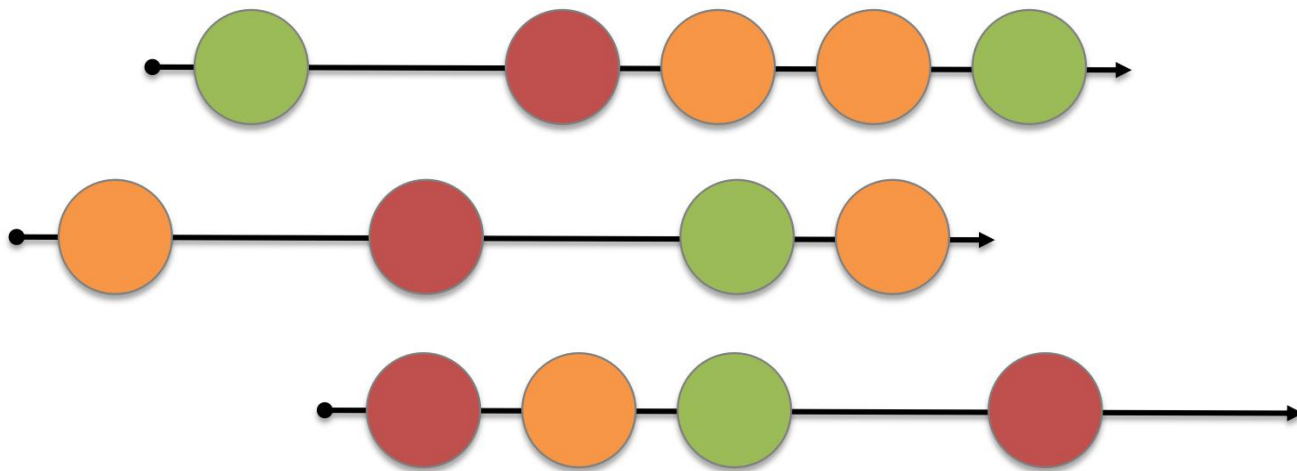
- *TF-IDF* - Event Sequences



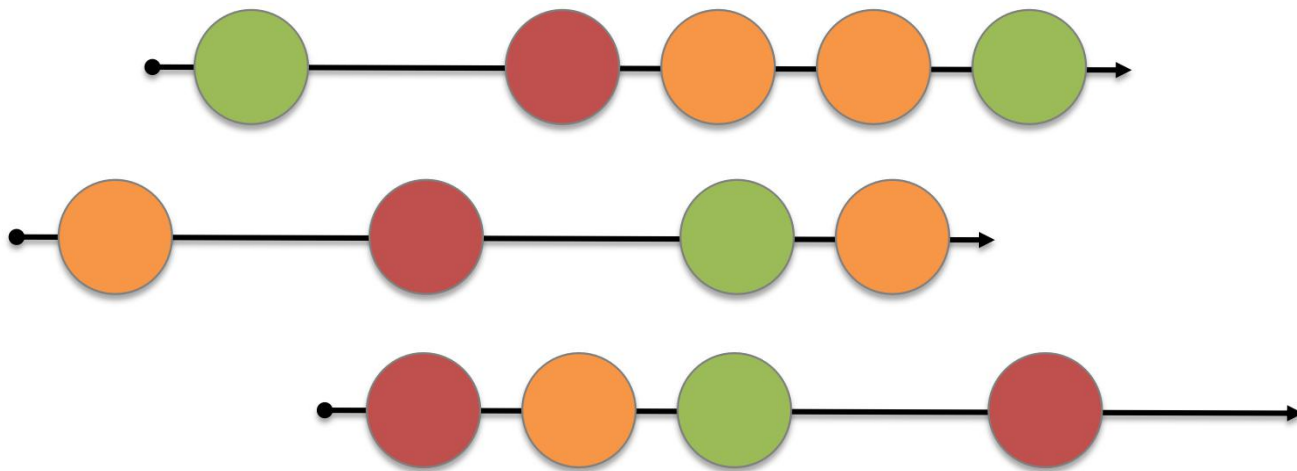
# Sequence Preprocessing



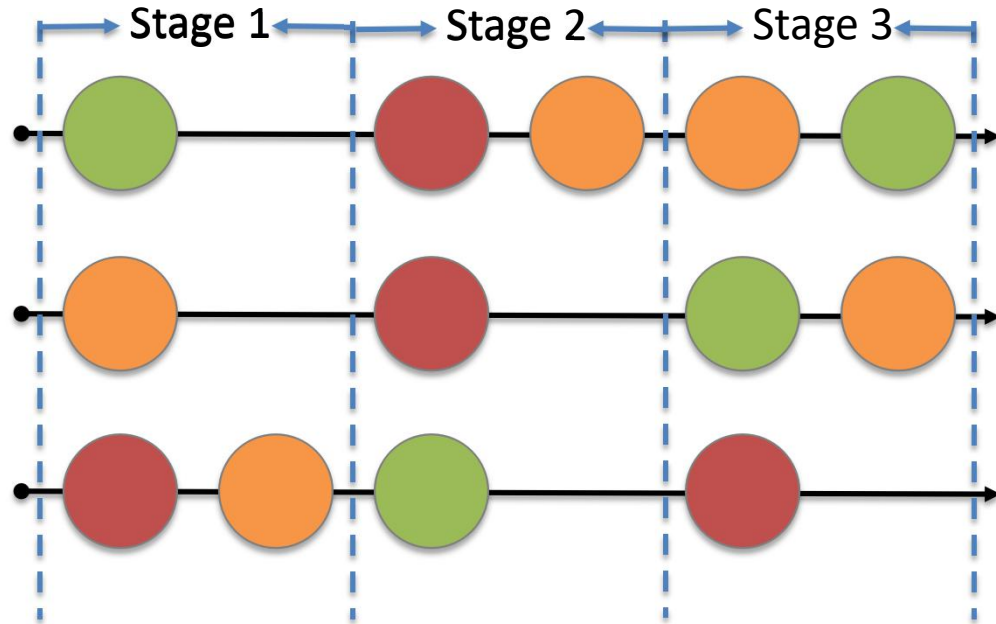
# Sequence Preprocessing



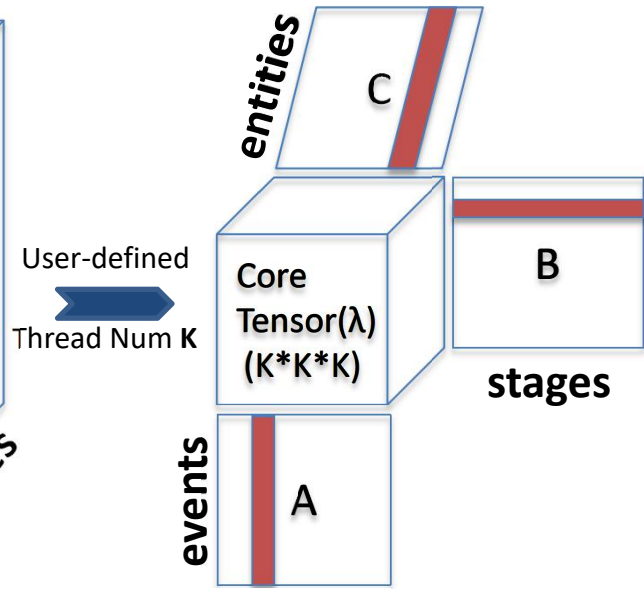
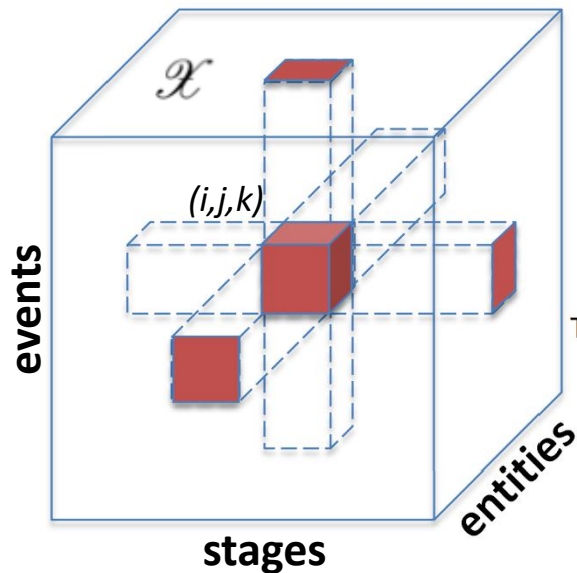
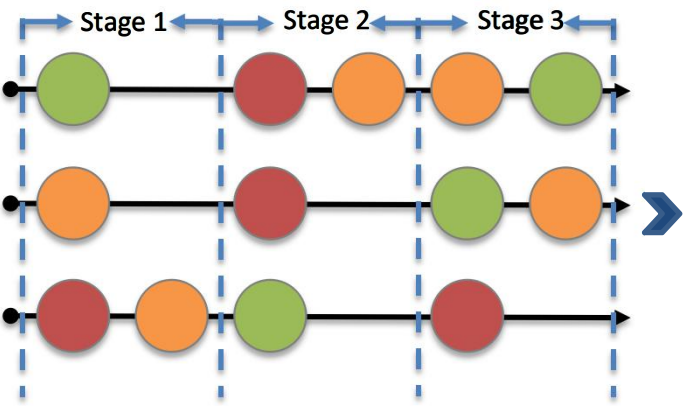
# Sequence Preprocessing



# Sequence Preprocessing



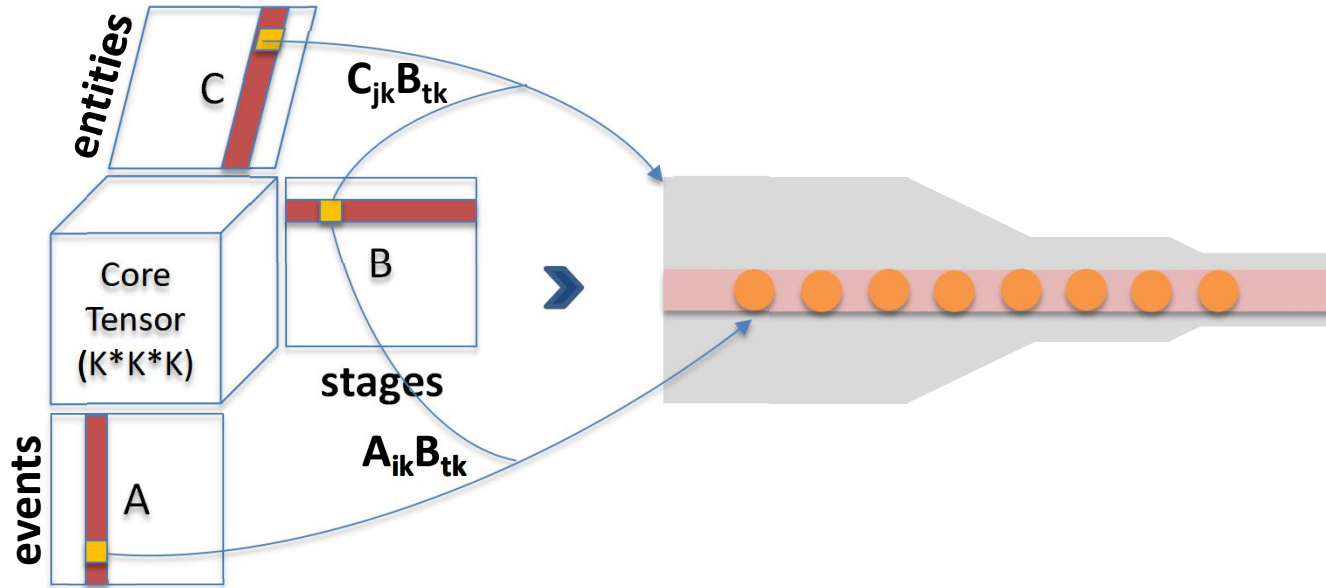
# Tensor Analysis



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

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

# Tensor Analysis

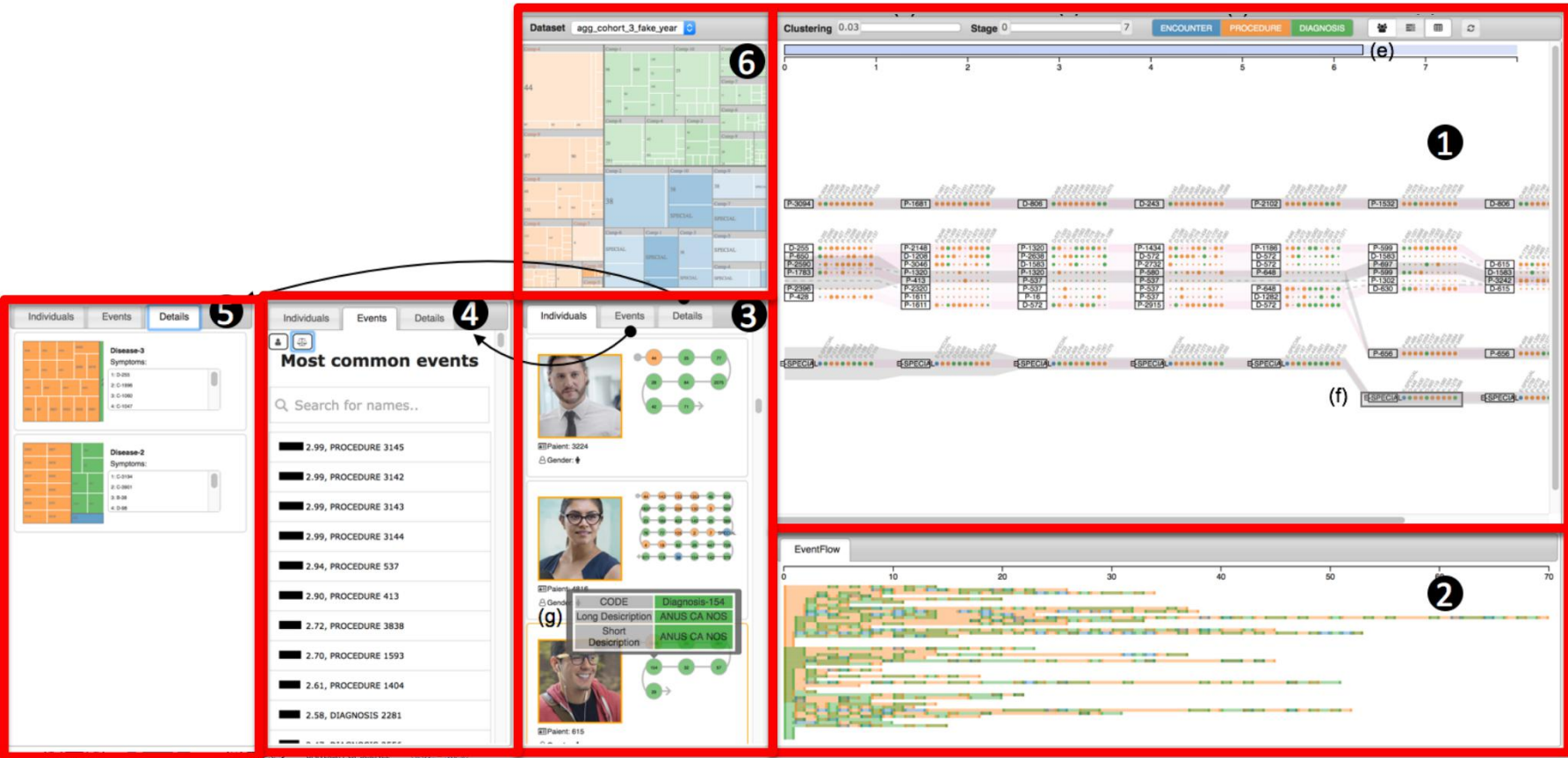




*Introduction*  
**Visualization Design**  
*Evaluation*

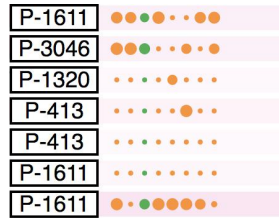


# User Interface

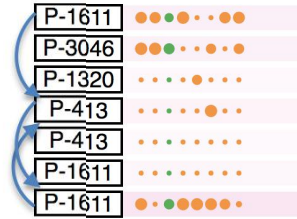


# Creating the Thread View

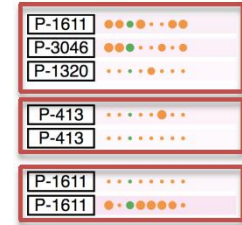
## Visualize Threads



## 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	●●●●●●

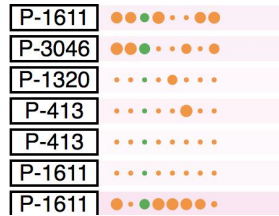
## Rearrange Threads

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

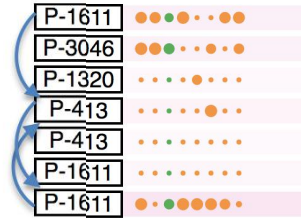
Thread	P-2148	P-1320	P-1434	P-1186
	●●●●●●	●●●●●●	●●●●●●	●●●●●●
P-1611	●●●●●●	P-2638	D-572	D-572
P-3046	●●●●●●	D-1583	P-2732	D-572
P-1320	●●●●●●	P-1320	P-1434	P-648
P-413	●●●●●●	P-537	P-537	P-648
P-413	●●●●●●	P-537	P-537	P-648
P-1611	●●●●●●	P-537	P-537	D-1282
P-1611	●●●●●●	D-572	P-2915	D-572

# Creating the Thread View

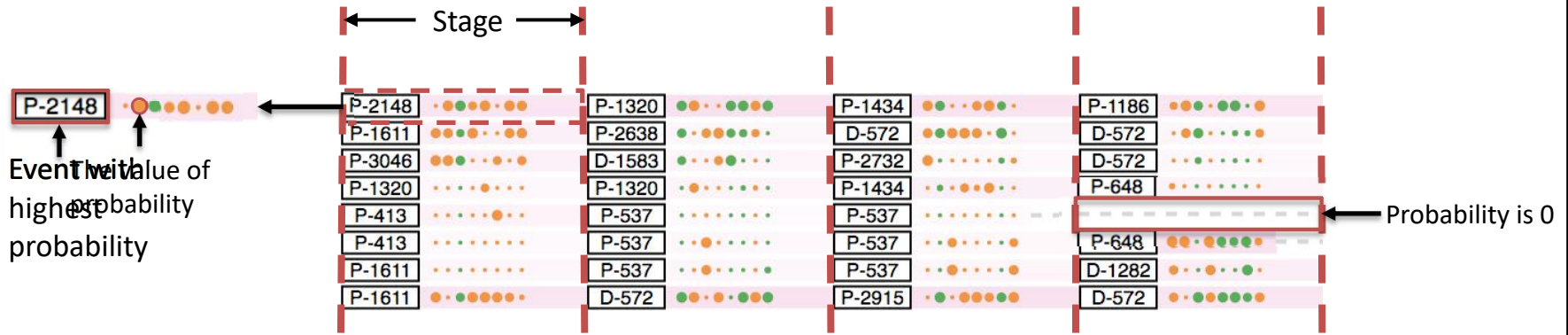
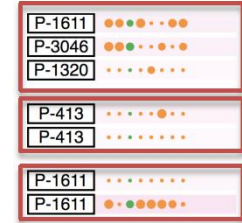
## Visualize Threads



## Layout Threads

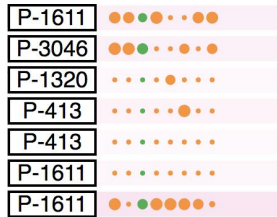


## Rearrange Threads

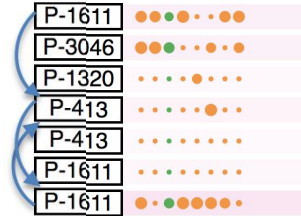


# Creating the Thread View

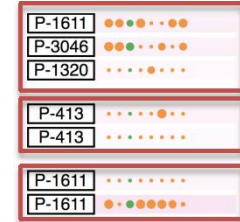
## Visualize Threads



## Layout Threads



## Rearrange Threads



Minimize the distance of similar threads

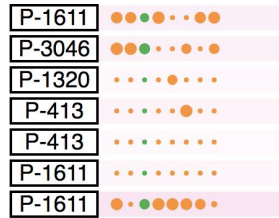
$$\sum_{t=0}^T (\alpha \sum_{i < j} w_{ij}(t) \|y_i(t) - y_j(t)\|^2 + (1 - \alpha) \sum_i \|y_i(t) - y_i(t-1)\|^2)$$

Reduce thread crossing

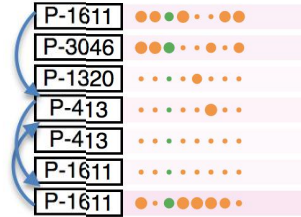
Balance Two Terms

# Creating the Thread View

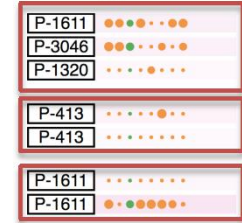
## Visualize Threads



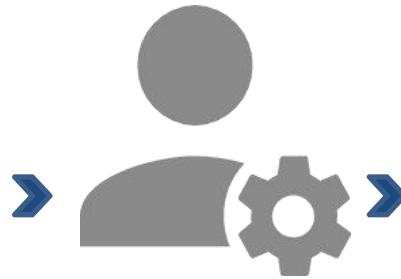
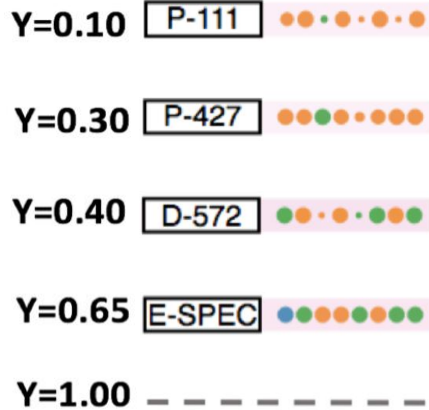
## Layout Threads



## Rearrange Threads

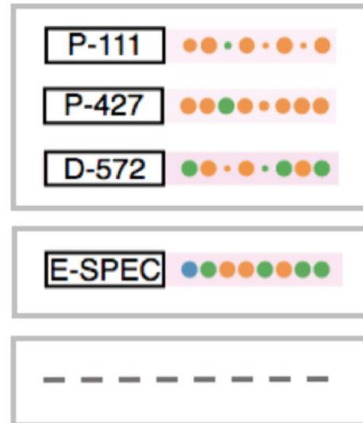


## Initial layout

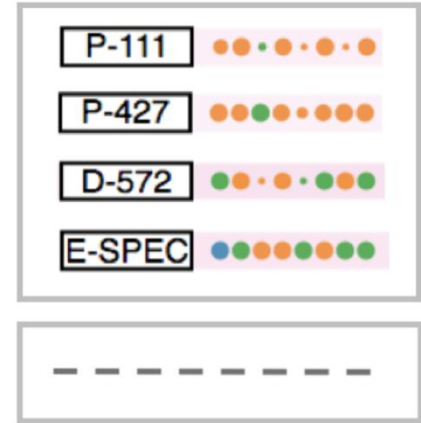


User Defined  
Threshold

## Cluster Threshold = 0.2



## Cluster Threshold = 0.25

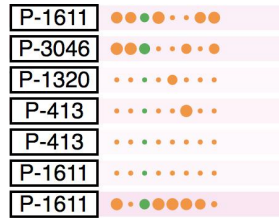




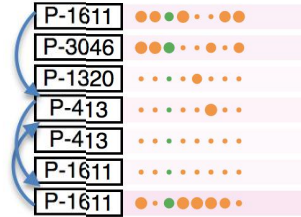


# Creating the Thread View

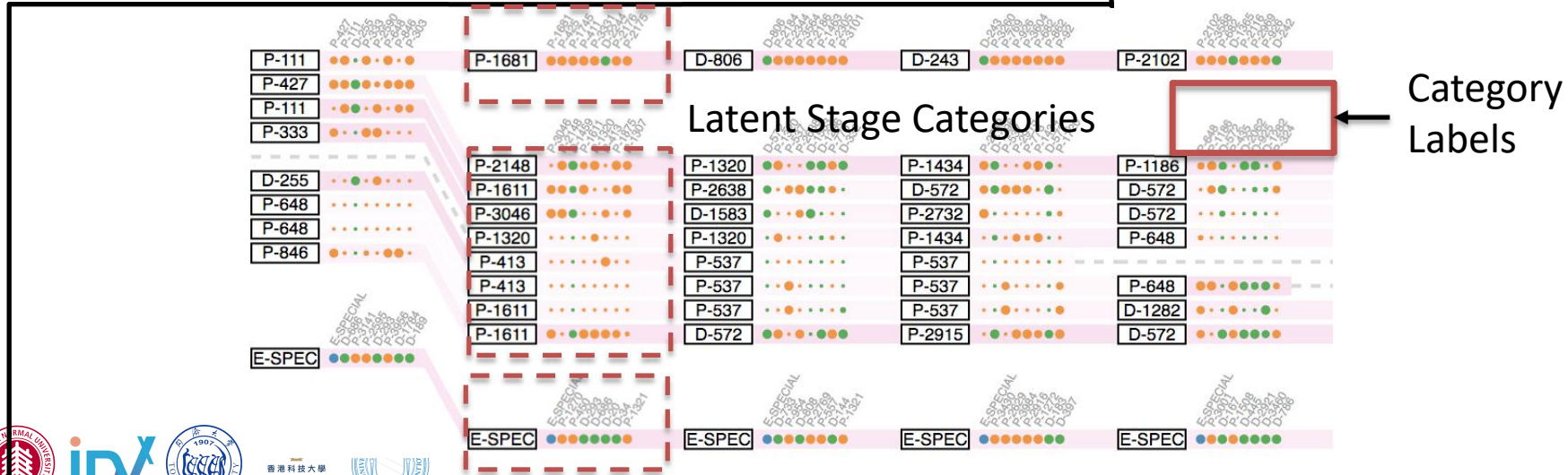
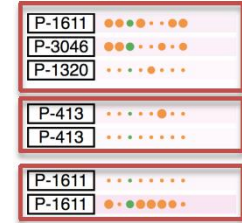
## Visualize Threads



## Layout Threads



## Rearrange Threads



# Creating the Thread View

## Visualize Threads

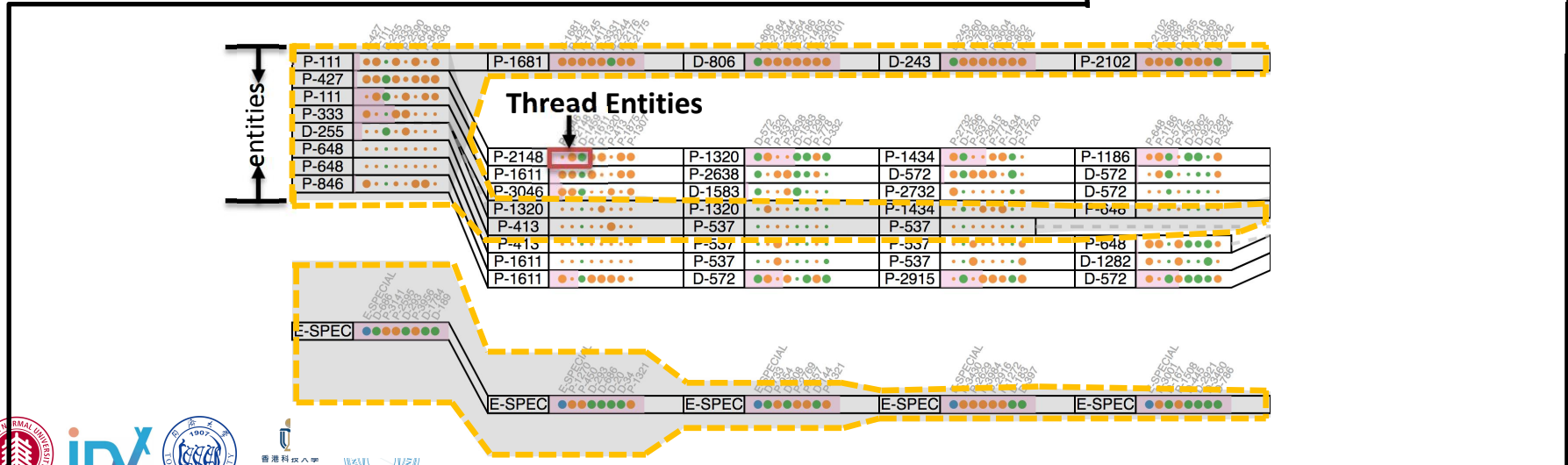
P-1611	●●●●●●
P-3046	●●●●●●
P-1320	●●●●●●
P-413	●●●●●●
P-413	●●●●●●
P-1611	●●●●●●
P-1611	●●●●●●

## Layout Threads

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

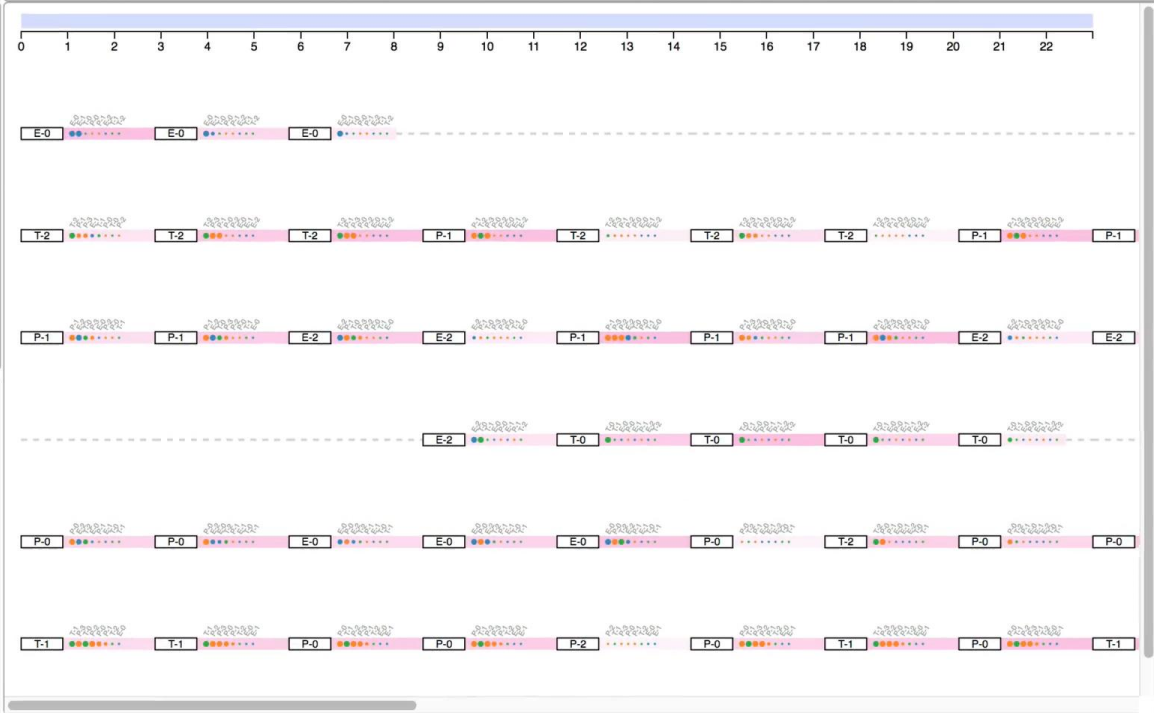
## Rearrange Threads

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





Individuals Events Details



EventFlow



# *Introduction*

# *Visualization Design*

# ***Evaluation***

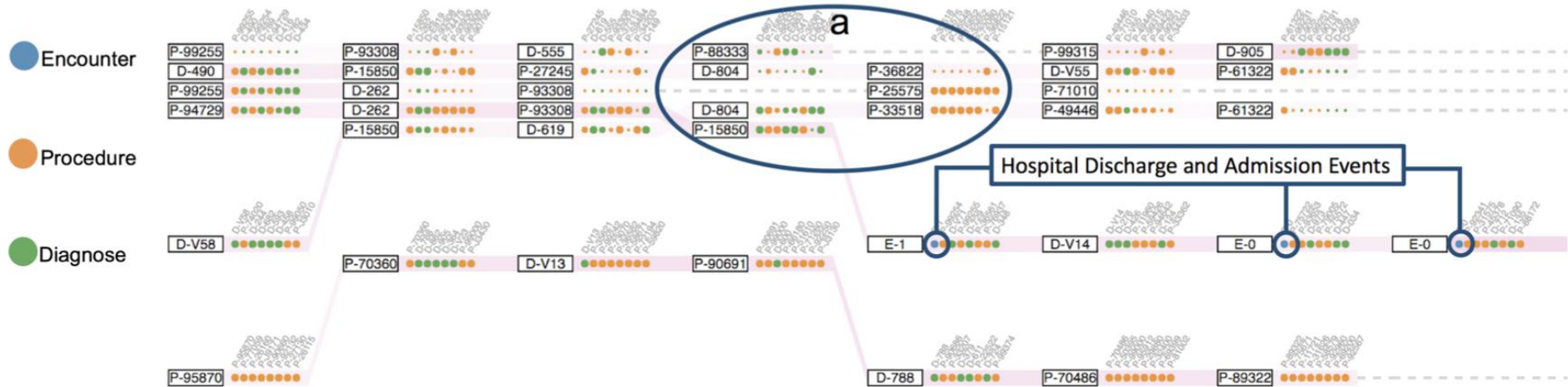


# Usage Scenario: COPD Cohort



- 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

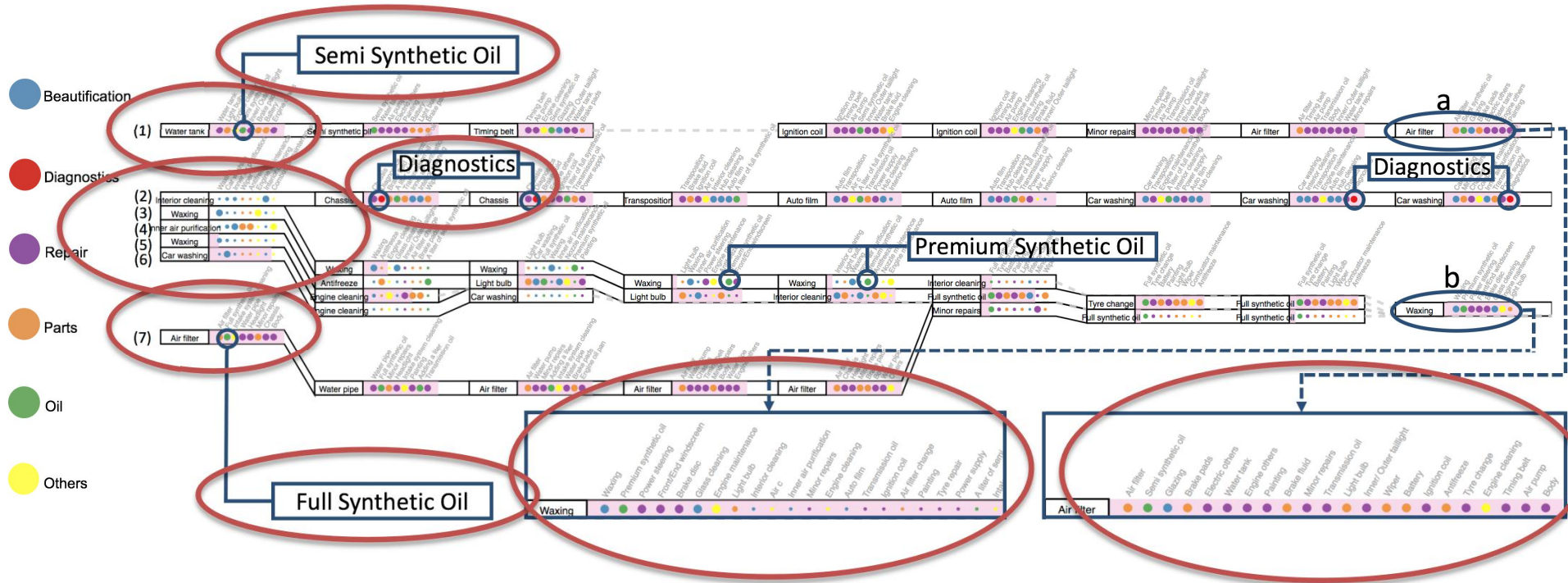
# Usage Scenario: Car Maintenance



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



# Usage Scenario: Car Maintenance

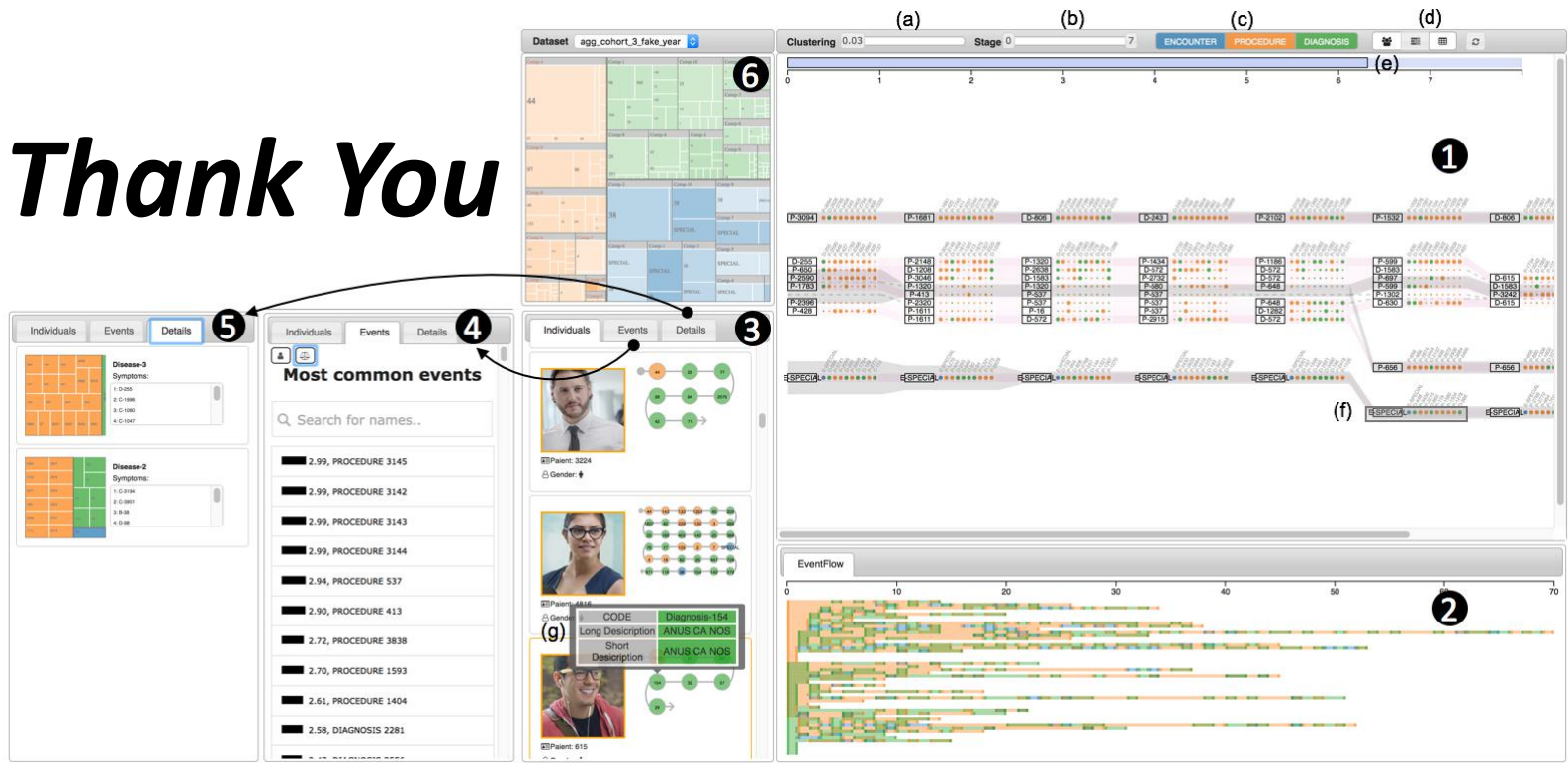


# Conclusion

- 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



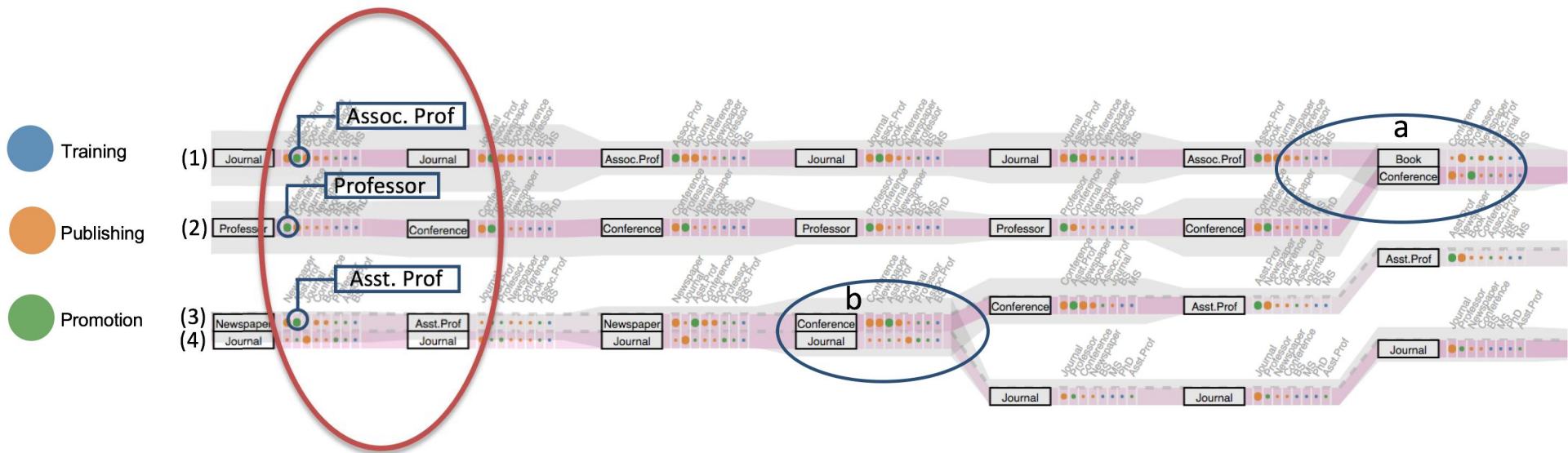
**EventThread: Visual Summarization and Stage Analysis of Event Sequence Data**

# Usage Scenario: Academic Behaviors



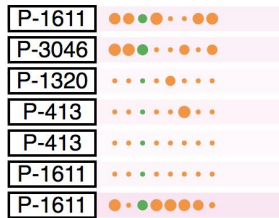
- 40 individuals
- 23 years
- 10 event types, classified into 3 high-level categories: training, publishing, promotion

# Usage Scenario: Academic Behaviors

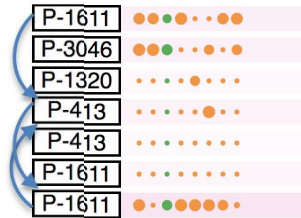


# Creating the Thread View

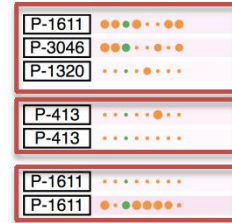
## Visualize Threads



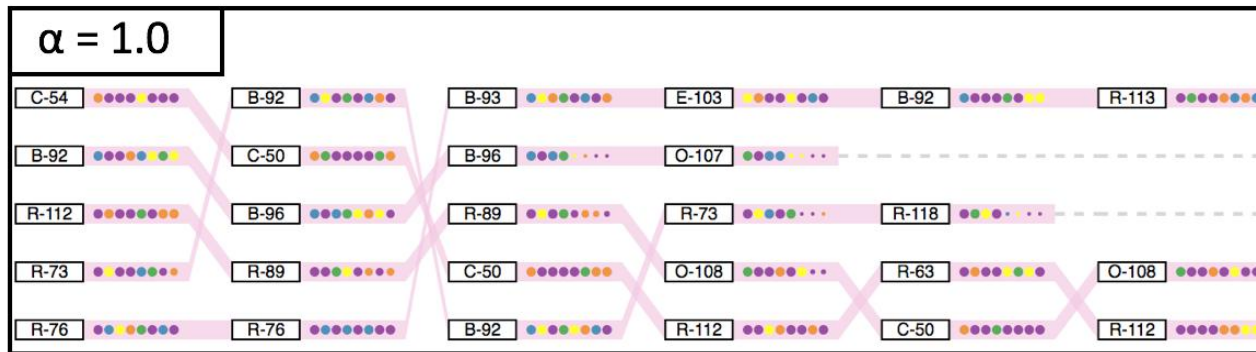
## Layout Threads



## Rearrange Threads

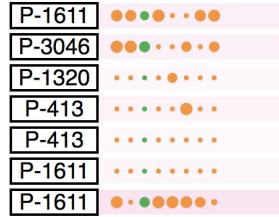


$$\sum_{t=0}^T (\alpha \sum_{i < j} w_{ij}(t) \|y_i(t) - y_j(t)\|^2 + (1 - \alpha) \sum_i \|y_i(t) - y_i(t-1)\|^2)$$

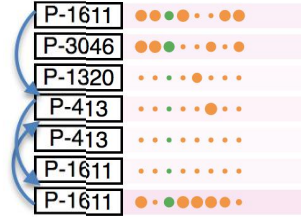


# Creating the Thread View

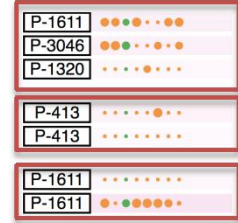
## Visualize Threads



## Layout Threads



## Rearrange Threads



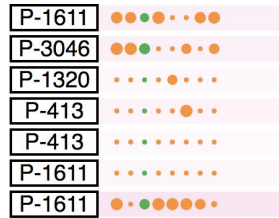
$$\sum_{t=0}^T (\alpha \sum_{i < j} w_{ij}(t) \|y_i(t) - y_j(t)\|^2 + (1 - \alpha) \sum_i \|y_i(t) - y_i(t-1)\|^2)$$

$\alpha = 0$

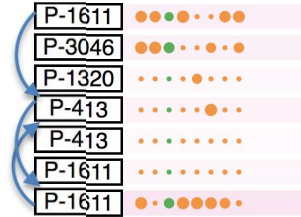


# Creating the Thread View

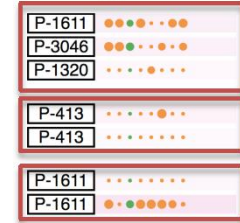
## Visualize Threads



## Layout Threads



## Rearrange Threads



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

$\alpha = 0.5$

